

Stochastik für Informatiker

Holger Kösters

Zusammenfassung

Dies ist die Grundlage für meine Vorlesung „Stochastik für Informatiker“ an der Universität Rostock. Dies ist eine vorläufige Version, die kleinere und größere Fehler enthalten kann. Hinweise auf Unklarheiten und Unrichtigkeiten sind daher willkommen: `holger.koesters@uni-rostock.de` Dies ist eine skizzenhafte Übersicht ohne vollständige Erläuterungen, Begründungen, Beweise usw.; sie kann den Besuch der Vorlesung nur ergänzen, aber nicht ersetzen!

– nicht zur Weiterverbreitung bestimmt –

Abkürzungen:

W. = Wahrscheinlichkeit(s)

EW = Erwartungswert

Ablaufplan:

1. Einführung (Was ist Stochastik?)
2. Diskrete W.räume (Rechenregeln für W.maße)
3. Diskrete W.räume (Kombinatorik) (Zufallsgrößen auf diskreten W.räumen)
4. Diskrete W.räume (Wichtige diskrete W.verteilungen)
5. Allgemeine W.räume (Wichtige stetige W.verteilungen)
6. Allgemeine W.räume (Verteilungsfunktionen) (Zufallsgrößen auf allgemeinen W.räumen)
7. Bedingte Wahrscheinlichkeiten
8. Unabhängigkeit
9. Erwartungswert (Rechenregeln für Erwartungswerte)
10. Erwartungswert, Varianz und Kovarianz
11. Grenzwertsätze für Summen von u.i.v. Zufallsgrößen
12. Statistik (Grundbegriffe) (Schätztheorie)
13. Statistik (Schätztheorie) (Testtheorie)
14. Statistik (Konfidenzbereiche) (Regression)

¹*Date:* January 31, 2020

²*File:* stochinf-024.tex

Inhaltsverzeichnis

0	Einführung (Was ist Stochastik?)	3
1	Diskrete Wahrscheinlichkeitsräume	11
1.1	Diskrete Wahrscheinlichkeitsräume	11
1.2	Zufallsgrößen auf diskreten W.räumen	20
2	Allgemeine Wahrscheinlichkeitsräume	30
2.1	Allgemeine Wahrscheinlichkeitsräume	30
2.2	Zufallsgrößen auf allgemeinen W.räumen	41
3	Bedingte Wahrscheinlichkeiten und Unabhängigkeit	48
3.1	Bedingte Wahrscheinlichkeiten	48
3.2	Unabhängigkeit von Ereignissen	55
3.3	Unabhängigkeit von Zufallsgrößen	57
4	Erwartungswert und Varianz	67
4.1	Erwartungswert	67
4.2	Der Erwartungswert von diskreten Zufallsgrößen	69
4.3	Der Erwartungswert von allgemeinen Zufallsgrößen	75
4.4	Varianz und Kovarianz	78
5	Grenzwertsätze für Summen von u.i.v. Zufallsgrößen	84
5.1	Das Gesetz der großen Zahlen	84
5.2	Der zentrale Grenzwertsatz	88
6	Statistik	94
6.1	Einführende Beispiele	94
6.2	Grundbegriffe der Statistik	96
6.3	Punktschätztheorie	100
6.4	Bereichsschätztheorie	107
6.5	Testtheorie	113
A	Grundwissen Mathematik	121
A.1	Mengen	121
A.2	Tupel	124
A.3	Abbildungen	125
A.4	Folgen und Reihen	126
A.5	Funktionen	129
A.6	Exponentialfunktion und Logarithmusfunktion	133
B	Tabellen	135
C	Symbolverzeichnis	139

0 Einführung (Was ist Stochastik?)

Begriff „Stochastik“.

- altgr. στόχος Ziel / Vermutung; στοχαστική τέχνη „Kunst des Vermutens“
- Teilgebiet der Mathematik, das sich mit der math. Beschreibung, Untersuchung und Auswertung von zufallsabhängigen Vorgängen befasst; besteht aus den Teilgebieten *Wahrscheinlichkeitstheorie* und *mathematische Statistik*
 - typische Frage in der Wahrscheinlichkeitstheorie: Eine „faire“ Münze wird 100mal geworfen. Wie groß ist die W., dass dabei (genau) 70mal Kopf fällt?
 - typische Frage in der mathematischen Statistik: Eine Münze wird 100mal geworfen; dabei fällt (genau) 70mal Kopf. Lässt dies den Schluss zu, dass die Münze nicht „fair“ ist?
- Lehre vom Zufall

Warum „Stochastik“ ?

Viele Beobachtungen hängen (zumindest teilweise) vom Zufall ab oder können vereinfachend als zufällig angesehen werden:

- Ergebnisse bei Glücksspielen (Würfel, Münzen, Kugeln, ...)
- Anzahl von Geburten / Sterbefällen
- Anzahl von Erdbeben / Unfällen / radioaktiven Zerfällen / Rechnerabstürzen / Server-Aufrufen / defekten Produkten / verkauften Produkten / ...
- Niederschlags-W./-menge/-dauer am nächsten Tag / im nächsten Monat
- Aktienkurs am nächsten Tag / im nächsten Jahr
- Laufzeit eines Algorithmus
- Reaktionszeit eines IT-Systems
- Lebensdauer eines technischen Geräts
- Beschaffenheit (z. B. Zusammensetzung) eines Materials
- Messfehler bei physikalischen oder technischen Experimenten

Um in diesen Situationen „gute“ Entscheidungen treffen zu können, ist es häufig notwendig, die vorliegenden Unsicherheiten quantitativ zu beschreiben und zu untersuchen; dazu benötigt man die Sprache der Mathematik bzw. insbesondere der Stochastik.

weiteres Stichwort: Data Science

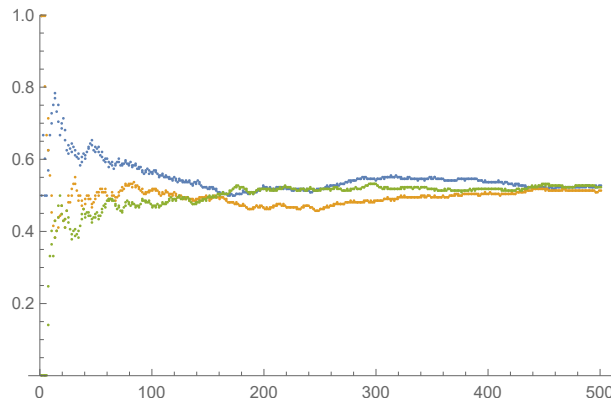
Eigenschaften (idealer) zufallsabhängiger Vorgänge.

- es gibt mehrere mögliche Versuchsausgänge (Ergebnisse)
- jede Durchführung liefert ein eindeutiges Ergebnis
- es ist nicht vorhersagbar, welches Ergebnis eintritt
- das Experiment kann (zumindest im Prinzip) unter den gleichen Bedingungen wiederholt werden
- ob der Vorgang echt zufällig ist (\leftarrow *Indeterminiertheit*) oder nur zufällig erscheint (\leftarrow *Unkenntnis*), bleibt dabei ungeklärt

Zufallsabhängige Vorgänge sind dadurch gekennzeichnet, dass (zunächst) keine erkennbaren Gesetzmäßigkeiten vorliegen.

Wiederholt man denselben zufallsabhängigen Vorgang mehrmals (und unabhängig voneinander), so treten erfahrungsgemäß Regelmäßigkeiten auf:

Beispiel: Wirft man eine Münze (mit den Seiten „Kopf“ und „Zahl“) n -mal, so lassen sich die einzelnen Ergebnisse nicht vorhersagen – die Erfahrung lehrt uns allerdings, dass für großes n die Anzahl ($=$: *absolute Häufigkeit*) von „Kopf“ etwa np bzw. der Anteil ($=$: *relative Häufigkeit*) von „Kopf“ etwa $\frac{np}{n} = p$ beträgt, wobei $p \in [0, 1]$ eine Konstante ist, die nur von der Münze abhängt.



Entwicklung der relativen Häufigkeit von Kopf beim Münzwurf (3 Versuchsreihen)

Es lässt sich also eine Stabilisierung von relativen Häufigkeiten beobachten. (*Empirisches Gesetz der großen Zahlen*)

Diese Gesetzmäßigkeiten sollen in der Stochastik genauer beschrieben und untersucht werden.

Mathematische Beschreibung (einfacher) zufallsabhängiger Vorgänge.

- Menge Ω der möglichen Versuchsausgänge (*Ergebnisse*)
- Zahlen $f(\omega)$, $\omega \in \Omega$, welche die „Chance des Eintretens“ der einzelnen Versuchsausgänge beschreiben (*Wahrscheinlichkeiten*);

Erläuterung:

- Idee: Wahrscheinlichkeit \approx relative Häufigkeit
- Wird der zufallsabhängige Vorgang n -mal wiederholt und tritt dabei k -mal das Ergebnis $\omega \in \Omega$ auf, so heißt $H_n(\omega) = k$ *absolute Häufigkeit von ω* und $h_n(\omega) = k/n$ *relative Häufigkeit von ω* .
- $f(\omega) = p$ bedeutet, dass $h_n(\omega)$ für großes n „etwa“ bei p liegt. (Dies ist eine Annahme!)
- Aus den Eigenschaften von relativen Häufigkeiten ergeben sich die Eigenschaften von Wahrscheinlichkeiten: $f(\omega) \geq 0$ für alle $\omega \in \Omega$, $\sum_{\omega \in \Omega} f(\omega) = 1$. (Dabei sei Ω zunächst als endlich vorausgesetzt.)

Bemerkungen:

- Häufig interessieren nicht nur die W.en von einzelnen *Ergebnissen*, sondern auch die W.en von *Ereignissen*, die verschiedene Ergebnisse umfassen. (Beim Wurf eines gewöhnlichen Würfels könnte man z.B. als Grundmenge $\Omega = \{1, 2, 3, 4, 5, 6\}$ wählen und das Ereignis A : „gerade Zahl“ betrachten, das die Ergebnisse 2, 4, 6 umfasst. Die W.en dieser Ereignisse erhält man dadurch, dass man die W.en der zugehörigen Ergebnisse addiert.
- Wir legen hier (zur Motivation und zur Interpretation von Wahrscheinlichkeiten) den *frequentistischen Wahrscheinlichkeitsbegriff* zugrunde. Es gibt auch andere Wahrscheinlichkeitsbegriffe, vgl. dazu die Bemerkungen zur Bestimmung von Wahrscheinlichkeiten weiter unten.

Beispiele.

1. Eine Münze (mit den Seiten „Kopf“ und „Zahl“) wird geworfen.

Ansatz 1: Wir unterstellen, dass die Münze „symmetrisch“ / „fair“ ist.

- $\Omega = \{K, Z\}$, wobei $K \triangleq$ „Kopf“, $Z \triangleq$ „Zahl“
- $f(K) = 1/2$, $f(Z) = 1/2$ (wegen *Symmetrie* / *Fairness*)

Ansatz 2: Wir unterstellen, dass die Münze „gefälscht“ / „gezinkt“ / „unfair“ ist.

- $\Omega = \{K, Z\}$, wobei $K \triangleq$ „Kopf“, $Z \triangleq$ „Zahl“
- $f(K) = p$, $f(Z) = 1 - p$ (wobei $p \in [0, 1]$ *unbekannter Parameter*)

Bemerkung: Alternativ können wir $\Omega = \{0, 1\}$ wählen, wobei $0 \triangleq$ „Zahl“, $1 \triangleq$ „Kopf“.

2. *Drei faire Münzen werden in der Hand geschüttelt und dann gleichzeitig geworfen. Mit welcher W. erscheint „genau einmal Kopf“?*

Ansatz 1: Münzen nicht unterscheidbar

- $\Omega = \{(K, K, K), (K, K, Z), (K, Z, Z), (Z, Z, Z)\}$, wobei Teilergebnisse „sortiert“
- $f(K, K, K) = ???, f(K, K, Z) = ???, f(K, Z, Z) = ???, f(Z, Z, Z) = ???$

Die Annahme, dass die einzelnen Ergebnisse die gleiche Chance besitzen, erscheint hier nicht angemessen: Führt man das Experiment wiederholt durch, so stellt man nämlich fest, dass die relativen Häufigkeiten von (K, K, Z) und (K, Z, Z) etwa bei $\frac{3}{8}$, die relativen Häufigkeiten von (K, K, K) und (Z, Z, Z) hingegen etwa bei $\frac{1}{8}$ liegen. Um dies zu erklären, versehen wir die Münzen gedanklich mit den Nummern 1, 2, 3. Dann gibt es bei (K, K, Z) und (K, Z, Z) jeweils 3 Möglichkeiten, wie diese Ergebnisse zustande kommen, bei (K, K, K) und (Z, Z, Z) dagegen nur jeweils 1 Möglichkeit. Die Annahme, dass die einzelnen Möglichkeiten die gleiche Chance besitzen, führt daher auf die folgende Festlegung:

- $f(K, K, K) = \frac{1}{8}, f(K, K, Z) = \frac{3}{8}, f(K, Z, Z) = \frac{3}{8}, f(Z, Z, Z) = \frac{1}{8}$ (wegen Symmetrie auf der Ebene der Möglichkeiten, wie die Ergebnisse zustande kommen)
 - A : „genau einmal Kopf“, $A = \{(K, Z, Z)\}$
 - $\mathbb{P}(A) = f(K, Z, Z) = \frac{3}{8}$
- \rightsquigarrow „Genau einmal Kopf“ erscheint mit W. $\mathbb{P}(A) = \frac{3}{8}$.

Ansatz 2: Münzen unterscheidbar

Hier nehmen wir z. B. an, dass wir faire Münzen mit unterschiedlichen Jahreszahlen verwenden; dies sollte natürlich keinen Einfluss auf die gesuchte W. haben!

- $\Omega = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{K, Z\} \text{ für alle } i = 1, 2, 3\}$, wobei $\omega_i \triangleq$ (Teil-)Ergebnis bei der i -ten Münze
 - $f(\omega) = \frac{1}{8}$ für alle $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$ (wegen Symmetrie)
 - A : „genau einmal Kopf“, $A = \{(K, Z, Z), (Z, K, Z), (Z, Z, K)\}$
 - $\mathbb{P}(A) = f(K, Z, Z) + f(Z, K, Z) + f(Z, Z, K) = \frac{3}{8}$
- \rightsquigarrow „Genau einmal Kopf“ erscheint mit W. $\mathbb{P}(A) = \frac{3}{8}$.

Merkregel: Umfasst ein Ereignis mehrere Ergebnisse, so erhalten wir die W. des Ereignisses, indem wir die W.en der zugehörigen Ergebnisse addieren.

Bemerkung: Ob man das erste oder zweite Modell wählt, kann man sich aussuchen. Wichtig ist, dass die Beschreibung von A und die Berechnung von $\mathbb{P}(A)$ zum Modell passen. Bei Ansatz 2 lässt sich die Wahl der W.'en $f(\omega)$ einfacher / klarer begründen; dafür ist die Menge Ω etwas größer / unübersichtlicher.

3. Eine verbogene Münze wird dreimal nacheinander geworfen. Mit welcher W. erscheint „genau einmal Kopf“?

Ansatz:

- $\Omega = \{(\omega_1, \omega_2, \omega_3) : \omega_i \in \{K, Z\} \text{ für alle } i = 1, 2, 3\}$, wobei $\omega_i \triangleq$ (Teil-)Ergebnis im i -ten Wurf
- $f(\omega) = p_{\omega_1} p_{\omega_2} p_{\omega_3}$ für alle $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$, wobei $p_K := p$, $p_Z := 1 - p$ und $p \in [0, 1]$ ein unbekannter Parameter ist, der die W. für Kopf beim einmaligen Münzwurf angibt (wegen Unabhängigkeit)

ausführliche Begründung: über relative Häufigkeiten

Angenommen, der Versuch (d.h. der 3-fache Münzwurf) wird n mal wiederholt.

Von diesen n Versuchen liefern etwa np_{ω_1} Versuche ω_1 im 1. Wurf.

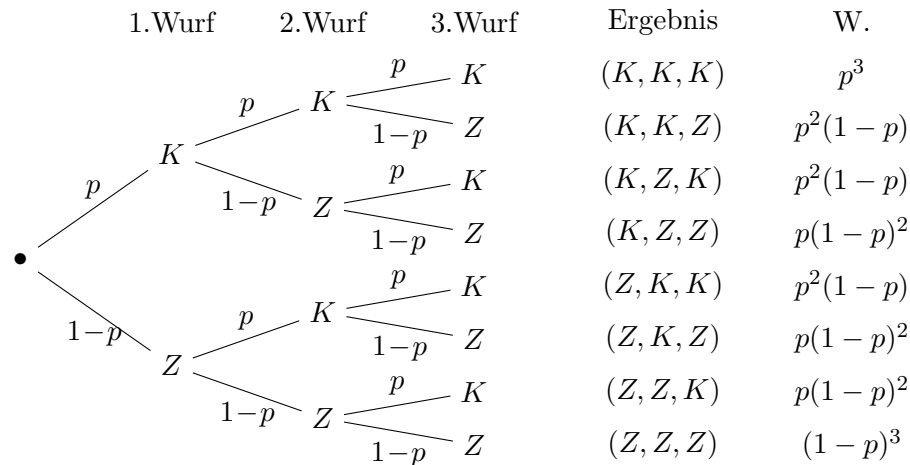
Von diesen np_{ω_1} Versuchen liefern etwa $np_{\omega_1} p_{\omega_2}$ Versuche ω_2 im 2. Wurf.

Von diesen $np_{\omega_1} p_{\omega_2}$ Versuchen liefern etwa $np_{\omega_1} p_{\omega_2} p_{\omega_3}$ Versuche ω_3 im 3. Wurf.

Damit ist die Festlegung $f(\omega) := p_{\omega_1} p_{\omega_2} p_{\omega_3}$ nahe liegend.

Dabei haben wir angenommen, dass die einzelnen (Teil-)Ergebnisse *unabhängig voneinander* sind, also z. B. das (Nicht-)Erscheinen von Kopf im 1. Wurf keinen Einfluss auf das (Nicht-)Erscheinen von Kopf im 2. Wurf hat.

Diese Informationen lassen sich schön in einem W.baum darstellen:



- A : „genau einmal Kopf“, $A = \{(K, Z, Z), (Z, K, Z), (Z, Z, K)\}$

- $\mathbb{P}(A) = f(K, Z, Z) + f(Z, K, Z) + f(Z, Z, K) = 3p(1-p)^2$

\rightsquigarrow „Genau einmal Kopf“ erscheint mit W. $3p(1-p)^2$.

Merkregel: Ist ein Gesamtexperiment aus mehreren Teilexperimenten zusammengesetzt, die *unabhängig voneinander* durchgeführt werden, so erhalten wir die W.en für die Gesamtergebnisse, indem wir die W.en für die zugehörigen Teilergebnisse multiplizieren.

4. Wie oft muss man eine verbogene Münze werfen, bis das erste Mal „Kopf“ auftritt?

Ansatz:

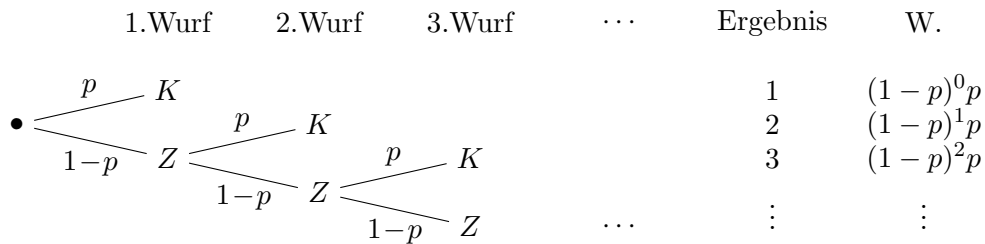
- $\Omega = \mathbb{N} = \{1, 2, 3, \dots\}$, wobei $\omega \triangleq$ Anzahl der Versuche
- $f(\omega) = (1-p)^{\omega-1}p$, $\omega \in \Omega$

Dabei ist $p \in]0, 1]$ wieder ein unbekannter Parameter.

ausführliche Begründung: über Hilfsexperiment

Um einen Ansatz für die W. zu erhalten, dass „Kopf“ erstmalig im n -ten Versuch auftritt, betrachten wir den n -maligen Münzwurf. Dieses Experiment lässt sich durch $\tilde{\Omega} := \{(\tilde{\omega}_1, \dots, \tilde{\omega}_n) : \tilde{\omega}_i \in \{K, Z\} \text{ für alle } i = 1, \dots, n\}$ und $\tilde{f}(\tilde{\omega}_1 \cdots \tilde{\omega}_n) := p_{\tilde{\omega}_1} \cdots p_{\tilde{\omega}_n}$ beschreiben, wobei p_K und p_Z wie in Beispiel 3 gewählt seien. Die W., dass „Kopf“ erstmalig im n -ten Versuch auftritt, beträgt $\tilde{f}(\underbrace{Z, \dots, Z}_{(n-1)\text{-mal}}, K) = (1-p)^{n-1}p$.

Auch diese Informationen lassen sich schön in einem (unendlich großen) W.baum darstellen:



Wir überprüfen sicherheitshalber, dass die gewählten W.en $f(\omega)$ sich zu 1 addieren:

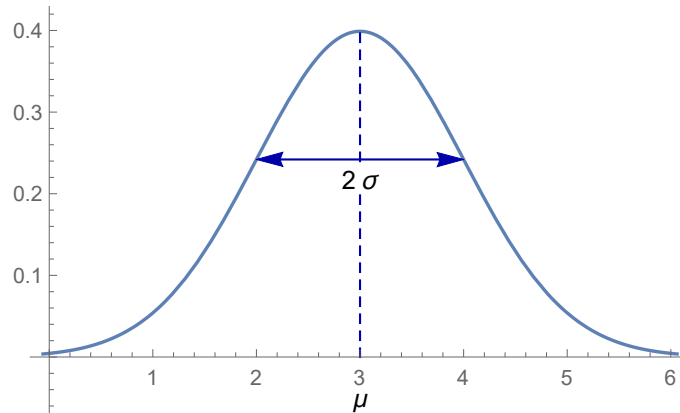
$$\sum_{\omega \in \Omega} f(\omega) = \sum_{n=1}^{\infty} p(1-p)^{n-1} = p \sum_{n=1}^{\infty} (1-p)^{n-1} \underset{\text{geom. Reihe}}{=} p \cdot \frac{1}{1-(1-p)} = 1.$$

Beachte, dass wir sind in allen Beispielen *systematisch* vorgegangen, indem wir erst eine mathematische Beschreibung der Situation angegeben und dann die W.en der uns interessierenden Ereignisse berechnet haben. Dies erscheint hier recht umständlich; bei schwierigeren Beispielen zahlt sich diese Vorgehensweise aber aus!

Bestimmung von Wahrscheinlichkeiten.

1. theoretische Argumente

- Symmetrie
- Unabhängigkeit
- Verteilungsfamilien, *Beispiel*: „Gauß’sche Glockenkurve“



2. statistische Experimente

- unbekannte Wahrscheinlichkeiten / Parameter anhand von Beobachtungen schätzen

Beispiel: Münzwurf mit einer verbogenen Münze

die Münze wird n mal geworfen, die unbekannte Wahrscheinlichkeit p für Kopf soll auf der Grundlage der dabei entstehenden n Ergebnisse geschätzt werden

$$\begin{aligned} \text{Schätzung für } p &= \text{relative Häufigkeit von Kopf} = \frac{\text{absolute Häufigkeit von Kopf}}{\text{Anzahl der Münzwürfe}} \\ \hat{p} &= h_n(\text{Kopf}) = \frac{H_n(\text{Kopf})}{n} \end{aligned}$$

(Ist das eine vernünftige Schätzung?)

3. subjektive Wahrscheinlichkeiten

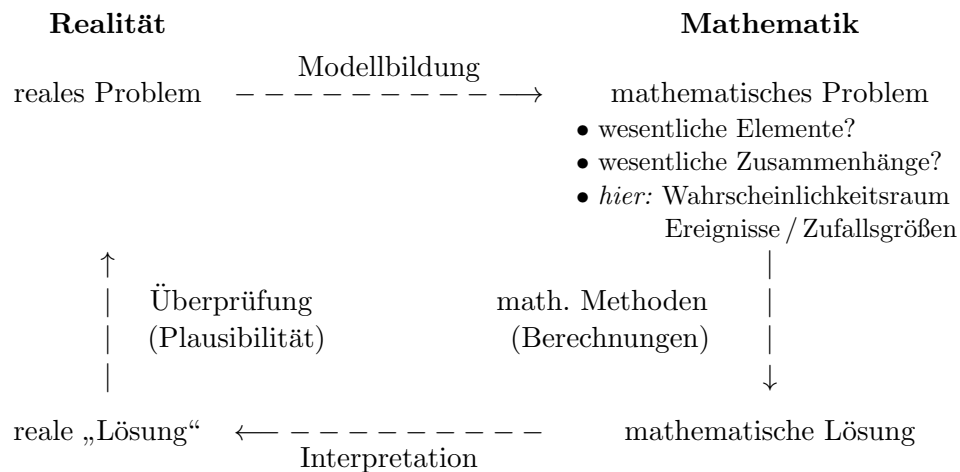
- persönliche Einschätzungen / Überzeugungen

Hier lassen sich (wie im Alltag) auch Vorgängen Wahrscheinlichkeiten zuordnen, die nur ein einziges Mal stattfinden oder die schon stattgefunden haben! (*Beispiele*: Wird Deutschland Fußball-Weltmeister? War Herr X der Mörder?)

4. Kombination der vorherigen Methoden, z. B. von 1. + 2.

- theoretische Argumente → Reduktion auf eine Verteilungsfamilie
- statistische Experimente → Schätzung der unbekannten Parameter

Modellbildungsprozess.



- Beschreibung der Wirklichkeit durch ein *mathematisches Modell*
- jedes Modell beruht auf vereinfachenden Annahmen
- die Wahl des Modells lässt sich nicht „rein mathematisch“ begründen
- die Wahl des Modells ist i. d. R. nicht eindeutig (vgl. Beispiel 2)

Zusammenfassung:

- Wir beschreiben (einfache) zufallsabhängige Vorgänge, indem wir die Menge Ω der möglichen Ergebnisse sowie die W.en $f(\omega)$ der einzelnen Ergebnisse $\omega \in \Omega$ angeben.
- Wir wollen darauf achten, die Wahl von Ω und $f(\omega)$ kurz zu begründen.
- Die W.en von Ereignissen lassen sich anschließend durch Addition der W.en der zugehörigen Ergebnisse bestimmen.
- Wir interpretieren W.en als relative Häufigkeiten bei sehr vielen Versuchswiederholungen.
- Die Wahl des Modells hängt i. d. R. auch von den getroffenen Annahmen ab und ist i. d. R. nicht eindeutig.

1 Diskrete Wahrscheinlichkeitsräume

1.1 Diskrete Wahrscheinlichkeitsräume

Mit diskreten W.räumen lassen sich einfache zufallsabhängige Vorgänge beschreiben, bei denen die Menge der möglichen Versuchsausgänge *abzählbar* (d. h. endlich oder abzählbar-unendlich) ist. Dabei heißt eine Menge Ω *abzählbar-unendlich*, wenn es eine bijektive Abbildung $\varphi : \mathbb{N} \rightarrow \Omega$ gibt, d. h. wenn sie sich in der Form $\Omega = \{\omega_1, \omega_2, \omega_3, \dots\}$ schreiben lässt, wobei jedes Element von Ω genau einmal vorkommt.

Ist Ω eine beliebige Menge, so bezeichnen wir mit $\mathfrak{P}(\Omega)$ die *Potenzmenge*, d. h. die Menge aller Teilmengen $A \subseteq \Omega$ (einschließlich der leeren Menge \emptyset und der Menge Ω selbst).

Definition 1.1 (Diskreter W.raum). Es seien $\Omega \neq \emptyset$ eine abzählbare Menge und $f : \Omega \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften $f(\omega) \geq 0$ für alle $\omega \in \Omega$ und $\sum_{\omega \in \Omega} f(\omega) = 1$. Dann heißt das Tripel $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ mit der Abbildung $\mathbb{P} : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$, $\mathbb{P}(A) := \sum_{\omega \in A} f(\omega)$ *diskreter W.raum* (zum Paar (Ω, f)).

Bemerkung 1.2 (Umgang mit Summen). Beachte, dass die Summe $\sum_{\omega \in A} f(\omega)$, die in der Definition von $\mathbb{P}(A)$ auftritt, endlich viele oder abzählbar-unendlich viele Terme umfassen kann. Eine solche Summe berechnen wir, indem wir eine *beliebige* Reihenfolge $\{\omega_1, \dots, \omega_n\}$ (für endliches A) bzw. $\{\omega_1, \omega_2, \omega_3, \dots\}$ (für abzählbar-unendliches A) der Elemente von A wählen, wobei jedes Element von A genau einmal vorkommt, und die Summe

$$\sum_{i=1}^n f(\omega_i) \quad \text{bzw.} \quad \sum_{i=1}^{\infty} f(\omega_i) := \lim_{n \rightarrow \infty} \sum_{i=1}^n f(\omega_i)$$

bilden. Dabei hängt der Wert der Summe nicht von der Wahl der Reihenfolge ab, da man Summen mit nicht-negativen Summanden beliebig umordnen kann, ohne den Wert der Summe zu verändern (vgl. Kapitel A.4). – Ähnliche Umordnungen werden wir in den folgenden Beweisen stillschweigend verwenden. Sei etwa $A \subseteq \Omega$ und $A = \bigcup_{i \in I} A_i$, wobei I eine abzählbare Indexmenge ist und jedes Element von A in genau einer Teilmenge A_i liegt. Dann gilt

$$\sum_{\omega \in A} f(\omega) \stackrel{\text{Umordnen}}{=} \sum_{i \in I} \sum_{\omega \in A_i} f(\omega).$$

Im Folgenden seien (Ω, f) und $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ wie in Definition 1.1, sofern nichts anderes angegeben ist. Die anschauliche Bedeutung der auftretenden Mengen und Abbildungen ergibt sich aus den Beispielen in Kapitel 0 und motiviert die folgenden Bezeichnungen:

Bemerkung 1.3 (Bezeichnungen).

Es seien $\omega \in \Omega$ und $A \in \mathfrak{P}(\Omega)$.

Ω Ergebnismenge / Ergebnisraum / Grundmenge / Grundraum /
Stichprobenraum

ω	Ergebnis
$f(\omega)$	W. des Ergebnisses ω
f	W.dichte / W.funktion (gibt die W.en der einzelnen Ergebnisse an)
A	Ereignis
$\mathbb{P}(A)$	W. des Ereignisses A
\mathbb{P}	W.maß / W.verteilung (gibt die W.en der einzelnen Ereignisse an)
$\{\omega\}$	Elementarereignis
$\mathbb{P}(\{\omega\}) = f(\omega)$	Elementar-W.
Ω	sicheres Ereignis
\emptyset	unmögliches Ereignis
A tritt ein.	Es erscheint ein Ergebnis $\omega \in A$.
A tritt nicht ein.	Es erscheint ein Ergebnis $\omega \notin A$.

Ein wichtiger Spezialfall ist dadurch gegeben, dass der W.raum endlich ist (d.h. dass die Menge Ω endlich ist) und dass jedes Ergebnis die gleiche Chance des Eintretens besitzt:

Bemerkung 1.4 (Gleichverteilung). Ist $\Omega \neq \emptyset$ eine endliche Menge und gilt $f(\omega) = 1/|\Omega|$ für alle $\omega \in \Omega$, so heißt die zugehörige W.verteilung \mathbb{P} (*diskrete*) *Gleichverteilung auf Ω* , kurz \mathcal{U}_Ω . In diesem Fall gilt

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} \quad \forall A \in \mathfrak{P}(\Omega),$$

d. h. die Bestimmung von W. lässt sich auf die Bestimmung von Anzahlen zurückführen.

Beachte, dass die Formel in Bemerkung 1.4 nur anwendbar ist, wenn eine Gleichverteilung vorliegt. Manchmal erweist es sich daher als nützlich, den W.raum – wie z. B. in Beispiel 2 in der Einleitung – geschickt zu wählen, so dass man mit der Gleichverteilung arbeiten kann.

In der Situation von Bemerkung 1.4 spricht man häufig auch von *Laplace-Experimenten*. Wir werden diese Bezeichnungsweise allerdings vermeiden, weil für uns *Experimente* die realen Vorgänge sind, während wir die mathematischen Modelle *W.räume* nennen.

Bemerkung 1.5.

Da Ereignisse formal Teilmengen von Ω sind, können wir mittels mengentheoretischer Operationen aus gegebenen Ereignissen A, B, C_n ($n \in \mathbb{N}$) neue Ereignisse konstruieren:

Modell	Bezeichnung	Interpretation
A^c	Komplement / Gegenereignis	A tritt nicht ein.
$A \cap B$	Durchschnitt	A und B treten ein.
$A \cup B$	Vereinigung	A oder B treten ein.

$A \Delta B$	symmetrische Differenz	Entweder A oder B tritt ein.
$A \setminus B$	mengentheoretische Differenz	A tritt ein, B tritt nicht ein.
$\bigcap_{n=1}^{\infty} C_n$	(abzählbarer) Durchschnitt	Alle Ereignisse C_n treten ein.
$\bigcup_{n=1}^{\infty} C_n$	(abzählbare) Vereinigung	Mindestens ein Ereignis C_n tritt ein.
$\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} C_m$		Unendlich viele Ereignisse C_n treten ein.
$\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} C_m$		Fast alle Ereignisse C_n treten ein.

Es gelten dann die üblichen Rechenregeln für Mengenoperationen, vgl. Anhang.

Eine Familie $(A_i)_{i \in I} \subseteq \mathfrak{P}(\Omega)$ heißt *paarweise disjunkt*, wenn für alle $i, j \in I$ mit $i \neq j$ $A_i \cap A_j = \emptyset$ gilt.

Satz 1.6 (Eigenschaften von W.maßen).

Für jeden diskreten W.raum $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ gilt:

- (1) $\forall A \in \mathfrak{P}(\Omega) : \mathbb{P}(A) \geq 0$ (Nicht-Negativität)
- (2) $\mathbb{P}(\Omega) = 1$ (Normiertheit)
- (3) $\forall A_1, \dots, A_n \in \mathfrak{P}(\Omega) : A_1, \dots, A_n$ paarweise disjunkt $\Rightarrow \mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$
(Additivität)
- (3') $\forall A_1, A_2, \dots \in \mathfrak{P}(\Omega) : A_1, A_2, \dots$ paarweise disjunkt $\Rightarrow \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$
(σ -Additivität)
- (4) $\forall A \in \mathfrak{P}(\Omega) : \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ (Regel von der Gegen-W.)
- (5) $\forall A \in \mathfrak{P}(\Omega) : \mathbb{P}(A) \leq 1$
- (6) $\mathbb{P}(\emptyset) = 0$
- (7) $\forall A, B \in \mathfrak{P}(\Omega) : A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ (Monotonie)
- (8) $\forall A, B \in \mathfrak{P}(\Omega) : A \subseteq B \Rightarrow \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ (Subtraktivität)
- (9) $\forall A_1, \dots, A_n \in \mathfrak{P}(\Omega) : \mathbb{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$ (Subadditivität)
- (9') $\forall A_1, A_2, \dots \in \mathfrak{P}(\Omega) : \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ (σ -Subadditivität)
- (10) $\forall A_1, \dots, A_n \in \mathfrak{P}(\Omega) : \mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$
(Siebformel von Sylvester-Poincaré / Einschluss-Ausschluss-Prinzip)

Bemerkung (Spezialfälle von (10)).

$$n = 2: \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$$

$$n = 3: \mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) \\ - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3)$$

Beweis von Satz 1.6.

(1) – (3') müssen mit der Definition von \mathbb{P} gezeigt werden.

(1) klar, da Summen nicht-negativer Zahlen nicht-negativ sind.

(2) klar, da $\sum_{\omega \in \Omega} f(\omega) = 1$ vorausgesetzt ist.

(3) Dies folgt daraus, dass sich Summen nicht-negativer Zahlen beliebig umordnen lassen, ohne den Wert der Summe zu verändern:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \stackrel{\text{Def.}}{=} \sum_{\omega \in \bigcup_{i=1}^n A_i} f(\omega) \stackrel{\text{Umordnen}}{=} \sum_{i=1}^n \sum_{\omega \in A_i} f(\omega) \stackrel{\text{Def.}}{=} \sum_{i=1}^n \mathbb{P}(A_i).$$

Für die mittlere Gleichung beachte, dass jedes $\omega \in \bigcup_{i=1}^n A_i$ auf jeder Seite genau einmal vorkommt; dafür ist natürlich die Voraussetzung wesentlich, dass die Ereignisse A_i paarweise disjunkt sind.

(3') Dies zeigt man analog zu (3).

(4) – (9) können (mittels geeigneter *Zerlegungen*) auf (1) – (3') zurückgeführt werden.

(4) Es gilt

$$\mathbb{P}(A) + \mathbb{P}(A^c) \stackrel{(3)}{=} \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) \stackrel{(2)}{=} 1,$$

woraus die Behauptung durch Umstellen folgt.

(5) folgt aus (1) und (4).

(6) folgt aus (2) und (4).

(7) Es gilt

$$\mathbb{P}(A) \stackrel{(1)}{\leq} \mathbb{P}(A) + \mathbb{P}(B \setminus A) \stackrel{(3)}{=} \mathbb{P}(A \cup (B \setminus A)) = \mathbb{P}(B),$$

wobei im letzten Schritt die Voraussetzung $A \subseteq B$ eingegangen ist.

(8) Wie schon im Beweis von (7) gesehen, gilt

$$\mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(B),$$

daraus folgt die Behauptung durch Umstellen.

(9) Wir verwenden, dass sich allgemeine Vereinigungen auf disjunkte Vereinigungen zurückführen lassen. Sei dazu $B_i := A_i \setminus (A_1 \cup \dots \cup A_{i-1})$, $i = 1, \dots, n$. Dann gilt $\bigcup_{i=1}^n A_i = \bigcup_{i=1}^n B_i$, B_1, \dots, B_n paarweise disjunkt sowie $B_i \subseteq A_i$, $i = 1, \dots, n$, und damit

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^n B_i\right) \stackrel{(3)}{=} \sum_{i=1}^n \mathbb{P}(B_i) \stackrel{(7)}{\leq} \sum_{i=1}^n \mathbb{P}(A_i).$$

(9') Dies zeigt man analog zu (9), wobei man im mittleren Schritt (3') statt (3) benutzt.

- (10) Dies zeigt man durch vollständige Induktion nach n .¹ Der Fall $n = 1$ ist klar, und der Fall $n = 2$ ergibt sich durch Kombination der Gleichungen

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1 \cup (A_2 \setminus A_1)) \stackrel{(3)}{=} \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1)$$

und

$$\mathbb{P}(A_2) = \mathbb{P}((A_1 \cap A_2) \cup (A_2 \setminus A_1)) \stackrel{(3)}{=} \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_2 \setminus A_1).$$

Für den Induktionsschritt verwenden wir, dass sich die Behauptung mit der Abkürzung $A_I := \bigcap_{i \in I} A_i$ auch in der Form

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{\emptyset \subsetneq I \subseteq \{1, \dots, n\}} (-1)^{|I|-1} \mathbb{P}(A_I)$$

darstellen lässt. Ist diese Behauptung für jede n -fache Vereinigung richtig, wobei $n \geq 2$, so folgt für jede $(n+1)$ -fache Vereinigung

$$\begin{aligned} & \mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) \\ = & \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cup A_{n+1}\right) \\ \stackrel{\text{Fall } n=2}{=} & \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^n A_i\right) \cap A_{n+1}\right) \\ \stackrel{\text{Distr.gesetz für Mengen}}{=} & \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^n \underbrace{(A_i \cap A_{n+1})}_{=: B_i}\right) \\ \stackrel{\text{I.V.}}{=} & \sum_{\emptyset \subsetneq I \subseteq \{1, \dots, n\}} (-1)^{|I|-1} \mathbb{P}(A_I) + \mathbb{P}(A_{n+1}) - \sum_{\emptyset \subsetneq I \subseteq \{1, \dots, n\}} (-1)^{|I|-1} \mathbb{P}\left(\underbrace{A_I \cap A_{n+1}}_{=: B_I}\right) \\ \stackrel{\text{Indexvers. } J=I \cup \{n+1\}}{=} & \sum_{\emptyset \subsetneq I \subseteq \{1, \dots, n\}} (-1)^{|I|-1} \mathbb{P}(A_I) + \mathbb{P}(A_{n+1}) + \sum_{\{n+1\} \subsetneq J \subseteq \{1, \dots, n+1\}} (-1)^{|J|-1} \mathbb{P}(A_J) \\ = & \sum_{\emptyset \subsetneq I \subseteq \{1, \dots, n+1\}} (-1)^{|I|-1} \mathbb{P}(A_I), \end{aligned}$$

d. h. die Behauptung ist auch für jede $(n+1)$ -fache Vereinigung richtig.

□

¹ Dieser Beweis ist in der Vorlesung übersprungen worden.

Beispiel 1.7 (Würfelwurf). Ein ‘fairer’ Würfel wird 4-mal geworfen.

- (i) Mit welcher W. fällt mindestens einmal die Sechs?
- (ii) Mit welcher W. fällt genau einmal die Sechs?
- (iii) Mit welcher W. ist das größte Ergebnis eine Vier?

Wir wählen $\Omega = \{(\omega_1, \omega_2, \omega_3, \omega_4) : \omega_i \in \{1, \dots, 6\} \text{ für } i = 1, 2, 3, 4\}$ (wobei $\omega_i \triangleq$ Augenzahl im i -ten Wurf) und $f(\omega) := 1/|\Omega|$ für alle $\omega \in \Omega$ oder $\mathbb{P} :=$ Gleichverteilung auf Ω (aus Symmetriegründen).

- (i) • A : „mindestens einmal die Sechs“, $A = \{\omega \in \Omega : \exists i \in \{1, 2, 3, 4\} : \omega_i = 6\}$
 - Idee: Übergang zum Komplement
 - A^c : „nie die Sechs“, $A^c = \{\omega \in \Omega : \forall i \in \{1, 2, 3, 4\} : \omega_i \neq 6\}$
 - $\mathbb{P}(A^c) = \frac{|A^c|}{|\Omega|} = \frac{5^4}{6^4} = \frac{625}{1296}$
 - $\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{625}{1296} = \frac{671}{1296} [\approx 0.518]$
- (ii) • B : „genau einmal die Sechs“, $B = \{\omega \in \Omega : \exists! i \in \{1, 2, 3, 4\} : \omega_i = 6\}$
 - Idee: geschickte Zerlegung
 - B_i : „die Sechs genau im i -ten Wurf“, $B_i = \{\omega \in \Omega : (\omega_i = 6) \wedge (\forall j \neq i : \omega_j \neq 6)\}$
 - $\mathbb{P}(B_i) = \frac{|B_i|}{|\Omega|} = \frac{5^3}{6^4} = \frac{125}{1296}$
 - $\mathbb{P}(B) = \mathbb{P}(B_1 \cup \dots \cup B_4) \underset{\text{Additivität}}{=} \mathbb{P}(B_1) + \dots + \mathbb{P}(B_4) = 4 \cdot \frac{125}{1296} = \frac{500}{1296} [\approx 0.386]$
- (iii) • C : „größte Augenzahl = 4“, $C = \{\omega \in \Omega : \max_{i=1, \dots, 4} \omega_i = 4\}$
 - Idee: geschickte Darstellung
 - C_i : „größte Augenzahl $\leq i$ “, $C_i = \{\omega \in \Omega : (\forall j : \omega_j \leq i)\}$
 - $\mathbb{P}(C_i) = \frac{|C_i|}{|\Omega|} = \frac{i^4}{6^4}$
 - $\mathbb{P}(C) = \mathbb{P}(C_4 \setminus C_3) \underset{\text{Subtraktivität}}{=} \mathbb{P}(C_4) - \mathbb{P}(C_3) = \frac{256}{1296} - \frac{81}{1296} = \frac{175}{1296} [\approx 0.135]$

Eine Folge $(A_n)_{n \in \mathbb{N}} \subseteq \mathfrak{P}(\Omega)$ heißt *aufsteigend* bzw. *absteigend*, falls für alle $n \in \mathbb{N}$ $A_n \subseteq A_{n+1}$ bzw. $A_n \supseteq A_{n+1}$ gilt.

Lemma 1.8 (Charakterisierung der σ -Additivität). Es sei $\Omega \neq \emptyset$ eine nicht-leere Menge und $\mathbb{P} : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$ eine Abbildung mit den Eigenschaften (1) – (3) in Satz 1.6. Dann sind äquivalent:

- (a) (σ -Additivität) Für jede Folge $(A_n)_{n \in \mathbb{N}}$ von paarweise disjunkten Mengen in $\mathfrak{P}(\Omega)$ gilt $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.
- (b) (Stetigkeit von unten) Für jede aufsteigende Folge $(A_n)_{n \in \mathbb{N}}$ in $\mathfrak{P}(\Omega)$ gilt $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcup_{n=1}^{\infty} A_n)$.
- (c) (Stetigkeit von oben) Für jede absteigende Folge $(A_n)_{n \in \mathbb{N}}$ in $\mathfrak{P}(\Omega)$ gilt $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_{n=1}^{\infty} A_n)$.

Ist $(A_n)_{n \in \mathbb{N}} \subseteq \mathfrak{P}(\Omega)$ eine aufsteigende bzw. absteigende Folge und $A = \bigcup_{n=1}^{\infty} A_n$ bzw. $A = \bigcap_{n=1}^{\infty} A_n$, so schreibt man auch $A_n \uparrow A$ bzw. $A_n \downarrow A$. Mit diesen Bezeichnungen kann man die Eigenschaften (b) und (c) auch in der suggestiven Form

$$\left[A_n \uparrow A \Rightarrow \mathbb{P}(A_n) \uparrow \mathbb{P}(A) \right] \quad \text{bzw.} \quad \left[A_n \downarrow A \Rightarrow \mathbb{P}(A_n) \downarrow \mathbb{P}(A) \right]$$

formulieren.

Praktisch bedeuten die Eigenschaften (b) und (c), dass wir „kompliziertere“ Ereignisse durch „einfachere“ Ereignisse von innen bzw. von außen approximieren können und dann die zugehörigen W.en durch Grenzwertbildung erhalten können. Beachte die Analogie zur Volumen-Bestimmung in der Geometrie!

*Beweis von Lemma 1.8.*² Gilt (a) und ist $(A_n)_{n \in \mathbb{N}}$ eine aufsteigende Folge in $\mathfrak{P}(\Omega)$, so ist die Folge $(B_n)_{n \in \mathbb{N}}$ mit $B_n := A_n \setminus (A_1 \cup \dots \cup A_{n-1})$, $n \in \mathbb{N}$, eine Folge paarweise disjunkter Mengen in $\mathfrak{P}(\Omega)$, und es folgt

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) \stackrel{(a)}{=} \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(B_i) \stackrel{(3)}{=} \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n B_i\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n),$$

also (b). Gilt umgekehrt (b) und ist $(A_n)_{n \in \mathbb{N}}$ eine Folge paarweise disjunkter Mengen in $\mathfrak{P}(\Omega)$, so ist die Folge $(B_n)_{n \in \mathbb{N}}$ mit $B_n := A_1 \cup \dots \cup A_n$, $n \in \mathbb{N}$, eine aufsteigende Folge in $\mathfrak{P}(\Omega)$, und es folgt

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) \stackrel{(b)}{=} \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \stackrel{(3)}{=} \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i),$$

also (a). Die Äquivalenz von (b) und (c) ergibt sich durch Übergang zum Komplement, d. h. mit Regel (4) in Satz 1.6. \square

Kombinatorik.

Motivation. Liegt die Gleichverteilung auf Ω ($\Omega \neq \emptyset$ endlich) vor, so müssen wir zur Berechnung von $\mathbb{P}(A)$ die Anzahl der Elemente von Ω sowie von A bestimmen. Manchmal lässt sich dieses Problem auf die folgende Problemstellung zurückführen:

Aus einer Menge mit n Elementen (o. E. $\{1, \dots, n\}$) wird k -mal ein Element ausgewählt. Wie viele Möglichkeiten gibt es?

² Dieser Beweis ist in der Vorlesung übersprungen worden.

Satz 1.9. Die Anzahl der Möglichkeiten, aus einer Menge mit n Elementen k -mal ein Element auszuwählen, ist durch die folgende Tabelle gegeben:

	mit Wiederholung	ohne Wiederholung
geordnete Stichproben	n^k (Fall I)	$\frac{n!}{(n-k)!}$ (Fall II)
ungeordnete Stichproben	$\binom{n+k-1}{k}$ (Fall IV)	$\binom{n}{k}$ (Fall III)

Dabei kommt es bei „[un]geordneten Stichproben“ auf die Reihenfolge [nicht] an, und im Fall „ohne Wiederholung“ sei stets $k \leq n$ vorausgesetzt.

Statt von *geordneten Stichproben* / *ungeordneten Stichproben* sprechen wir auch von *Variationen* / *Kombinationen*.

Formale Beschreibung der Mengen der möglichen Stichproben.

Fall I (Geordnete Stichproben mit Wiederholung)

$$\Omega_I = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\} \text{ für } i = 1, \dots, k\} =: \{1, \dots, n\}^k$$

($\omega_i \triangleq$ Auswahl im i ten Schritt)

$$|\Omega_I| = n \cdot n \cdot \dots \cdot n = n^k$$

Fall II (Geordnete Stichproben ohne Wiederholung)

$$\Omega_{II} = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\} \text{ für } i = 1, \dots, k \text{ und } \omega_i \neq \omega_j \text{ für } 1 \leq i < j \leq k\}$$

($\omega_i \triangleq$ Auswahl im i ten Schritt)

$$|\Omega_{II}| = n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

Fall III (Ungeordnete Stichproben ohne Wiederholung)

$$\Omega_{III} = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\} \text{ für } i = 1, \dots, k \text{ und } \omega_1 < \dots < \omega_k\}$$

($\omega_i \triangleq$ das i te-kleinste ausgewählte Element)

$$|\Omega_{III}| = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Alternative:

$$\Omega'_{III} = \{\{\omega_1, \dots, \omega_k\} : \omega_i \in \{1, \dots, n\} \text{ für } i = 1, \dots, k \text{ und } \omega_i \neq \omega_j \text{ für } 1 \leq i < j \leq k\}$$

($\{\omega_1, \dots, \omega_k\} \triangleq$ Menge der ausgewählten Elemente)

Fall IV (Ungeordnete Stichproben mit Wiederholung)

$$\Omega_{IV} = \{(\omega_1, \dots, \omega_k) : \omega_i \in \{1, \dots, n\} \text{ für } i = 1, \dots, k \text{ und } \omega_1 \leq \dots \leq \omega_k\}$$

($\omega_i \triangleq$ das i te-kleinste ausgewählte Element)

$$|\Omega_{IV}| = \binom{n+k-1}{k}$$

Dabei bezeichnet $n! := 1 \cdot \dots \cdot n$ („ n Fakultät“) das Produkt der natürlichen Zahlen von 1 bis n .

Beachte, dass wir bei ungeordneten Stichproben die ausgewählten Elemente sortieren können, da die Reihenfolge unwesentlich ist, und damit die Stichprobe durch ein aufsteigend geordnetes Tupel beschreiben können!

*Beweisskizze zu Satz 1.9.*³

Wir verwenden die folgende *Grundregel der Kombinatorik*:

Gibt es bei einem k -stufigen Auswahlverfahren

- im 1. Schritt n_1 Auswahlmöglichkeiten,
 - im 2. Schritt n_2 Auswahlmöglichkeiten,
 - \vdots
 - im k . Schritt n_k Auswahlmöglichkeiten,
- so gibt es insgesamt $n_1 \cdot n_2 \cdot \dots \cdot n_k$ Auswahlmöglichkeiten.

Fall I: nach Grundregel der Kombinatorik

Fall II: nach Grundregel der Kombinatorik

Fall III: Jede geordnete Stichprobe lässt sich auf eindeutige Weise dadurch erzeugen, dass man als Erstes die Elemente (also die zugrunde liegende ungeordnete Stichprobe) und als Zweites die Reihenfolge (also die Anordnung) bestimmt. Nach der Grundregel der Kombinatorik und Bem. 1.10 gilt daher $|\Omega_{\text{II}}| = |\Omega_{\text{III}}| \cdot k!$, also $|\Omega_{\text{III}}| = |\Omega_{\text{II}}| / k!$. Durch Einsetzen des Ergebnisses für Fall II folgt die Behauptung.

Fall IV: Dies ergibt sich nun aus der Beobachtung, dass die Abbildung $f : \mathbb{N}^k \rightarrow \mathbb{N}^k$ mit $f(\omega_1, \omega_2, \dots, \omega_k) := (\omega_1 + 0, \omega_2 + 1, \dots, \omega_k + k - 1)$ eine Bijektion von Ω_{IV} mit den Parametern n und k auf Ω_{III} mit den Parametern $n + k - 1$ und k vermittelt. Dabei haben wir ein weiteres Grundprinzip der Kombinatorik benutzt, nämlich dass aus der Existenz einer Bijektion $g : A \rightarrow B$ zwischen zwei endlichen Mengen A und B schon $|A| = |B|$ folgt. \square

Bemerkung 1.10. In Fall II ist der Fall $n = k$ von besonderem Interesse; hier ist Ω_{II} die Menge der Permutationen (Umordnungen) der n Elemente, und es gilt $|\Omega_{\text{II}}| = n!$.

Bemerkung 1.11. Um sich einen Eindruck von der Größe von $n!$ (für großes n) bzw. von $\binom{n}{k}$ (für großes k und $n - k$) zu verschaffen, kann man die *Stirling-Formel* verwenden:

$$n! \sim \sqrt{2\pi n} (n/e)^n \quad (n \rightarrow \infty) \quad \left[\iff \lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} (n/e)^n} = 1 \right].$$

Bemerkung 1.12. Die Problemstellung in Satz 1.9 lässt sich unterschiedlich interpretieren:

1. Urnenmodell

Aus einer Urne mit n Kugeln (mit den Nummern $1 - n$) wird k -mal gezogen.

- mit / ohne Zurücklegen \triangleq mit / ohne Wiederholung
- mit / ohne Berücksichtigung der Reihenfolge \triangleq geordnete / ungeordnete Stichproben

2. Teilchen-Fächer-Modell

k Teilchen werden auf n Fächer (mit den Nummern $1 - n$) verteilt.

- mit / ohne Mehrfachbelegung \triangleq mit / ohne Wiederholung
- unterscheidbare / nicht-unterscheidbare Teilchen \triangleq geordnete / ungeordnete Stichproben

³ Dieser Beweis ist in der Vorlesung übersprungen worden, weil er in der Vorlesung „Diskrete Strukturen und Optimierung“ behandelt wird.

Zusammenhang: Wähle für jedes Teilchen das Fach, in welches das Teilchen gelegt wird.

Beispiele 1.13.

- (a) (*Wörter*) Aus einem Alphabet mit n „Buchstaben“ werden „Wörter“ der Länge n gebildet, wobei Buchstaben mehrfach verwendet werden dürfen. Wie groß ist die W., dass ein rein zufälliges Wort keinen Buchstaben doppelt enthält?
- (b) (*Murmeln*) 6 nicht-unterscheidbare Murmeln werden auf 3 unterscheidbare Dosen verteilt. Wie viele Möglichkeiten gibt es?
- (c) (*Murmeln*) 6 nicht-unterscheidbare Murmeln werden auf 3 nicht-unterscheidbare Dosen verteilt. Wie viele Möglichkeiten gibt es? (*!!! Achtung Falle !!!*)
- (d) (*Wörter*) Aus den Buchstaben des Wortes ANANAS werden „Wörter“ der Länge 6 gebildet, wobei jeder Buchstabe genau so oft verwendet werden muss, wie er im Wort ANANAS vorkommt. Wie viele Möglichkeiten gibt es?
Wir wählen nacheinander die Positionen für die A's, N's und S's. $\rightsquigarrow \binom{6}{3} \binom{3}{2} \binom{1}{1} = 60$.
- (e) (*Lotto*) Sie geben beim Lotto „6 aus 49“ (mit Superzahl) einen Tipp ab. Wie groß ist die W. für die Gewinnklassen I, II, III, ...?
- (f) (*Münzwurf*) Es werden n faire Münzen gleichzeitig geworfen. Mit welcher W. erhalten wir (genau) k -mal Kopf, $k = 0, \dots, n$?

1.2 Zufallsgrößen auf diskreten W.räumen

Motivation. In Beispiel 1.13(f) gilt für die Mengen A_k : „(genau) k -mal Kopf“ $\mathbb{P}(A_k) \geq 0$ für alle $k = 0, \dots, n$ sowie $\sum_{k=0}^n \mathbb{P}(A_k) = \mathbb{P}(\bigcup_{k=0}^n A_k) = \mathbb{P}(\Omega) = 1$, da die Mengen A_0, \dots, A_n paarweise disjunkt mit $\bigcup_{k=0}^n A_k = \Omega$ sind. Interessieren wir uns ausschließlich für die Anzahl von Kopf, so können wir daher zu einem neuen diskreten W.raum mit der Grundmenge $\tilde{\Omega} := \{0, \dots, n\}$ und den Elementar-W.'en $\tilde{f}(k) := \mathbb{P}(A_k) = \binom{n}{k}/2^n$, $k = 0, \dots, n$ übergehen.

Diese Situation, dass wir uns in einem gegebenen „umfangreicheren“ W.raum nur für einen speziellen Aspekt interessieren und zu einem „einfacheren“ W.raum übergehen, kommt in der Stochastik sehr häufig vor. Wir wollen dies nun allgemein formulieren, wobei wir den relevanten Aspekt mit Hilfe einer Abbildung beschreiben. Ist dabei $X : \Omega \rightarrow \mathcal{X}$ eine Abbildung und ist $B \subseteq \mathcal{X}$ beliebig, so bezeichnet

$$\{X \in B\} := X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$$

das Urbild von B unter X . Dieses Urbild beschreibt gerade das Ereignis, dass die Zufallsgröße X einen Wert in der Menge B annimmt. Im Spezialfall $B = \{x\}$ schreibt man einfach $\{X = x\}$ statt $\{X \in \{x\}\}$. Zudem lässt man die Mengenklammern häufig weg, wenn diese Ereignisse als Argumente von W.maßen auftreten.

Definition 1.14 (Zufallsgröße). Es seien $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ ein diskreter W.raum und $X : \Omega \rightarrow \mathcal{X}$ eine Abbildung mit Werten in einer abzählbaren Menge \mathcal{X} . Dann heißt X *Zufallsgröße* oder *Zufallsvariable* (mit Werten in \mathcal{X}), und die Abbildung $\mathbb{P}_X : \mathfrak{P}(\mathcal{X}) \rightarrow \mathbb{R}$ mit $\mathbb{P}_X(B) := \mathbb{P}(X \in B)$ für alle $B \in \mathfrak{P}(\mathcal{X})$ heißt *induzierte Verteilung* (oder auch nur *Verteilung*) von X unter \mathbb{P} .

Interpretation: Zufallsgrößen „verarbeiten“ / „verdichten“ / „filtern“ die Information über den Ausgang eines Zufallsexperimentes.

Beispiele 1.15 (Beispiele für Zufallsgrößen).

- (a) In Beispiel 1.13 (f) beschreibt die Zufallsgröße $X : \Omega \rightarrow \mathcal{X} := \{0, \dots, n\}$ mit $X(\omega) := \sum_{i=1}^n \omega_i$ die Anzahl von Kopf, und die induzierte Verteilung ist dann wegen

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(\bigcup_{k \in B} \{X = k\}\right) = \sum_{k \in B} \binom{n}{k} / 2^n \quad \forall B \subseteq \mathcal{X}$$

gerade die diskrete W.verteilung zum Paar (\mathcal{X}, f_X) mit $\mathcal{X} := \{0, \dots, n\}$ und $f_X(k) := \binom{n}{k} / 2^n$, $k = 0, \dots, n$.

- (b) (vgl. Übungen) Ist (Ω, f) mit $\Omega := \{1, \dots, 6\}^2$ ($\omega_1, \omega_2 \triangleq$ gewürfelte Augenzahlen) und $f(\omega) := \frac{1}{36} \forall \omega \in \Omega$ (wegen Symmetrie) das Modell für den gleichzeitigen Wurf zweier fairer Würfel, so beschreibt die Zufallsgröße $S : \Omega \rightarrow \mathcal{S} := \{2, \dots, 12\}$ mit $S(\omega_1, \omega_2) := \omega_1 + \omega_2$ die Augensumme, und die induzierte Verteilung ist dann wegen

$$\mathbb{P}_S(B) = \mathbb{P}(S \in B) = \mathbb{P}\left(\bigcup_{k \in B} \{S = k\}\right) = \sum_{k \in B} \frac{6 - |k-7|}{36} \quad \forall B \subseteq \mathcal{S}$$

gerade die diskrete W.verteilung zum Paar (\mathcal{S}, f_S) mit $\mathcal{S} := \{2, \dots, 12\}$ und $f_S(k) := \frac{6 - |k-7|}{36}$, $k = 2, \dots, 12$.

Das folgende Lemma zeigt, dass die Bezeichnung *induzierte Verteilung* gerechtfertigt ist:

Lemma 1.16. In der Situation von Definition 1.14 ist \mathbb{P}_X eine diskrete W.verteilung auf \mathcal{X} , genauer die diskrete W.verteilung auf \mathcal{X} mit der W.dichte $f_X(x) := \mathbb{P}(X = x)$, $x \in \mathcal{X}$.

Beweis. Für alle $B \in \mathfrak{P}(\mathcal{X})$ gilt

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}\left(\bigcup_{x \in B} \{X = x\}\right) \stackrel{\text{disjunkte Vereinigung}}{=} \sum_{x \in B} \mathbb{P}(X = x) = \sum_{x \in B} f_X(x).$$

Damit ist \mathbb{P}_X das diskrete W.maß zur W.dichte f_X . (Beachte dazu: Wegen $f_X(x) = \mathbb{P}(X = x) \geq 0$ für alle $x \in \mathcal{X}$ sowie $\sum_{x \in \mathcal{X}} f_X(x) = \mathbb{P}(X \in \mathcal{X}) = 1$ (in der Rechnung $B := \mathcal{X}$ einsetzen!) ist f_X in der Tat eine W.dichte.) \square

Merkregel:

Die Verteilung einer Zufallsgröße $X : \Omega \rightarrow \mathcal{X}$ auf einem diskreten W.raum bestimmen wir, indem wir die zugehörigen Elementar-W.en $\mathbb{P}(X = x)$, $x \in \mathcal{X}$, berechnen.

Lemma 1.17. Sind $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ ein diskreter W.raum, $X : \Omega \rightarrow \mathcal{X}$ eine Zufallsgröße und $h : \mathcal{X} \rightarrow \mathcal{Y}$ eine Abbildung (wobei \mathcal{X} und \mathcal{Y} abzählbar), so ist auch $h(X) := h \circ X : \Omega \rightarrow \mathcal{Y}$ eine Zufallsgröße, und es gilt $\mathbb{P}_{h(X)} = (\mathbb{P}_X)_h$.

Beweis. Für alle $C \in \mathfrak{P}(\mathcal{Y})$ gilt

$$(h \circ X)^{-1}(C) = \{\omega \in \Omega : h(X(\omega)) \in C\} = \{\omega \in \Omega : X(\omega) \in h^{-1}(C)\} = X^{-1}(h^{-1}(C))$$

und damit

$$\mathbb{P}_{h(X)}(C) = \mathbb{P}((h \circ X)^{-1}(C)) = \mathbb{P}(X^{-1}(h^{-1}(C))) = \mathbb{P}_X(h^{-1}(C)) = (\mathbb{P}_X)_h(C).$$

□

Beispiel 1.18 (Bestimmung von Verteilungen von Zufallsgrößen).

Bei einem Glücksspiel werden zwei faire Würfel gleichzeitig geworfen. Es interessiert allein die Summe S der geworfenen Augenzahlen. Der Gewinn G bestimmt sich nach der folgenden Regel:

$$\begin{aligned} S = 12 & \Rightarrow G = 12 \\ S = 9, 10, 11 & \Rightarrow G = 2 \\ S < 9 & \Rightarrow G = 0 \end{aligned}$$

Gesucht ist die Verteilung von G .

Seien dazu (Ω, f) und $S : \Omega \rightarrow \mathcal{S}$ wie in Beispiel 1.15. Dann wird der Gewinn durch die Zufallsgröße $G = h \circ S : \Omega \rightarrow \{0, 2, 12\}$ mit

$$h : \mathcal{S} \rightarrow \{0, 2, 12\}, \quad h(s) := \begin{cases} 12, & s = 12 \\ 2, & 9 \leq s \leq 11 \\ 0, & s < 9 \end{cases}$$

beschrieben. Die Verteilung von G ergibt sich aus der schon bekannten Verteilung von S :

x	2	3	4	5	6	7	8	9	10	11	12
$f_S(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

y	0	2	12
$f_G(y)$	$\frac{26}{36}$	$\frac{9}{36}$	$\frac{1}{36}$

Beispielsweise gilt $f_G(2) = \mathbb{P}(G = 2) = \mathbb{P}(S \in \{9, 10, 11\}) = f_S(9) + f_S(10) + f_S(11) = \frac{9}{36} = \frac{1}{4}$. □

Oft ist man an mehreren Zufallsgrößen gleichzeitig interessiert. Seien etwa $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ ein diskreter W.raum und X und Y Zufallsgrößen mit Werten in abzählbaren Mengen \mathcal{X} bzw. \mathcal{Y} . Dann kann man die „zusammengesetzte“ Zufallsgröße (X, Y) mit Werten in der abzählbaren Menge $\mathcal{X} \times \mathcal{Y}$ betrachten. Also ist $\mathbb{P}_{(X, Y)}$ eine Verteilung auf $\mathcal{X} \times \mathcal{Y}$, und die Elementar-W.en sind von der Form $\mathbb{P}((X, Y) = (x, y)) = \mathbb{P}(X = x, Y = y)$, wobei $x \in \mathcal{X}$, $y \in \mathcal{Y}$ und $\{X = x, Y = y\} := \{X = x\} \cap \{Y = y\}$.

Wenn die Verteilung von (X, Y) bereits bekannt ist, lassen sich die Verteilungen von X und von Y (die sog. *Randverteilungen*) wie folgt bestimmen:

Lemma 1.19 (Bestimmung von Randverteilungen). *Sind $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ ein diskreter W.raum und $(X, Y) : \Omega \rightarrow \mathcal{X} \times \mathcal{Y}$ eine zusammengesetzte Zufallsgröße (wobei \mathcal{X}, \mathcal{Y} abzählbar), so gilt*

$$\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y)$$

für alle $x \in \mathcal{X}$.

Beweis.

$$\mathbb{P}(X = x) = \mathbb{P}\left(\bigcup_{y \in \mathcal{Y}} \{X = x, Y = y\}\right) \stackrel[\text{Vereinigung}]{\text{disjunkte}} = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y).$$

□

Eine analoge Aussage gilt natürlich für Y statt X .

Allgemeiner kann man die Situation betrachten, dass die W.dichte einer Zufallsgröße (X_1, \dots, X_n) mit Werten in einer abzählbaren Menge $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$ gegeben ist und eine *Randverteilung* \mathbb{P}_{X_i} ($i = 1, \dots, n$) bzw. sogar eine *mehrdimensionale Randverteilung* $\mathbb{P}_{(X_{i_1}, \dots, X_{i_m})}$ ($1 \leq i_1 < \dots < i_m \leq n$) gesucht ist. Hier gilt:

Merkregel:

Ist die W.dichte einer zusammengesetzten Zufallsgröße (X_1, \dots, X_n) mit Werten in einer abzählbaren Menge $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$ gegeben, so erhalten wir die W.dichten der Randverteilungen, indem wir über die „freien“ Variablen aufsummieren.

Beispiel 1.20 (Bestimmung von Randverteilungen). Zwei faire Münzen, auf deren Seiten die Zahlen 0 und 1 stehen, werden gleichzeitig geworfen; es interessieren uns die Summe X und das Produkt Y der Einzelergebnisse.

Wenn die gemeinsame Verteilung von X und Y bereits in Form der folgenden Tabelle gegeben ist (*Übung!!!*), erhält man die Randverteilungen von X bzw. Y , indem man die Zeilensummen bzw. die Spaltensummen bildet:

$\mathbb{P}(X = x, Y = y)$	$y = 0$	$y = 1$	$\mathbb{P}(X = x)$
$x = 0$	1/4	0	1/4
$x = 1$	1/2	0	1/2
$x = 2$	0	1/4	1/4
$\mathbb{P}(Y = y)$	3/4	1/4	1

(Die Eins unten rechts dient nur der Kontrolle, da die W.en $\mathbb{P}(X = x)$ bzw. $\mathbb{P}(Y = y)$ sich jeweils zu Eins addieren müssen, und kann weggelassen werden.) □

Bemerkung 1.21. In vielen Situationen – etwa wenn man W.'en der Form $\mathbb{P}_X(B)$ bestimmen will / muss – ist es ausreichend, den „induzierten“ W.raum $(\mathcal{X}, \mathfrak{P}(\mathcal{X}), \mathbb{P}_X)$ zu kennen. Oft verzichtet man dann darauf, den zugrunde liegenden W.raum $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ anzugeben, und gibt nur die Verteilung von X an: „Sei X eine Zufallsgröße mit der Verteilung $P \dots$ “. Dies bedeutet, dass man mit *irgendeinem* W.raum $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ und *irgendeiner* Zufallsgröße X arbeiten kann, so dass die Verteilung von X mit P übereinstimmt. Ist P ein diskretes W.maß mit der Grundmenge \mathcal{X} , so kann man z.B. immer $(\Omega, \mathfrak{P}(\Omega), \mathbb{P}) := (\mathcal{X}, \mathfrak{P}(\mathcal{X}), P)$ und $X = \text{id}_\Omega$ (identische Abbildung) wählen.

Insbesondere gibt es zwei Arten, wie Zufallsgrößen typischerweise auftreten:

1. Es sind der zugrunde liegende W.raum $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ und die Zufallsgröße X gegeben, und die (teilweise) Bestimmung der Verteilung von X ist i. d. R. das erste Problem, mit dem man sich befassen muss.
2. Es sind „nur“ eine nicht näher spezifizierte Zufallsgröße X sowie ihre Verteilung P (oder eine anschauliche Beschreibung, aus der sich die Verteilung P direkt ergibt) angegeben, und man muss / kann ausgehend von P „weiterrechnen“.

Beispiele für diskrete Wahrscheinlichkeitsverteilungen.

Wir wollen weitere Beispiele für Zufallsgrößen betrachten und in diesem Zusammenhang eine Reihe wichtiger W.verteilungen kennen lernen. Dabei bezeichnen wir die Grundmenge und die W.dichte der induzierten W.verteilung oft mit $\tilde{\Omega}$ und \tilde{f} .

Bernoulli-Verteilung.

Es wird ein (u. U. komplexes) Zufallsexperiment durchgeführt, und es interessiert allein die Frage, ob dabei ein bestimmtes Ereignis $A \in \mathfrak{P}(\Omega)$ eintritt oder nicht. Ist der W.raum $(\Omega, \mathfrak{P}(\Omega), \mathbb{P})$ gegeben, so lässt sich die Situation durch die Zufallsgröße

$$\mathbf{1}_A : \Omega \rightarrow \{0, 1\} \quad \mathbf{1}_A(\omega) := \begin{cases} 1 & ; \omega \in A \\ 0 & ; \omega \notin A \end{cases}$$

(Indikatorfunktion von A) beschreiben; ihre Verteilung ist durch $\tilde{\Omega} := \{0, 1\}$ und $\tilde{f}(1) = \mathbb{P}(A)$, $\tilde{f}(0) = 1 - \mathbb{P}(A)$ gegeben.

Definition 1.22 (Bernoulli-Verteilung). Sei $p \in [0, 1]$. Die diskrete W.-verteilung zu $\Omega = \{0, 1\}$ und $f(1) = p$, $f(0) = 1 - p$ heißt *Bernoulli-Verteilung* zum Parameter p .

Interpretation: $1 \triangleq$ Ereignis tritt ein / „Erfolg“; $0 \triangleq$ Ereignis tritt nicht ein / „Misserfolg“; $p \triangleq$ Erfolgs-W.

Binomialverteilung.

Frage: Ein Bernoulli-Experiment (d.h. ein Zufallsexperiment, das sich durch eine Bernoulli-Verteilung beschreiben lässt) wird n -mal „unabhängig voneinander“ wiederholt. Mit welcher Wahrscheinlichkeit treten genau k Erfolge auf? ($0 \leq k \leq n$)

Antwort: Das n -stufige Bernoulli-Experiment lässt sich durch $\Omega = \{0, 1\}^n$ und $f(\omega) = f(\omega_1, \dots, \omega_n) = p^{\omega_1}(1-p)^{1-\omega_1} \dots p^{\omega_n}(1-p)^{1-\omega_n} = p^{\sum_{i=1}^n \omega_i} (1-p)^{n-\sum_{i=1}^n \omega_i}$ beschreiben, wobei zur Begründung der Wahl von f auf Beispiel 3 in Kapitel 0 verwiesen sei. Die Anzahl der Erfolge wird dann durch die Zufallsgröße

$$X : \Omega \rightarrow \{0, \dots, n\} \quad X(\omega) := \sum_{i=1}^n \omega_i;$$

beschrieben; ihre Verteilung ist durch $\tilde{\Omega} := \{0, \dots, n\}$ und

$$\tilde{f}(k) = \mathbb{P}(X^{-1}(\{k\})) = \sum_{\omega \in X^{-1}(\{k\})} f(\omega) = |X^{-1}(\{k\})| p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

($k = 0, \dots, n$) gegeben.

Definition 1.23 (Binomialverteilung).

Sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Die durch

$$f(k) := \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, \dots, n)$$

gegebene diskrete W. verteilung auf $\{0, \dots, n\}$ heißt *Binomialverteilung* mit den Parametern n und p , kurz $\mathcal{B}(n, p) = \mathcal{B}_{n,p}$.

Merkregel:

Die Binomialverteilung beschreibt die Anzahl der „Erfolge“ bei unabh. Versuchswiederholungen.

Frage: Aus einer Urne mit r roten und s schwarzen Kugeln wird n -mal mit Zurücklegen gezogen. Mit welcher W. wird genau k -mal eine rote Kugel gezogen? ($0 \leq k \leq n$)

Antwort 1: (*mittels Rechnung*) Wir nehmen an, dass die roten Kugeln von 1 bis r , die schwarzen Kugeln von $r+1$ bis $r+s$ durchnummeriert sind, und wählen $\Omega = \{1, \dots, r+s\}^n$, $\mathbb{P} = \mathcal{U}_\Omega$ (wegen Symmetrie) und $A_k := \{\omega \in \Omega : |\{i : \omega_i \leq r\}| = k\}$. Es gilt dann

$$\mathbb{P}(A_k) = \frac{|A_k|}{|\Omega|} = \frac{\binom{n}{k} r^k s^{n-k}}{(r+s)^n} = \binom{n}{k} \left(\frac{r}{r+s}\right)^k \left(\frac{s}{r+s}\right)^{n-k} = \mathcal{B}_{n,r/(r+s)}(\{k\});$$

die gesuchte W. ist also $\mathcal{B}_{n,r/(r+s)}(\{k\}) = \binom{n}{k} \left(\frac{r}{r+s}\right)^k \left(\frac{s}{r+s}\right)^{n-k}$.

Antwort 2: (*mittels Begründung*) Die Zufallsgröße X gebe an, wie oft rot gezogen wird. Da mit Zurücklegen gezogen wird, sind die Ergebnisse der einzelnen Ziehungen unabhängig voneinander, und X ist $\mathcal{B}_{n,r/(r+s)}$ -verteilt (als Anzahl der „Erfolge“ bei n unabhängigen Versuchswiederholungen mit Erfolgs-W. $\frac{r}{r+s}$). Die gesuchte W. ist also $\mathbb{P}_X(\{k\}) = \mathcal{B}_{n,r/(r+s)}(\{k\}) = \binom{n}{k} \left(\frac{r}{r+s}\right)^k \left(\frac{s}{r+s}\right)^{n-k}$.

Hypergeometrische Verteilung.

Frage: Aus einer Urne mit r roten und s schwarzen Kugeln wird n -mal ohne Zurücklegen gezogen. Mit welcher W. wird genau k -mal eine rote Kugel gezogen? ($0 \leq k \leq n \leq r+s$)

Antwort: Wir nehmen an, dass die roten Kugeln von 1 bis r , die schwarzen Kugeln von $r+1$ bis $r+s$ durchnummeriert sind, und wählen $\Omega = \{\omega \subseteq \{1, \dots, r+s\} : |\omega| = n\}$ und $\mathbb{P} = \mathcal{U}_\Omega$ (wegen Symmetrie). Wie oft rot gezogen wird, wird dann durch die Zufallsgröße

$$X : \Omega \rightarrow \{0, \dots, n\} \quad X(\omega) := |\omega \cap \{1, \dots, r\}|$$

beschrieben; ihre Verteilung ist durch $\tilde{\Omega} := \{0, \dots, n\}$ und

$$\tilde{f}(k) = \mathbb{P}(X^{-1}(\{k\})) = \frac{|X^{-1}(\{k\})|}{|\Omega|} = \frac{\binom{r}{k} \binom{s}{n-k}}{\binom{r+s}{n}}$$

($k = 0, \dots, n$) gegeben. Dabei sei (wie üblich) $\binom{k}{j} := 0$ gesetzt, falls $j < 0$ oder $j > k$ gilt.

Definition 1.24 (Hypergeometrische Verteilung).

Seien $r, s \in \mathbb{N}$ und $1 \leq n \leq r+s$. Dann heißt die durch

$$f(k) := \frac{\binom{r}{k} \binom{s}{n-k}}{\binom{r+s}{n}} \quad (k = 0, \dots, n)$$

gegebene diskrete W.verteilung auf $\{0, \dots, n\}$ *hypergeometrische Verteilung* mit den Parametern n, r, s , kurz $\mathcal{H}(n, r, s) = \mathcal{H}_{n,r,s}$.

Beachte, dass hier die Wahl und die Reihenfolge der Parameter in der Literatur nicht einheitlich sind – so werden auch $n, r+s, r$ oder $r+s, r, n$ als Parameter verwendet.

Merkregel:

Wird aus einer Urne mit r roten Kugeln und s schwarzen Kugeln n -mal gezogen und interessiert die Anzahl X der Ziehungen, bei denen rot gezogen wird, so ist X $\mathcal{B}(n, \frac{r}{r+s})$ - bzw. $\mathcal{H}(n, r, s)$ -verteilt, wenn mit bzw. ohne Zurücklegen gezogen wird.

Approximation der hypergeometrischen Verteilung durch die Binomialverteilung.

Sind in der obigen Situation r und s „sehr groß“ im Vergleich zu n , so sollte es keine Rolle spielen, ob mit oder ohne Zurücklegen gezogen wird.

Satz 1.25. Sind $(r_N)_{N \in \mathbb{N}}, (s_N)_{N \in \mathbb{N}}$ Folgen natürlicher Zahlen mit $\lim_{N \rightarrow \infty} r_N = \infty$, $\lim_{N \rightarrow \infty} s_N = \infty$ und $\lim_{N \rightarrow \infty} r_N / (r_N + s_N) = p$ und ist $n \in \mathbb{N}$ fest, so gilt $\lim_{N \rightarrow \infty} \mathcal{H}(n, r_N, s_N)(\{k\}) = \mathcal{B}(n, p)(\{k\})$ für alle $k = 0, \dots, n$.

Beweis. Für jedes feste $k \in \mathbb{N}_0$ gilt

$$\frac{1}{N^k} \binom{N}{k} = \frac{1}{N^k} \frac{N!}{(N-k)! k!} = \frac{N}{N} \cdot \frac{N-1}{N} \cdot \dots \cdot \frac{N-k+1}{N} \frac{1}{k!} \xrightarrow{N \rightarrow \infty} \frac{1}{k!}.$$

Damit folgt für alle $k = 0, \dots, n$ (durch geschicktes Erweitern)

$$\begin{aligned} \mathcal{H}(n, r_N, s_N)(\{k\}) &= \frac{\binom{r_N}{k} \binom{s_N}{n-k}}{\binom{r_N+s_N}{n}} = \frac{\frac{1}{r_N^k} \binom{r_N}{k} \frac{1}{s_N^{n-k}} \binom{s_N}{n-k}}{\frac{1}{(r_N+s_N)^n} \binom{r_N+s_N}{n}} \frac{r_N^k}{(r_N+s_N)^k} \frac{s_N^{n-k}}{(r_N+s_N)^{n-k}} \\ &\xrightarrow{N \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} = \mathcal{B}(n, p)(\{k\}). \end{aligned}$$

□

Poisson-Verteilung.

Frage: Wie verhält sich bei der Binomialverteilung die W. für (genau) k Erfolge, wenn man die Anzahl n der unabhängigen Versuchswiederholungen vergrößert und gleichzeitig die Erfolgs-W. p_n verkleinert, so dass $\lim_{n \rightarrow \infty} (np_n) = \lambda \in [0, \infty[$? (*Bemerkung:* Wir werden später sehen, dass np_n die mittlere Anzahl der Erfolge ist, so dass es nahe liegt, diesen Wert stabil zu halten.)

Antwort: Für jedes feste $k \in \mathbb{N}_0$ gilt

$$\lim_{n \rightarrow \infty} \left(\binom{n}{k} p_n^k (1-p_n)^{n-k} \right) = \lim_{n \rightarrow \infty} \left(\frac{1}{n^k} \binom{n}{k} (np_n)^k \left(1 - \frac{np_n}{n}\right)^{n-k} \right) = \frac{1}{k!} \lambda^k e^{-\lambda},$$

wobei einerseits die bereits verwendete Regel $\lim_{n \rightarrow \infty} \left(\frac{1}{n^k} \binom{n}{k} \right) = \frac{1}{k!}$ und andererseits die aus der Analysis bekannte Regel $\lim_{n \rightarrow \infty} a_n = a \Rightarrow \lim_{n \rightarrow \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$ eingegangen ist.

Beachte, dass diese Werte wegen $e^{-\lambda} \lambda^k / k! \geq 0$ für alle $k \in \mathbb{N}_0$ und $\sum_{k=0}^{\infty} e^{-\lambda} \lambda^k / k! = 1$ (*Exponentialreihe!*) in der Tat ein W.maß definieren. Wir definieren daher:

Definition 1.26 (Poisson-Verteilung).

Sei $\lambda \in [0, \infty[$. Die durch

$$f(k) := e^{-\lambda} \frac{\lambda^k}{k!} \quad (k \in \mathbb{N}_0)$$

gegebene diskrete W.verteilung auf \mathbb{N}_0 heißt *Poisson-Verteilung* mit dem Parameter λ , kurz $\mathcal{P}(\lambda) = \mathcal{P}_\lambda$.

Approximation der Binomialverteilung durch die Poisson-Verteilung

Es ergibt sich aus unserer „Herleitung“, dass die Poisson-Verteilung zur Approximation der Binomialverteilung verwendet werden kann, wenn n „groß“ und p „klein“ ist:

Satz 1.27 (Poisson'scher Grenzwertsatz). *Ist $(p_n)_{n \in \mathbb{N}}$ eine Folge in $[0, 1]$ mit $\lim_{n \rightarrow \infty} (np_n) = \lambda \in [0, \infty[$, so gilt $\lim_{n \rightarrow \infty} \mathcal{B}_{n, p_n}(\{k\}) = \mathcal{P}_\lambda(\{k\})$ für alle $k \in \mathbb{N}_0$.*

Bemerkung 1.28. Die Poisson-Verteilung liefert ein einfaches Modell zur Beschreibung der Anzahl der Ereignisse (Erdbeben, Unfälle, Telefonanrufe, E-Mails, Druckaufträge, Rechnerabstürze, ...) in einem festen Zeitraum.

Begründung. Sei $t > 0$ fest. Wir zerlegen das Intervall $I =]0, t]$ in n Teilintervalle $I_k =]\frac{(k-1)t}{n}, \frac{kt}{n}]$, $k = 1, \dots, n$, n „groß“, und treffen die folgenden Annahmen:

- (i) In jedem Teilintervall tritt höchstens 1 Ereignis auf.
- (ii) Die W. p hierfür ist proportional zur Intervalllänge, also $p = \lambda t/n$. ($\lambda \triangleq$ Intensität)
- (iii) Ereignisse in verschiedenen Intervallen sind „unabhängig voneinander“.

Dann ist die Anzahl der Ereignisse im Intervall $I =]0, t]$ (als Anzahl der „Erfolge“ bei unabhängigen Versuchswiederholungen) näherungsweise $\mathcal{B}_{n, \lambda t/n}$ - bzw. $\mathcal{P}_{\lambda t}$ -verteilt. \square

Geometrische Verteilung.

Frage: Ein Bernoulli-Experiment wird „unabhängig voneinander“ wiederholt, bis der erste Erfolg eintritt. Mit welcher W. geschieht dies im k -ten Versuch?

Antwort: Das k -stufige Bernoulli-Experiment lässt sich durch

$$\Omega = \{0, 1\}^k \quad \text{und} \quad f(\omega) = f(\omega_1, \dots, \omega_k) = p^{\omega_1} (1-p)^{1-\omega_1} \dots p^{\omega_k} (1-p)^{1-\omega_k}$$

beschreiben. Es interessiert dann das Ereignis $A_k := \{(0, \dots, 0, 1)\}$; seine W. ist $\mathbb{P}(A_k) = (1-p)^{k-1}p$.

Beachte, dass diese Werte im Fall $p \in]0, 1]$ wegen $(1-p)^{k-1}p \geq 0$ für alle $k \in \mathbb{N}$ und $\sum_{k=1}^{\infty} (1-p)^{k-1}p = 1$ (*geometrische Reihe!*) in der Tat ein W.maß definieren. Wir definieren daher:

Definition 1.29 (Geometrische Verteilung).

Sei $p \in]0, 1]$. Die durch

$$f(k) := (1-p)^{k-1}p \quad (k \in \mathbb{N})$$

gegebene diskrete W.verteilung auf \mathbb{N} heißt *geometrische Verteilung* mit dem Parameter p , kurz $\mathcal{G}(p) = \mathcal{G}_p$.

Merkregel:

Die geom. Verteilung beschreibt die Anzahl der Versuche bis zum ersten „Erfolg“ bei unabhängigen Versuchswiederholungen.

Beachte, dass die Literatur auch hier nicht einheitlich ist – oft wird auch die um 1 verschobene Verteilung auf \mathbb{N}_0 als geometrische Verteilung bezeichnet; hier zählt man die Anzahl der Misserfolge bis zum ersten Erfolg.

Zusammenfassung:

- Zufallabhängige Vorgänge, bei denen die Menge der möglichen Ergebnisse abzählbar ist, lassen sich durch diskrete W.räume beschreiben.
- Wir geben diskrete W.räume an, indem wir die Menge Ω der möglichen Ergebnisse sowie die W.dichte $f : \Omega \rightarrow [0, 1]$ angeben.
- Mit Hilfe der Rechenregeln für W.maße lassen sich die W.en von komplizierteren Ereignissen häufig auf die W.en von einfacheren Ereignissen zurückführen.
- Liegt eine diskrete Gleichverteilung vor, so lässt sich die Berechnung von W.en auf die Berechnung von Anzahlen zurückführen. (\leadsto Kombinatorik)
- Eine Zufallsgröße auf einem diskreten W.raum ist eine Abbildung $X : \Omega \rightarrow \mathcal{X}$ (mit Werten in einer abzählbaren Menge \mathcal{X}).
- Die Verteilung einer solchen Zufallsgröße bestimmen wir, indem wir die dazugehörigen Elementar-W.en $\mathbb{P}(X = x)$, $x \in \mathcal{X}$, bestimmen.
- Wir haben eine Reihe wichtiger diskreter W.verteilungen kennengelernt (Binomialverteilung, hypergeometrische Verteilung, Poisson-Verteilung, geometrische Verteilung).

2 Allgemeine Wahrscheinlichkeitsräume

2.1 Allgemeine Wahrscheinlichkeitsräume

Wir wollen uns nun zufallsabhängigen Vorgängen zuwenden, bei denen sich die Menge der möglichen Versuchsausgänge durch ein Intervall $(a, b) \subseteq \mathbb{R}$ mit $a < b$, also durch eine überabzählbare Menge, beschreiben lässt. Diese Situation tritt häufig dann auf, wenn kontinuierliche Größen (Zeiten, Längen, ...) gemessen werden, die vom Zufall beeinflusst sind. Diese Situation lässt sich nicht direkt durch einen diskreten W.raum beschreiben, da eine überabzählbare Menge vorliegt. Allerdings können wir sie *zumindest näherungsweise* durch einen diskreten W.raum beschreiben, indem wir das Problem diskretisieren: Dazu unterteilen wir die Zeit bzw. den Raum in viele kleine Abschnitte und notieren nur noch, in welchem Abschnitt das Ergebnis liegt. Anschließend untersuchen wir, was passiert, wenn wir eine „sehr feine“ Unterteilung zugrunde legen. Zur Motivation betrachten wir einige Beispiele:

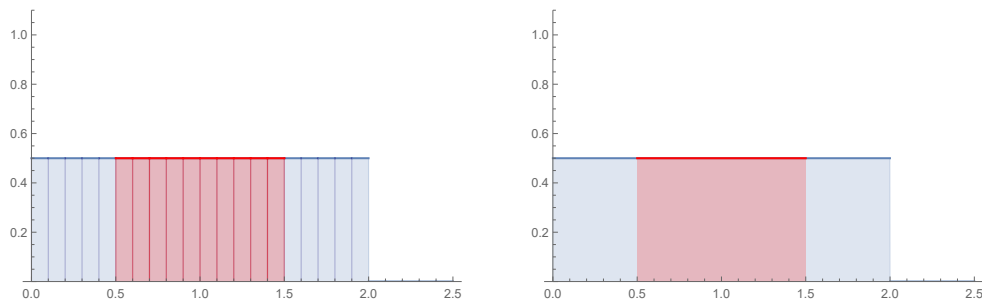
Beispiel 2.1.

(a) (Zufälliger Punkt)

Es soll ein zufälliger Punkt aus dem Intervall $]0, T]$ ausgewählt werden, wobei $T \in \mathbb{N}$ (der Einfachheit halber) eine feste natürliche Zahl sei und Intervalle $I \subseteq]0, T]$ derselben Länge dieselbe W. besitzen sollen. Wir unterteilen das Intervall $]0, T]$ (für großes $n \in \mathbb{N}$) in Abschnitte der Länge $1/n$. Dann bietet sich zur Beschreibung der diskrete W.raum $(\Omega_n, \mathfrak{P}(\Omega_n), \mathbb{P}_n)$ mit $\Omega_n := \{1, \dots, Tn\}$ und $\mathbb{P}_n := \mathcal{U}(\Omega_n)$ (d.h. $f_n(k) := \frac{1}{Tn}$ für alle $k \in \Omega_n$) an. Die W. eines Intervalls $I =]a, b] \subseteq]0, T]$ entspricht dann *näherungsweise* der Summe $\sum_{k \in \Omega_n : k/n \in I} f_n(k)$ bzw. (für $n \rightarrow \infty$) dem Integral

$$\int_I \frac{1}{T} dx = \int_a^b \frac{1}{T} dx = \left[\frac{x}{T} \right]_{x=a}^{x=b} = \frac{b-a}{T},$$

wie die nachstehende Graphik verdeutlicht:



Veranschaulichung der Approximation für $T = 2$, $n = 10$, $I =]0.5, 1.5]$

Dabei haben wir die Elementar-W.en von \mathbb{P}_n durch Flächeninhalte von Rechtecken (mit der Breite $\frac{1}{n}$ und der Höhe $\frac{1}{T}$) veranschaulicht. – Damit können wir für $n \rightarrow \infty$

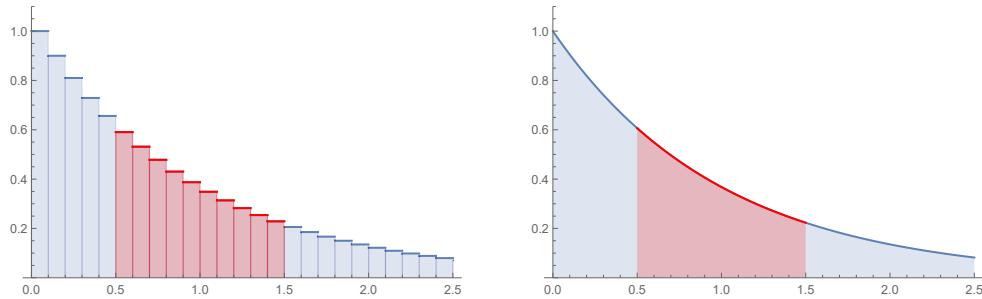
die Grundmenge $\Omega =]0, T]$ und die Funktion $f(x) = \frac{1}{T}$ wählen und die W. von I durch Integration von f über I berechnen. (Dieser Ansatz bleibt auch für $T \notin \mathbb{N}$ sinnvoll.)

(b) (*Wartezeitproblem*)

Wir warten an einer (wenig befahrenen) Straße auf das nächste Auto, wobei wir die Zeit in Stunden messen. Wir unterteilen das Intervall $(0, \infty)$ (für großes $n \in \mathbb{N}$) in Abschnitte der Länge $1/n$. Dann bietet sich zur Beschreibung (unter den Annahmen aus Bemerkung 1.28) der diskrete W.raum $(\Omega_n, \mathfrak{P}(\Omega_n), \mathbb{P}_n)$ mit $\Omega_n := \{1, 2, 3, \dots\}$ und $\mathbb{P}_n := \mathcal{G}(\lambda/n)$ (d.h. $f_n(k) = (1 - \frac{\lambda}{n})^{k-1} \frac{\lambda}{n}$ für alle $k \in \Omega_n$) an, wobei λ/n die W. für die Ankunft eines Autos in einem Abschnitt angibt. Die W. eines Intervalls $I =]a, b] \subseteq (0, \infty)$ entspricht dann *näherungsweise* der Summe $\sum_{k \in \Omega_n : k/n \in I} f_n(k)$ bzw. (für $n \rightarrow \infty$) dem Integral

$$\int_I \lambda e^{-\lambda x} dx = \int_a^b \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{x=a}^{x=b} = e^{-\lambda a} - e^{-\lambda b},$$

wie die nachstehende Graphik verdeutlicht:



Veranschaulichung der Approximation für $\lambda = 1$, $n = 10$, $I =]0.5, 1.5]$

Dabei haben wir die Elementar-W.en von \mathbb{P}_n wieder durch Flächeninhalte von Rechtecken veranschaulicht. – Damit können wir für $n \rightarrow \infty$ die Grundmenge $\Omega = (0, \infty)$ und die Funktion $f(x) = \lambda e^{-\lambda x}$ wählen und die W. von I durch Integration von f über I berechnen.

Ohne Beweis sei angemerkt, dass sich der Übergang von Summen zu Integralen, den wir hier nur graphisch begründet haben, auch rechnerisch begründen lässt; im Hintergrund steht dabei die folgende Beziehung zwischen Integralen und Summen:

Ist f auf $]a, b]$ Riemann-integrierbar, so gilt für großes $n \in \mathbb{N}$

$$\int_a^b f(x) dx \approx \sum_{k \in \mathbb{Z} : \frac{k}{n} \in]a, b]} f\left(\frac{k}{n}\right) \frac{1}{n}.$$

Da sich mit Integralen in der Regel einfacher rechnen lässt als mit Summen, wollen wir von nun an direkt ein stetiges Modell (statt einer diskreten Approximation) wählen:

Merkregel:

Gegeben ein Intervall $\Omega \subseteq \mathbb{R}$ und eine Funktion $f : \Omega \rightarrow \mathbb{R}$ mit den Eigenschaften

$$f(x) \geq 0 \text{ für alle } x \in \Omega \quad \text{und} \quad \int_{\Omega} f(x) dx = 1$$

(wobei die Existenz des – ggf. uneigentlichen – Riemann-Integrals vorauszusetzen ist) können wir das *stetige W.maß* \mathbb{P} betrachten, das jedem Intervall $I \subseteq \Omega$ die W.

$$\mathbb{P}(I) = \int_I f(x) dx$$

zuordnet. Die Funktion f wird auch als *W.dichte* von \mathbb{P} bezeichnet.

Beachte, dass diese Konstruktion von stetigen W.mäßen analog zur Konstruktion von diskreten W.mäßen ist, nur dass jetzt Integrale statt Summen gebildet werden. Es ist allerdings anzumerken, dass die Werte $f(x)$ hier nicht als Elementar-W.’en interpretiert werden dürfen: Für jede Einpunktmenge $\{x\} \subseteq I$ gilt nämlich

$$\mathbb{P}(\{x\}) = \int_{\{x\}} f(y) dy = 0,$$

was im Allgemeinen nicht dem Wert von $f(x)$ entspricht. Es kann sogar $f(x) > 1$ gelten, wie die W.dichte $f(x) = \lambda e^{-\lambda x}$ aus Beispiel 2.1 (b) mit $\lambda > 1$ zeigt. Daher muss man bei stetigen W.mäßen immer „in Intervallen“ statt „in Einpunktmengen“ denken.

Beispiele 2.2 (Beispiele für stetige W.verteilungen auf \mathbb{R}).

Dichte	Name der W.verteilung	Parameter
$f(x) = \frac{1}{b-a} \mathbf{1}_{(a,b)}(x)$	<i>Gleichverteilung auf $\langle a, b \rangle$</i> $\mathcal{U}(a, b)$	$-\infty < a < b < +\infty$
$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x)$	<i>Exponentialverteilung</i> $\mathcal{E}(\lambda)$	$\lambda > 0$
$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-(x-\mu)^2/2\sigma^2}$	<i>Normalverteilung</i> $\mathcal{N}(\mu, \sigma^2)$	$\mu \in \mathbb{R}, \sigma \in (0, \infty)$

Die (Nicht-Standard-!)Notation $\langle a, b \rangle$ soll andeuten, dass man sich aussuchen kann, ob man die Randpunkte hinzunimmt oder nicht. Beachte, dass wir hier als Grundmenge stets $\Omega = \mathbb{R}$ wählen und die Dichte bei Bedarf mit einer geeigneten Indikatorfunktion $\mathbf{1}_I$ multiplizieren; alternativ könnten wir auch (wie in Beispiel 2.1) als Grundmenge $\Omega = I$ wählen und die Indikatorfunktion in der Dichte weglassen.

Die Gleichverteilung und die Exponentialverteilung sind uns bereits in Beispiel 2.1 (a) bzw. (b) begegnet; insbesondere folgt daraus, dass sich die Exponentialverteilung zur Modellierung von Wartezeiten anbietet. Eine Bedingung für das Auftreten der Normalverteilung wird später durch den *Zentralen Grenzwertsatz* gegeben.

Beweis, dass die Funktionen in Beispiel 2.2 W.dichten sind.

(a) (*Gleichverteilung*)

Es gilt $f(x) \geq 0$ für alle $x \in \mathbb{R}$ sowie

$$\int_{\mathbb{R}} \frac{1}{b-a} \mathbf{1}_{(a,b)}(x) dx = \int_a^b \frac{1}{b-a} dx = \left[\frac{x}{b-a} \right]_{x=a}^{x=b} = \frac{b}{b-a} - \frac{a}{b-a} = 1.$$

(b) (*Exponentialverteilung*)

Es gilt $f(x) \geq 0$ für alle $x \in \mathbb{R}$ (da exp positiv) sowie

$$\int_{\mathbb{R}} \lambda e^{-\lambda x} \mathbf{1}_{(0,\infty)}(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_{x=0}^{x=\infty} = (-0) - (-1) = 1.$$

(c) (*Normalverteilung*)

Es gilt $f(x) \geq 0$ für alle $x \in \mathbb{R}$ (da exp positiv) sowie

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \stackrel{\substack{y=(x-\mu)/\sigma \\ dy=dx/\sigma}}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} dy = 1;$$

die letzte Gleichheit lässt sich nicht so einfach nachrechnen und soll daher als bekannt vorausgesetzt werden. □

Natürlich wollen wir nicht nur Intervallen $I \subseteq \mathbb{R}$, sondern auch allgemeineren Teilmengen $A \subseteq \mathbb{R}$ Wahrscheinlichkeiten zuordnen, unter anderem um die Rechenregeln für W. maße aus Satz 1.6 anwenden zu können. Dabei stößt man allerdings auf Probleme:

Bemerkung 2.3 (Zufälliger Punkt). Wir betrachten Beispiel 2.1 (a) speziell für $T = 1$, d. h. wir wählen $\Omega =]0, 1]$ und fordern, dass die W. einer Menge $A \subset \Omega$ proportional zur „Länge“ dieser Menge ist. Als W. maß hätten wir gern eine Abbildung $\mathbb{P} : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$, welche die aus Satz 1.6 bekannten Eigenschaften

- (i) $\forall A \in \mathfrak{P}(\Omega) : \mathbb{P}(A) \geq 0$ (*Nicht-Negativität*)
- (ii) $\mathbb{P}(\Omega) = 1$ (*Normiertheit*)
- (iii) $\forall A_1, A_2, \dots \in \mathfrak{P}(\Omega) : A_1, A_2, \dots$ paarweise disjunkt $\Rightarrow \mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ (*σ -Additivität*)

besitzt und darüber hinaus sowie die „Symmetrie-Bedingung“

- (iv) $\forall A \in \mathfrak{P}(\Omega) \forall t \in (0, 1) : A \subseteq (0, 1) \wedge A + t \subseteq (0, 1) \Rightarrow \mathbb{P}(A) = \mathbb{P}(A + t)$

erfüllt. Dabei sei $A + t := \{\omega + t : \omega \in A\}$. Anschaulich bedeutet Bedingung (iv), dass sich die W. nicht verändert, wenn wir die Menge A innerhalb des Intervalls $\Omega = (0, 1)$ verschieben.

Problem: Es lässt sich (mit Hilfe des Auswahlaxioms aus der Mengenlehre) zeigen, dass eine Abbildung $\mathbb{P} : \mathfrak{P}(\Omega) \rightarrow \mathbb{R}$ mit den Eigenschaften (i) – (iv) nicht existiert!

Lösung: Wir ersetzen den Definitionsbereich $\mathfrak{P}(\Omega)$ durch ein kleineres Mengensystem $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ und suchen eine Abbildung $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$, so dass die Eigenschaften (i) – (iv) für alle Mengen in \mathcal{A} gelten. Dabei fordern wir, dass das System \mathcal{A} zum einen gewisse „einfache“ Mengen (hier z. B. alle Intervalle) sowie die Mengen \emptyset und Ω enthält, denen offensichtliche Wahrscheinlichkeiten zugeordnet werden können, und zum anderen abgeschlossen unter den Mengenoperationen $^c, \cup, \cap, \setminus, \Delta, \bigcup_{n=1}^{\infty}, \bigcap_{n=1}^{\infty}$ ist, damit die Rechenregeln aus Satz 1.6 sinnvoll bleiben. \square

Dies führt zu den folgenden Begriffsbildungen:

Definition 2.4 (σ -Algebra). Sei $\Omega \neq \emptyset$. Ein Mengensystem $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ heißt σ -Algebra über Ω , falls gilt:

- (i) $\Omega \in \mathcal{A}$
- (ii) $\forall A \in \mathcal{A} : A^c \in \mathcal{A}$
- (iii) $\forall A_1, A_2, \dots \in \mathcal{A} : \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$

Bemerkung 2.5. Jede σ -Algebra ist abgeschlossen unter den Mengenoperationen $^c, \cup, \cap, \setminus, \Delta, \bigcup_{n=1}^{\infty}, \bigcap_{n=1}^{\infty}$.

Bemerkung 2.6. σ -Algebren werden i. d. R. dadurch beschrieben, dass man ein System von Mengen $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$ vorgibt und dann die kleinste σ -Algebra über Ω betrachtet, die \mathcal{E} enthält; diese σ -Algebra wird mit $\sigma(\mathcal{E})$ bzw. als *die von \mathcal{E} erzeugte σ -Algebra* bezeichnet.

Definition 2.7 (Wahrscheinlichkeitsmaß). Sei $\Omega \neq \emptyset$ und \mathcal{A} eine σ -Algebra über Ω . Eine Abbildung $\mathbb{P} : \mathcal{A} \rightarrow \mathbb{R}$ heißt *Wahrscheinlichkeitsmaß* auf \mathcal{A} , falls gilt:

- (i) $\forall A \in \mathcal{A} : \mathbb{P}(A) \geq 0$ (*Nicht-Negativität*)
- (ii) $\mathbb{P}(\Omega) = 1$ (*Normiertheit*)
- (iii) $\forall A_1, A_2, \dots \in \mathcal{A} : A_1, A_2, \dots$ paarweise disjunkt $\Rightarrow \mathbb{P}(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$ (σ -Additivität)

Bemerkung 2.8. Sei $\Omega \neq \emptyset$, \mathcal{A} eine σ -Algebra über Ω und \mathbb{P} ein W.maß auf \mathcal{A} . Satz 1.6 und Lemma 1.8 gelten entsprechend, *sofern* man überall nur Mengen aus \mathcal{A} zulässt.

Bemerkung 2.9. W.maße werden i. d. R. dadurch konstruiert, dass man ihre Werte auf einem (geeigneten) Erzeugendensystem \mathcal{E} von \mathcal{A} vorschreibt und sich überlegt, dass genau eine Fortsetzung zu einem W.maß auf \mathcal{A} existiert.

Definition 2.10 (Wahrscheinlichkeitsraum, Kolmogorov 1933).

Ein *W.raum* ist ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$, wobei Ω eine nicht-leere Menge, \mathcal{A} eine σ -Algebra über Ω und \mathbb{P} ein W.maß auf \mathcal{A} ist.

Wir bezeichnen die Elemente von Ω als *Ergebnisse*, die Elemente von \mathcal{A} (und nur diese!) als *Ereignisse* (oder auch als *messbare Mengen*) und die Werte von \mathbb{P} als *Wahrscheinlichkeiten*. Ferner bezeichnen wir \mathbb{P} auch als W.maß auf dem *messbaren Raum* (Ω, \mathcal{A}) oder – wenn die Wahl der σ -Algebra klar ist – als W.maß auf Ω .

Die Wahl der σ -Algebra bzw. des W.maßes ist im Allgemeinen ein schwieriges Problem. Glücklicherweise gibt es einige Konstruktionen, die die meisten Anwendungen abdecken:

Diskrete W.räume.

Sei $\Omega \neq \emptyset$ abzählbar.

Als \mathcal{E} wählen wir das System aller Einpunktmengen: $\mathcal{E} = \{\{\omega\} : \omega \in \Omega\}$. Dann gilt $\mathcal{A} = \sigma(\mathcal{E}) = \mathfrak{P}(\Omega)$.

Ist $f : \Omega \rightarrow \mathbb{R}$ eine Abbildung mit $f(\omega) \geq 0$ für alle $\omega \in \Omega$ und $\sum_{\omega \in \Omega} f(\omega) = 1$, so existiert genau ein W.maß \mathbb{P} auf \mathcal{A} mit $\mathbb{P}(\{x\}) = f(x)$ für alle $x \in \Omega$, nämlich das W.maß aus Definition 1.1. \mathbb{P} heißt *diskretes W.maß mit der (W.-)Dichte f* .

Genauer ist jedes W.maß auf \mathcal{A} von dieser Gestalt. (*Übung!*)

Stetige W.räume über \mathbb{R} .

Sei $\Omega = \mathbb{R}$.

Als \mathcal{E} wählen wir das System aller Intervalle der Form $]a, b]$: $\mathcal{E} = \{]a, b] : a, b \in \mathbb{R}\}$. Dann heißt $\mathcal{A} := \mathbb{B} := \sigma(\mathcal{E})$ *Borel- σ -algebra* über \mathbb{R} . Es ist nicht möglich (aber auch nicht nötig), \mathbb{B} explizit anzugeben.

Ist $f : \mathbb{R} \rightarrow \mathbb{R}$ eine „integrierbare“ Funktion mit $f(x) \geq 0$ für alle $x \in \mathbb{R}$ und $\int_{\mathbb{R}} f(x) dx = 1$, so existiert genau ein W.maß \mathbb{P} auf \mathcal{A} mit $\mathbb{P}(]a, b]) = \int_{]a, b]} f(x) dx$ für alle Intervalle $]a, b]$. \mathbb{P} heißt *stetiges W.maß mit der (W.-)Dichte f* .

Allerdings ist nicht jedes W.maß auf \mathcal{A} von dieser Gestalt. (*Übung!*)

Dabei heißt eine Funktion „integrierbar“, wenn sie uneigentlich Riemann-integrierbar ist.

Bemerkung 2.11. Sei \mathbb{P} ein stetiges W.maß über \mathbb{R} mit der Dichte f .

- (i) Es gilt $\mathbb{P}(\{x\}) = 0$ für alle $x \in \mathbb{R}$.
- (ii) Weiter gilt damit wegen der Additivität / σ -Additivität von \mathbb{P} auch $\mathbb{P}(A) = 0$ für jede endliche / abzählbare Teilmenge A von \mathbb{R} ; insbesondere kann \mathbb{P} damit kein diskretes W.maß sein!
- (iii) $f(x)$ lässt sich hier nicht mehr als (Elementar-)Wahrscheinlichkeit interpretieren; es kann sogar $f(x) > 1$ gelten!

- (iv) Man kann f (z.B. an endlich vielen Stellen) abändern, ohne \mathbb{P} zu verändern. Anders als bei diskreten Modellen ist die W.dichte f also durch das W.maß \mathbb{P} nicht (ganz) eindeutig bestimmt.
- (v) Es ist im Allgemeinen schwierig zu entscheiden, ob eine Menge $A \subseteq \mathbb{R}$ in \mathbb{B} liegt oder ob für eine Menge $A \in \mathbb{B}$ die W. $\mathbb{P}(A)$ durch Integration bestimmt werden kann. Da dies bei den in Anwendungen auftretenden Mengen in aller Regel der Fall ist, wollen wir dies immer stillschweigend voraussetzen:

Merkregel:

Ist \mathbb{P} ein stetiges W.maß über \mathbb{R} mit einer Dichte f , so können wir für die in Anwendungen auftretenden Mengen $A \subseteq \mathbb{R}$ die W.en $\mathbb{P}(A)$ bestimmen, indem wir die Dichte f über die Menge A integrieren, also das Integral $\int_A f(x) dx := \int_{\mathbb{R}} \mathbf{1}_A(x) f(x) dx$ berechnen.

- (vi) Man kann auch stetige W.maße über Teilmengen $\Omega \subseteq \mathbb{R}$ einführen; hier wählt man als σ -Algebra $\mathcal{A} := \mathbb{B}|_{\Omega} = \sigma(\{]a, b] \cap \Omega : a, b \in \mathbb{R}\})$, die *Borel- σ -algebra* über Ω , und als W.maß \mathbb{P} das (eindeutige) W.maß auf \mathcal{A} mit $\mathbb{P}(]a, b] \cap \Omega) = \int_{]a, b] \cap \Omega} f(x) dx$ für alle Intervalle $]a, b]$. Dabei sei $f : \Omega \rightarrow \mathbb{R}$ eine Funktion, die zu einer (W.-)Dichte mit den o.g. Eigenschaften wird, wenn wir sie durch $f(x) := 0$ für $x \notin \Omega$ auf \mathbb{R} fortsetzen.

Bei der Berechnung von Wahrscheinlichkeiten ist es gelegentlich nützlich, zuerst die Rechenregeln für W.maße anzuwenden und anschließend die Wahrscheinlichkeiten der verbleibenden (einfacheren) Mengen durch Integration zu bestimmen:

Beispiel 2.12 (Wartezeitproblem). Wir betrachten Beispiel 2.1 (b). Mit welcher W. fährt das erste Auto innerhalb der ersten Stunde oder nach der vierten Stunde vorbei?

Lösung:

Wir wissen bereits, dass sich der zufällige Zeitpunkt, zu dem das erste Auto vorbeifährt, durch die Exponentialverteilung $\mathcal{E}(\lambda)$ beschreiben lässt. Wir wählen daher $\Omega := \mathbb{R}$, $\mathcal{A} := \mathbb{B}$, $\mathbb{P} :=$ stetiges W.maß über \mathbb{R} mit der Dichte $f(x) = \lambda e^{-\lambda x} \mathbf{1}_{(0, \infty)}(x)$ und $A := (0, 1) \cup (4, \infty)$. Damit ergibt sich wegen $A = (0, \infty) \setminus [1, 4]$

$$\mathbb{P}(A) \underset{\text{Subtraktivität}}{=} \mathbb{P}((0, \infty)) - \mathbb{P}([1, 4]) = 1 - \int_1^4 \lambda e^{-\lambda x} dx = 1 - \left[-e^{-\lambda x} \right]_{x=1}^{x=4} = 1 - e^{-4\lambda} + e^{-\lambda}.$$

Alternativ hätten wir natürlich auch $\mathbb{P}(A) = \mathbb{P}((0, 1)) + \mathbb{P}((4, \infty))$ schreiben können und die W.en der einzelnen Intervalle berechnen können.

Eine weitere Möglichkeit bestünde darin, $\Omega := (0, \infty)$, $\mathcal{A} := \mathbb{B}|_{\Omega}$, $\mathbb{P} :=$ stetiges W.maß über $(0, \infty)$ mit der Dichte $f(x) = \lambda e^{-\lambda x}$ zu wählen; dies führt natürlich zum selben Ergebnis. □

Stetige W.räume über \mathbb{R}^n .

Sei $\Omega = \mathbb{R}^n$.

Als \mathcal{E} wählen wir das System aller n -dimensionaler Intervalle der Form $]a, b[:=]a_1, b_1[\times \cdots \times]a_n, b_n[$: $\mathcal{E} = \{]a, b[: a, b \in \mathbb{R}^n\}$. Dann heißt $\mathcal{A} := \mathbb{B}^n := \sigma(\mathcal{E})$ *Borel- σ -algebra über \mathbb{R}^n* .

Ist $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine „integrierbare“ Funktion mit $f(x) \geq 0$ für alle $x \in \mathbb{R}$ und $\int_{\mathbb{R}^n} f(x) dx = 1$, so existiert genau ein W.maß \mathbb{P} auf \mathcal{A} mit $\mathbb{P}(]a, b[) = \int_{]a, b[} f(x) dx$ für alle n -dimensionalen Intervalle $]a, b[$. \mathbb{P} heißt *stetiges W.maß mit der (W.-)Dichte f* .

Dabei wollen wir eine Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ als „integrierbar“ bezeichnen, wenn für jedes n -dimensionale Intervall $]a, b[:=]a_1, b_1[\times \cdots \times]a_n, b_n[$ ($a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R} \cup \{\pm\infty\}$) das iterierte Riemann-Integral existiert und jede Integrationsreihenfolge zum selben Ergebnis führt.

Für $n = 2$ ist also z. B.

$$\int_{]a, b[} f(x) dx = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_2 dx_1 = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2,$$

wobei die Doppelintegrale voraussetzungsgemäß existieren und übereinstimmen müssen. Für die in Anwendungen auftretenden Funktionen ist diese Bedingung üblicherweise erfüllt, so dass wir im Folgenden nicht weiter darauf eingehen.

Bemerkung 2.11 gilt hier entsprechend.

Beispiel 2.13 (Stetige Gleichverteilung). Es sei $B \subseteq \mathbb{R}^n$ eine Menge mit

$$\text{vol}(B) := \int_{\mathbb{R}^n} \mathbf{1}_B(x) dx \in (0, \infty)$$

(d. h. insbesondere muss $\mathbf{1}_B$ integrierbar sein). Dann heißt das stetige W.maß \mathbb{P} auf \mathbb{B}^n mit der Dichte $f(x) := \frac{1}{\text{vol}(B)} \mathbf{1}_B(x)$ (*stetige Gleichverteilung auf B* , kurz \mathcal{U}_B , und es gilt

$$\mathbb{P}(A) = \frac{\text{vol}(A \cap B)}{\text{vol}(B)}$$

für jedes n -dimensionale Intervall $A =]a, b[$ (bzw. allgemeiner sogar für jede Menge $A \in \mathbb{B}^n$, so dass $\text{vol}(A \cap B)$ definiert ist). Anschaulich ist $\text{vol}(A)$ das n -dim. Volumen von A ; die Bestimmung von W.en lässt sich also auf die Bestimmung von Volumina zurückführen. (Beachte die Analogie zu Bemerkung 1.4.) \square

Im Fall $n = 1$, $B = (a, b)$ erhalten wir gerade die Gleichverteilung auf (a, b) zurück.

Für einfache geometrische Figuren ist es i. d. R. einfacher, die Volumina mit den Formeln aus der Schule als mit dem Integral zu bestimmen:

Beispiel 2.14. Mit welcher W. liegt ein rein zufällig gewählter Punkt im Einheitskreis $K := \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ im 1. Quadranten (d. h. in $(0, \infty)^2$) ?

Lösung:

Wir wählen $\Omega := \mathbb{R}^2$, $\mathcal{A} := \mathbb{B}^2$, $\mathbb{P} :=$ Gleichverteilung auf K und $A := (0, \infty)^2$.

1. *Lösung:* Aus Symmetriegründen hat der Durchschnitt von K mit jedem Quadranten den Flächeninhalt $\frac{1}{4} \text{vol}(K)$. Damit folgt

$$\mathbb{P}(A) = \frac{\text{vol}(A \cap K)}{\text{vol}(K)} = \frac{\frac{1}{4} \text{vol}(K)}{\text{vol}(K)} = \frac{1}{4}.$$

(Genauer gilt nach der Formel für die Kreisfläche $\text{vol}(K) = \pi$ sowie $\text{vol}(A \cap K) = \frac{1}{4}\pi$, aber dies müssen wir hier nicht unbedingt wissen.)

2. *Lösung:* Die Dichte der Gleichverteilung auf K ist (wegen $\text{vol}(K) = \pi$) durch $f = \frac{1}{\pi} \mathbf{1}_K$ gegeben. Damit folgt

$$\begin{aligned} \mathbb{P}(A) &= \int_A f(x) dx = \int_{\mathbb{R}^2} \mathbf{1}_A(x) \frac{1}{\pi} \mathbf{1}_K(x) dx = \int_{\mathbb{R}^2} \frac{1}{\pi} \mathbf{1}_{A \cap K}(x) dx = \int_0^1 \int_0^{\sqrt{1-x_1^2}} \frac{1}{\pi} dx_2 dx_1 \\ &= \int_0^1 \frac{1}{\pi} \sqrt{1-x_1^2} dx_1 \stackrel{\text{Formelsammlung}}{=} \left[\frac{1}{2\pi} (x \sqrt{1-x^2} + \arcsin(x)) \right]_0^1 = \frac{1}{4}. \end{aligned}$$

Im letzten Schritt in der ersten Zeile haben wir das 2-dimensionale Integral (entsprechend unserer Definition) durch ein iteriertes Integral ersetzt und die Indikatorfunktion durch die Wahl der Integrationsgrenzen „beseitigt“; dies macht man sich am besten anhand einer Skizze klar.

Dieser Lösungsweg ist aufwendiger; allerdings bringt er den Vorteil mit sich, dass er sich auf allgemeinere W.-Dichten f auf \mathbb{R}^2 und auf allgemeinere Teilmengen A von \mathbb{R}^2 übertragen lässt. \square

Bemerkung (Warnung). Beachte, dass wir die Bezeichnung „Dichte“ sowohl bei diskreten als auch bei stetigen W.mäßen verwenden. Welcher Typ gemeint ist, ergibt sich aus dem Zusammenhang. Wenn wir verdeutlichen wollen, welcher Typ gemeint ist, verwenden wir auch die genaueren Bezeichnungen „Zähl-Dichte“ im diskreten Fall bzw. „Riemann-Dichte“ im stetigen Fall. – Ebenso kann der Begriff „Gleichverteilung“ entweder ein diskretes oder ein stetiges W.maß bezeichnen, je nachdem ob die Grundmenge abzählbar oder kontinuierlich ist.

Beschreibung aller W.maße auf (\mathbb{R}, \mathbb{B}) .

Wir wollen nun *alle* W.maße auf (\mathbb{R}, \mathbb{B}) beschreiben:

Satz 2.15. *Ist \mathbb{P} ein W.maß auf (\mathbb{R}, \mathbb{B}) , so ist $F : \mathbb{R} \rightarrow \mathbb{R}$ mit $F(x) := \mathbb{P}(]-\infty, x])$ eine Abbildung mit den folgenden Eigenschaften:*

- (i) *F ist monoton wachsend.*
- (ii) *F ist rechtsseitig stetig.*
- (iii) *$\lim_{x \rightarrow -\infty} F(x) = 0$.*
- (iv) *$\lim_{x \rightarrow +\infty} F(x) = 1$.*

Umgekehrt gibt es zu jeder solchen Abbildung F (genau) ein W.maß \mathbb{P} auf (\mathbb{R}, \mathbb{B}) , so dass $\mathbb{P}(]-\infty, x]) = F(x)$ für alle $x \in \mathbb{R}$ gilt.

Beweisidee. Die Eigenschaften von F folgen aus denen von \mathbb{P} (vgl. dazu Bemerkung 2.8), nämlich (i) der Monotonie, (ii) der Stetigkeit von oben, (iii) der Stetigkeit von oben und $\mathbb{P}(\emptyset) = 0$, (iv) der Stetigkeit von unten und $\mathbb{P}(\Omega) = 1$.

Der Beweis der Umkehrung ist deutlich schwieriger: Die Grundidee besteht darin, für beschränkte Intervalle der Form $]a, b]$

$$\mathbb{P}(]a, b]) := F(b) - F(a)$$

zu setzen und sich zu überlegen, dass sich diese Abbildung auf eindeutige Weise zu einem W.maß \mathbb{P} auf \mathbb{B} fortsetzen lässt. \square

Definition 2.16 (Verteilungsfunktion). Eine Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ mit den Eigenschaften (i) – (iv) in Satz 2.15 heißt *Verteilungsfunktion* auf \mathbb{R} .

Genauer heißt bei gegebenem W.maß \mathbb{P} auf (\mathbb{R}, \mathbb{B}) die Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$ mit $F(x) = \mathbb{P}(]-\infty, x])$ für alle $x \in \mathbb{R}$ die *Verteilungsfunktion zum W.maß \mathbb{P}* (oder auch nur *Verteilungsfunktion von \mathbb{P}*) und umgekehrt bei gegebener Verteilungsfunktion F auf \mathbb{R} das W.maß \mathbb{P} auf (\mathbb{R}, \mathbb{B}) mit $\mathbb{P}(]a, b]) = F(b) - F(a)$ für alle $a, b \in \mathbb{R}$ mit $a < b$ das *W.maß zur Verteilungsfunktion F* .

Nach Satz 2.15 besteht also eine 1:1-Beziehung zwischen W.maßen auf (\mathbb{R}, \mathbb{B}) und Verteilungsfunktionen auf \mathbb{R} .

Bemerkung 2.17 (Wahrscheinlichkeitsmaße auf (\mathbb{R}, \mathbb{B})).

- (a) Ist \mathbb{P} ein W.maß auf (\mathbb{R}, \mathbb{B}) , so dass die zugehörige Verteilungsfunktion F eine „Sprungfunktion“ ist, die an den Stellen $x_1, x_2, x_3, \dots \in \mathbb{R}$ „springt“ und ansonsten konstant ist, so können wir

$$\Omega_0 := \{x_1, x_2, x_3, \dots\} \quad \text{und} \quad f_0 : \Omega_0 \rightarrow \mathbb{R}, \quad f_0(x) := F(x) - F(x-)$$

setzen (wobei $F(x-)$ den linksseitigen Grenzwert von F an der Stelle x bezeichnet), um \mathbb{P} als diskretes W.maß mit der Dichte f_0 nachzuweisen.

Die diskreten W.maße auf (\mathbb{R}, \mathbb{B}) entsprechen also den „Sprungfunktionen“ unter den Verteilungsfunktionen. (*Bemerkung:* Die Menge der Sprungstellen ist abzählbar, kann aber dicht in \mathbb{R} liegen.) Dabei werden die diskreten W.maße auf (\mathbb{R}, \mathbb{B}) zunächst auf $(\Omega_0, \mathfrak{P}(\Omega_0))$ mit $\Omega_0 \subset \mathbb{R}$ *abzählbar* definiert und anschließend auf (\mathbb{R}, \mathbb{B}) fortgesetzt, indem man $\mathbb{P}(B) := \mathbb{P}_0(B \cap \Omega_0)$ setzt.

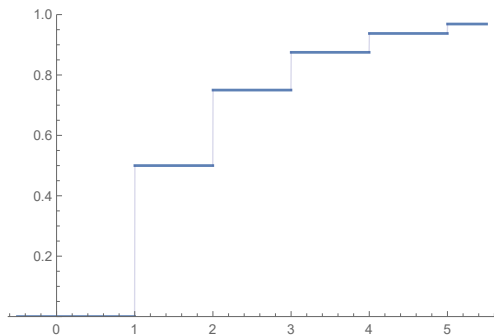
- (b) Ist \mathbb{P} ein W.maß auf (\mathbb{R}, \mathbb{B}) , so dass die zugehörige Verteilungsfunktion F stetig und (stückweise) stetig differenzierbar ist, so können wir

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) := F'(x)$$

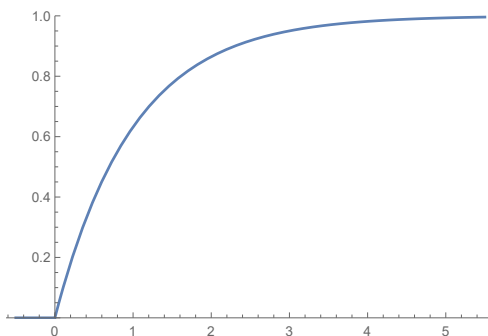
setzen (wobei wir $f(x) := 0$ setzen, wenn die Ableitung nicht existiert), um \mathbb{P} als stetiges W.maß mit der Dichte f nachzuweisen.

Die stetigen W.maße mit (stückweise) stetigen Dichten auf (\mathbb{R}, \mathbb{B}) entsprechen also den stetigen und (stückweise) stetig differenzierbaren Verteilungsfunktionen.

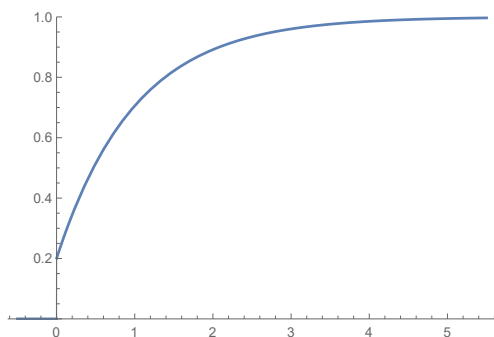
- (c) Es gibt weitere Typen von W.mäßen, z. B. „Mischtypen“ oder die *Cantor-Verteilung*. [...]



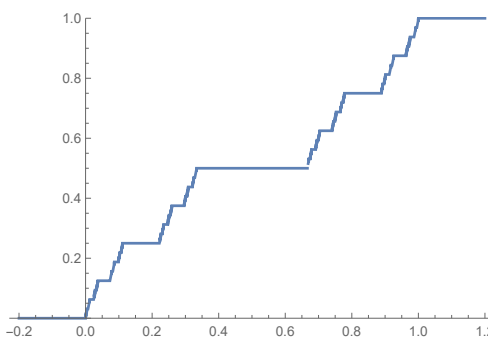
Verteilungsfunktion der $\mathcal{G}(\frac{1}{2})$ -Vtlg.



Verteilungsfunktion der $\mathcal{E}(1)$ -Vtlg.



Verteilungsfunktion einer Vtlg. vom Mischtyp



Verteilungsfunktion der Cantor-Vtlg.

Auf ähnliche Weise lassen sich W.maße auf $(\mathbb{R}^n, \mathbb{B}^n)$ mit Hilfe von *mehrdimensionalen Verteilungsfunktionen* $F : \mathbb{R}^n \rightarrow \mathbb{R}$ beschreiben. Allerdings ist die Charakterisierung, wann eine Funktion $F : \mathbb{R}^n \rightarrow \mathbb{R}$ eine Verteilungsfunktion, etwas aufwendiger, so dass wir nicht weiter darauf eingehen wollen.

Bemerkung (Berechnung von Wahrscheinlichkeiten bei allgemeinen W.maßen).

Ist ein W.maß \mathbb{P} durch seine Verteilungsfunktion F gegeben, so kann man die W. eines Intervalls $]a, b]$ mit der Formel

$$\mathbb{P}(]a, b]) = F(b) - F(a)$$

bestimmen und die W. eines allgemeinen Ereignisses $B \in \mathbb{B}$ dadurch bestimmen, dass man dieses (mit den üblichen mengentheoretischen Operationen) auf Intervalle zurückführt.

Ist \mathbb{P} ein diskretes oder stetiges W.maß, so ist es auch möglich, zunächst die Zähldichte bzw. die Riemann-Dichte zu bestimmen und anschließend eine Summe bzw. ein Integral über diese Dichte zu berechnen. Dieser Zugang⁴ ist (im Prinzip) auch noch anwendbar, wenn \mathbb{P} ein „Mischtyp“ einer diskreten und einer stetigen W.verteilung ist (vgl. dazu das dritte Beispiel auf Seite 40); hier besitzt der diskrete Anteil eine diskrete Dichte f_1 und der stetige Anteil eine Riemann-Dichte f_2 (wobei $\sum_{x \in \Omega_0} f_1(x) + \int_{\mathbb{R}} f_2(x) dx \stackrel{!}{=} 1$), und für „gutartige“ Mengen $A \in \mathbb{B}$ gilt

$$\mathbb{P}(A) = \sum_{x \in A \cap \Omega_0} f_1(x) + \int_A f_2(x) dx.$$

Im dritten Beispiel auf Seite 40 gilt etwa $\Omega_0 = \{0\}$, $f_1(0) = 0.2$, $f_2(x) = 0.8 e^{-x} \mathbf{1}_{(0, \infty)}(x)$ und damit $\mathbb{P}([0, 1]) = f_1(0) + \int_0^1 0.8 e^{-x} dx = 0.2 + 0.8(1 - e^{-1}) = 1 - 0.8 e^{-1}$. \square

2.2 Zufallsgrößen auf allgemeinen W.räumen

Wir wollen nun auf allgemeinen W.räumen Zufallsgrößen einführen. Auch diese sollten wieder eine induzierte Verteilung, gegeben durch $\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B))$, besitzen. Da \mathbb{P} nun nur noch auf \mathcal{A} definiert ist, brauchen wir eine Zusatzbedingung, damit nur Mengen aus \mathcal{A} als Urbildmengen auftreten:

Definition 2.18. Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein allgemeiner W.raum, $(\mathcal{X}, \mathcal{B})$ ein messbarer Raum und $X : \Omega \rightarrow \mathcal{X}$ eine \mathcal{A} - \mathcal{B} -messbare Abbildung (kurz: $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$), d. h. es gelte $X^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{B}$. Dann heißt X *Zufallsgröße* oder *Zufallsvariable* (mit Werten in $(\mathcal{X}, \mathcal{B})$), und die Abbildung $\mathbb{P}_X : \mathcal{B} \rightarrow \mathbb{R}$ mit $\mathbb{P}_X(B) := \mathbb{P}(X^{-1}(B))$ für alle $B \in \mathcal{B}$ heißt *induzierte Verteilung* (oder auch nur *Verteilung*) von X unter \mathbb{P} .

Im Folgenden schreiben wir auch wieder $\{X \in B\}$ statt $X^{-1}(B)$.

Lemma 2.19. In der Situation von Definition 2.18 ist \mathbb{P}_X ein W.maß auf $(\mathcal{X}, \mathcal{B})$.

⁴ Diese Anmerkung ist in der Vorlesung übersprungen worden.

Beweis. Man prüft nach, dass \mathbb{P}_X nicht-negativ, normiert und σ -additiv ist:

$$(a) \quad \forall B \in \mathcal{B} : \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) \underset{\mathbb{P} \text{ W.maß}}{\geq} 0.$$

$$(b) \quad \mathbb{P}_X(\mathcal{X}) = \mathbb{P}(X^{-1}(\mathcal{X})) = \mathbb{P}(\Omega) \underset{\mathbb{P} \text{ W.maß}}{=} 1.$$

$$(c) \quad \forall B_1, B_2, B_3, \dots \in \mathcal{B} : B_1, B_2, B_3, \dots \text{ paarweise disjunkt} \Rightarrow$$

$$\mathbb{P}_X\left(\bigcup_{k=1}^{\infty} B_k\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right)\right) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} X^{-1}(B_k)\right) \underset{\mathbb{P} \text{ W.maß}}{=} \sum_{k=1}^{\infty} \mathbb{P}(X^{-1}(B_k)) = \sum_{k=1}^{\infty} \mathbb{P}_X(B_k).$$

Dabei ist eingegangen, dass die Urbilder disjunkter Mengen wieder disjunkt sind. \square

Lemma 2.20. Sind $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum, $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$ eine Zufallsgröße und $h : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathcal{Y}, \mathcal{C})$ eine messbare Abbildung, so ist auch $h(X) := h \circ X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{Y}, \mathcal{C})$ eine Zufallsgröße, und es gilt $\mathbb{P}_{h(X)} = (\mathbb{P}_X)_h$.

Beweis. Dass $h \circ X$ eine Zufallsgröße ist, folgt aus

$$C \in \mathcal{C} \xrightarrow{h \text{ messbar}} h^{-1}(C) \in \mathcal{B} \xrightarrow{X \text{ messbar}} (h \circ X)^{-1}(C) = X^{-1}(h^{-1}(C)) \in \mathcal{A}.$$

Dass $\mathbb{P}_{h(X)} = (\mathbb{P}_X)_h$, zeigt man genauso wie in Lemma 1.17. \square

Ist der Grundraum Ω „diskret“ bzw. „kontinuierlich“, so verzichten wir im Folgenden häufig auf die Angabe der σ -Algebra; es ist dann stets die Potenzmenge $\mathfrak{P}(\Omega)$ auf Ω bzw. die Borel'sche σ -algebra $\mathbb{B}|_{\Omega}$ über Ω zugrunde zu legen.

Wir wollen nicht näher darauf eingehen, unter welchen Bedingungen eine Funktion messbar ist, da dies bei in Anwendungen auftretenden Funktionen i. d. R. der Fall ist. Man kann z. B. zeigen, dass jede (stückweise) stetige oder monotone Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ messbar (d. h. \mathbb{B} - \mathbb{B} -messbar) ist.

Wir wollen nun einige Überlegungen zur Bestimmung von induzierten Verteilungen anstellen. Dabei wollen wir vor allem den Fall betrachten, dass die Verteilung einer Zufallsgröße X mit Werten in $\mathcal{X} \subseteq \mathbb{R}$ gegeben ist und die Verteilung einer daraus entstehenden Zufallsgröße $Y = h(X)$ gesucht ist, wobei $h : \mathcal{X} \rightarrow \mathbb{R}$ messbar ist.

Zunächst noch eine Vereinbarung: Ist X eine Zufallsgröße mit Werten in \mathbb{R} , so wird die Verteilungsfunktion F_X von \mathbb{P}_X auch als *Verteilungsfunktion von X* bezeichnet:

$$F_X : \mathbb{R} \rightarrow \mathbb{R} \quad F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \in]-\infty, x]) =: \mathbb{P}(X \leq x).$$

Es sei daran erinnert, dass \mathbb{P}_X durch F_X eindeutig bestimmt ist. Entsprechend wird (im Falle der Existenz!) die Riemann-Dichte f_X von \mathbb{P}_X auch als *Riemann-Dichte von X* (oder auch nur als *Dichte von X*) bezeichnet.

Bemerkung 2.21 (Bestimmung von Verteilungen von Zufallsgrößen).

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum, X eine Zufallsgröße mit Werten in $\mathcal{X} \subseteq \mathbb{R}$ und $h : \mathcal{X} \rightarrow \mathbb{R}$ eine messbare Funktion. Dann gibt es die folgenden Ansätze zur Bestimmung der Verteilung von $Y := h(X)$:

(a) (*über Elementar-W.'en*)

Gibt es eine abzählbare Menge $\mathcal{Y} \subseteq \mathbb{R}$ mit $\mathbb{P}(Y \in \mathcal{Y}) = 1$, so ist \mathbb{P}_Y ein diskretes W.maß und kann am einfachsten über seine Elementar-W.en bestimmt werden. Dies ist insbesondere (aber nicht nur) der Fall, wenn \mathbb{P}_X ein diskretes W.maß ist.

(b) (*über Verteilungsfunktion*)

Allgemein kann man versuchen, die Verteilung \mathbb{P}_Y über die Verteilungsfunktion F_Y zu bestimmen, indem man $\mathbb{P}(Y \leq y)$ berechnet. Dies bietet sich vor allem an, wenn Y alle Werte in einem (echten) Intervall annimmt.

Ist die Verteilungsfunktion F_Y stetig und (stückweise) stetig differenzierbar, so besitzt \mathbb{P}_Y die Riemann-Dichte $f_Y := F'_Y$, vgl. dazu Bemerkung 2.17 (b). (Insbesondere besitzt Y also überhaupt eine Riemann-Dichte!)

(c) (*über Dichte*)⁵

Ist \mathbb{P}_X ein stetiges W.maß mit einer stetigen Dichte $f_X : \mathcal{X} \rightarrow \mathbb{R}$ auf einem Intervall $\mathcal{X} \subseteq \mathbb{R}$ und ist $h : \mathcal{X} \rightarrow \mathbb{R}$ eine streng monotone stetig differenzierbare Funktion mit $h'(x) \neq 0$ für alle $x \in \mathcal{X}$, so besitzt \mathbb{P}_Y die Dichte $f_Y : \mathcal{Y} \rightarrow \mathbb{R}$ auf dem Intervall $\mathcal{Y} := h(\mathcal{X})$ mit

$$f_Y(y) = f_X(h^{-1}(y)) |(h^{-1})'(y)| = f_X(h^{-1}(y)) \frac{1}{|h'(h^{-1}(y))|} \quad (y \in \mathcal{Y}).$$

Diese Formel ist auch als *Transformationsformel für Dichten* bekannt.

denn: Sind F_X und F_Y die Verteilungsfunktionen von \mathbb{P}_X und \mathbb{P}_Y , so gilt für den Fall, dass h streng monoton wachsend ist, für alle $y \in \mathcal{Y}$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq h^{-1}(y)) = F_X(h^{-1}(y)),$$

woraus die Behauptung durch Differentiation nach y (nach den Ableitungsregeln!) folgt. Der Fall, dass h streng monoton fallend ist, lässt sich ähnlich behandeln.

Ist h nur stückweise streng monoton, so kann man \mathcal{X} in mehrere Abschnitte zerlegen, auf denen h streng monoton ist, die o. a. Transformationsformel auf die einzelnen Abschnitte anwenden und die Ergebnisse addieren, vgl. das folgende Beispiel 2.22 (c).

⁵ Dieser Punkt ist in der Vorlesung übersprungen worden.

Beispiel 2.22 (Bestimmung von Verteilungen von Zufallsgrößen).

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und X eine Zufallsgröße mit Werten in \mathbb{R} .

(a) Ist $\mathbb{P}_X = \mathcal{E}(\lambda)$ und $Y := \lceil X \rceil$, so gilt $\mathbb{P}_Y = \mathcal{G}(1 - e^{-\lambda})$.

(b) Ist $\mathbb{P}_X = \mathcal{N}(0, 1)$ und $Y := \mu + \sigma X$ (wobei $\sigma \neq 0$), so gilt $\mathbb{P}_Y = \mathcal{N}(\mu, \sigma^2)$.

Allgemeiner gilt: Ist $\mathbb{P}_X = \mathcal{N}(\mu, \sigma^2)$ und $Y := aX + b$ (wobei $a \neq 0$), so gilt $\mathbb{P}_Y = \mathcal{N}(a\mu + b, a^2\sigma^2)$.

(c) Ist $\mathbb{P}_X = \mathcal{N}(0, 1)$ und $Y := X^2$, so gilt $\mathbb{P}_Y = \chi_1^2$.

Dabei bezeichnet χ_1^2 das stetige W.maß zur Dichte $f_1(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2} \mathbf{1}_{(0, \infty)}(x)$; χ_1^2 heißt *Chi-Quadrat-Verteilung zum Parameter 1*.

Beweis.

(a) Hier nimmt Y m. W. 1 Werte in \mathbb{N} an, so dass wir die Elementar-W.en bestimmen:
Für jedes $k \in \mathbb{N}$ gilt

$$\mathbb{P}(Y = k) = \mathbb{P}(k-1 < X \leq k) = \int_{k-1}^k \lambda e^{-\lambda t} dt = e^{-\lambda(k-1)} - e^{-\lambda k} = (e^{-\lambda})^{k-1} (1 - e^{-\lambda}).$$

Dies sind gerade die Elementar-W.en der geom. Verteilung zum Parameter $1 - e^{-\lambda}$.
 \Rightarrow Beh.

(b) Hier nimmt Y m. W. 1 Werte in \mathbb{R} an, so dass wir F_Y auf \mathbb{R} bestimmen.
Im Fall $\sigma > 0$ gilt

$$F_Y(y) = \mathbb{P}(\mu + \sigma X \leq y) \stackrel{\sigma > 0}{=} \mathbb{P}(X \leq \frac{y-\mu}{\sigma}) = \Phi(\frac{y-\mu}{\sigma}) \quad (y \in \mathbb{R})$$

und im Fall $\sigma < 0$ gilt

$$F_Y(y) = \mathbb{P}(\mu + \sigma X \leq y) \stackrel{\sigma < 0}{=} \mathbb{P}(X \geq \frac{y-\mu}{\sigma}) = 1 - \Phi(\frac{y-\mu}{\sigma}) \quad (y \in \mathbb{R}),$$

wobei $\Phi(x) := \int_{-\infty}^x \varphi(t) dt$ (mit $\varphi(t) := \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$) die Verteilungsfunktion der $\mathcal{N}(0, 1)$ -Verteilung bezeichnet (die sich leider nicht elementar angeben lässt).
Da F_Y stetig differenzierbar ist, besitzt \mathbb{P}_Y damit die Riemann-Dichte $f_Y = F'_Y$, also im Fall $\sigma > 0$

$$f_Y(y) = F'_Y(y) = \Phi'(\frac{y-\mu}{\sigma}) \frac{1}{\sigma} = \varphi(\frac{y-\mu}{\sigma}) \frac{1}{|\sigma|} = \varphi_{\mu, \sigma^2}(y) \quad (y \in \mathbb{R})$$

und im Fall $\sigma < 0$

$$f_Y(y) = F'_Y(y) = -\Phi'(\frac{y-\mu}{\sigma}) \frac{1}{\sigma} = \varphi(\frac{y-\mu}{\sigma}) \frac{1}{|\sigma|} = \varphi_{\mu, \sigma^2}(y) \quad (y \in \mathbb{R}),$$

wobei φ_{μ, σ^2} (in beiden Fällen!) die Dichte der $\mathcal{N}(\mu, \sigma^2)$ -Verteilung bezeichnet.
 \Rightarrow Beh.

Alternativ kann man hier auch die Transformationsformel für Dichten verwenden; mit $h(x) := \mu + \sigma x$ gilt $h^{-1}(y) = \frac{y-\mu}{\sigma}$ und damit

$$f_Y(y) = f_X(h^{-1}(y)) |(h^{-1})'(y)| = \varphi\left(\frac{y-\mu}{\sigma}\right) \frac{1}{|\sigma|} = \varphi_{\mu, \sigma^2}(y)$$

für alle $y \in \mathbb{R}$. \Rightarrow Beh.

Der allgemeine Fall lässt sich auf den bereits behandelten Spezialfall zurückführen: Nach Lemma 2.20 können wir o. E. annehmen, dass $X = \mu + \sigma Z$ mit $Z \sim \mathcal{N}(0, 1)$, und dann folgt $aX + b = a(\mu + \sigma Z) + b = (a\mu + b) + (a\sigma)Z \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

- (c) Hier nimmt Y m. W. 1 Werte in $(0, \infty)$ an, so dass wir F_Y auf $(0, \infty)$ bestimmen. Für alle $y > 0$ gilt

$$F_Y(y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(0 \leq X \leq +\sqrt{y}) + \mathbb{P}(-\sqrt{y} \leq X \leq 0) \underset{\text{Symmetrie}}{=} 2(\Phi(\sqrt{y}) - \Phi(0)).$$

Da F_Y auf $(0, \infty)$ stetig differenzierbar ist, besitzt \mathbb{P}_Y auf $(0, \infty)$ die Riemann-Dichte $f_Y(y) = F_Y'(y) = 2\varphi(\sqrt{y}) \frac{1}{2\sqrt{y}}$. \Rightarrow Beh.

Alternativ kann man hier wieder die Transformationsformel für Dichten verwenden; mit $h(x) := x^2$, $h_1 := h|_{(0, +\infty)}$, $h_2 := h|_{(-\infty, 0)}$ gilt $h_1^{-1}(y) = +\sqrt{y}$, $h_2^{-1}(y) = -\sqrt{y}$ und damit

$$f_Y(y) = f_X(h_1^{-1}(y)) |(h_1^{-1})'(y)| + f_X(h_2^{-1}(y)) |(h_2^{-1})'(y)| \underset{\text{Symmetrie}}{=} 2\varphi(\sqrt{y}) \frac{1}{2\sqrt{y}}$$

für alle $y > 0$. \Rightarrow Beh.

□

Häufig tritt auch die Situation auf, dass die Verteilung einer „zusammengesetzten“ Zufallsgröße $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ gegeben ist und die Verteilungen von X und von Y (die sog. *Randverteilungen*) gesucht sind. Besitzt (X, Y) eine diskrete Verteilung, kann man wie in Lemma 1.19 vorgehen (auch wenn der W.raum selbst nicht diskret ist). Besitzt (X, Y) eine stetige Verteilung, so gilt:

Lemma 2.23 (Bestimmung von Randverteilungen). *Ist $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ eine Zufallsgröße mit einer zweidimensionalen Riemann-Dichte $f_{(X, Y)}$, so besitzt X die eindimensionale Riemann-Dichte f_X mit*

$$f_X(x) = \int f_{(X, Y)}(x, y) dy$$

für alle $x \in \mathbb{R}$. (Insbesondere besitzt X also überhaupt eine Riemann-Dichte!)

Beweis. Für jedes Intervall $I \subseteq \mathbb{R}$ gilt

$$\begin{aligned} \mathbb{P}(X \in I) &= \mathbb{P}((X, Y) \in I \times \mathbb{R}) \overset{\text{gemeinsame Dichte}}{=} \int_{I \times \mathbb{R}} f_{(X, Y)}(x, y) d(x, y) \\ &\overset{\text{iteriertes Integral}}{=} \int_I \int_{\mathbb{R}} f_{(X, Y)}(x, y) dy dx \overset{\text{Definition von } f_X}{=} \int_I f_X(x) dx. \end{aligned}$$

□

Eine analoge Aussage gilt natürlich für Y statt X .

Allgemeiner kann man den Fall betrachten, dass die n -dimensionale Riemann-Dichte einer Zufallsgröße (X_1, \dots, X_n) mit Werten in \mathbb{R}^n gegeben ist und eine *Randverteilung* \mathbb{P}_{X_i} ($i = 1, \dots, n$) bzw. sogar eine *mehrdimensionale Randverteilung* $\mathbb{P}_{(X_{i_1}, \dots, X_{i_m})}$ ($1 \leq i_1 < \dots < i_m \leq n$) gesucht ist. Hier gilt:

Merkregel:

Ist die W.dichte einer zusammengesetzten Zufallsgröße (X_1, \dots, X_n) mit Werten in \mathbb{R}^n gegeben, so erhalten wir die W.dichten der Randverteilungen, indem wir über die „freien“ Variablen integrieren.

Beispiel 2.24 (Bestimmung von Randverteilungen). Die Zufallsgröße (X, Y) sei gleichverteilt auf der Menge

$$\Omega := \{(x, y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\},$$

(wobei $a, b > 0$), d. h. die Zufallsgröße besitze die zweidimensionale Riemann-Dichte

$$f_{(X,Y)}(x, y) = \frac{1}{\text{vol}(\Omega)} \mathbf{1}_\Omega(x, y) = \frac{1}{\pi ab} \mathbf{1}_{\{\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\}}.$$

(Dabei haben wir die Formel für den Flächeninhalt der Ellipse benutzt.) Dann sind die Verteilungen von X und Y durch die eindimensionalen Riemann-Dichten

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \begin{cases} = \int_{-b\sqrt{1-(x^2/a^2)}}^{+b\sqrt{1-(x^2/a^2)}} \frac{1}{\pi ab} dy = \frac{2}{\pi a^2} \sqrt{a^2 - x^2} & ; |x| \leq a \\ = 0 & ; |x| > a \end{cases}$$

und

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx \begin{cases} = \int_{-a\sqrt{1-(y^2/b^2)}}^{+a\sqrt{1-(y^2/b^2)}} \frac{1}{\pi ab} dx = \frac{2}{\pi b^2} \sqrt{b^2 - y^2} & ; |y| \leq b \\ = 0 & ; |y| > b \end{cases}$$

gegeben (*Halbkreisverteilung*). □

Anwendung: Simulation von Verteilungen.

Bei der *Simulation* versucht man, das Verhalten realer Systeme anhand eines Modells (etwa auf dem Computer) nachzuvollziehen. Bei der Simulation zufälliger Systeme ist man oft darauf angewiesen, Realisierungen von Zufallsgrößen mit vorgegebenen Verteilungen zu erzeugen. (Man spricht dann auch von der *Simulation von Verteilungen*.) In vielen Programmiersprachen gibt es eine Routine **rand**, die unabhängige $\mathcal{U}_{(0,1)}$ -verteilte (Pseudo-)Zufallszahlen liefert. Um von hier zu anderen Verteilungen zu gelangen, hilft manchmal die folgende Konstruktion weiter:

Satz 2.25 (Inversionsmethode). *Ist U eine $\mathcal{U}_{(0,1)}$ -verteilte Zufallsgröße, ist F eine Verteilungsfunktion und ist $G(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$ die sog. Pseudo-Inverse von F , so ist $\mathbb{P}_{G \circ U}$ das W.maß zur Verteilungsfunktion F .*

Beweis. Für alle $x \in \mathbb{R}$ und alle $u \in (0, 1)$ gilt

$$G(u) \leq x \quad \Leftrightarrow \quad F(x) \geq u.$$

„ \Leftarrow “ ist offensichtlich, und \Rightarrow folgt daraus, dass im Fall $G(u) \leq x$ eine absteigende Folge $(x_n)_{n \in \mathbb{N}}$ mit $F(x_n) \geq u$ und $x_n \rightarrow G(u)$ existiert, und damit folgt (mit den Eigenschaften von Verteilungsfunktionen!) $F(x) \geq F(G(u)) = \lim_{n \rightarrow \infty} F(x_n) \geq u$.

Wir bestimmen nun die Verteilungsfunktion von $G \circ U$: Für alle $x \in \mathbb{R}$ gilt

$$F_{G \circ U}(x) = \mathbb{P}(G(U) \leq x) \stackrel{\text{Vorüberlegung}}{=} \mathbb{P}(U \leq F(x)) = F_U(F(x)) \stackrel{\text{Übung}}{=} F(x).$$

Da ein W.maß durch seine Verteilungsfunktion eindeutig bestimmt ist, ist $\mathbb{P}_{G \circ U}$ damit das W.maß zur Verteilungsfunktion F . \square

Zusammenfassung:

- Zufallsabhängige Vorgänge, bei denen die Menge der möglichen Ergebnisse eine kontinuierliche Menge wie (z. B. ein Intervall) ist, lassen sich häufig durch stetige W.räume beschreiben.
- Wir geben stetige W.räume an, indem wir die Menge Ω sowie die Riemann-Dichte $f : \Omega \rightarrow \mathbb{R}$ angeben.
- Die W.en von einfachen Mengen (wie z. B. Intervallen) lassen sich berechnen, indem man die Riemann-Dichte f über diese Mengen integriert.
- Wir haben eine Reihe wichtiger stetiger W.verteilungen kennengelernt (Gleichverteilung, Exponentialverteilung, Normalverteilung).
- Eine Zufallsgröße auf einem allgemeinen W.raum ist eine (messbare) Abbildung $X : \Omega \rightarrow \mathcal{X}$.
- Meistens interessiert uns vor allem die Verteilung \mathbb{P}_X einer Zufallsgröße X ; diese lässt sich für eine reellwertige Zufallsgröße X z. B. über die Verteilungsfunktion $F_X(x) = \mathbb{P}(X \leq x)$ beschreiben.

3 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

3.1 Bedingte Wahrscheinlichkeiten

Häufig stehen gewisse Teilinformationen über den Ausgang eines Zufallsexperimentes zur Verfügung.

Beispiel 3.1 (Würfel). Ein gezinkter Würfel liefert die Ergebnisse $\omega \in \Omega = \{1, 2, 3, 4, 5, 6\}$ mit den folgenden (Elementar-)Wahrscheinlichkeiten:

ω	1	2	3	4	5	6
$f(\omega)$	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{10}{25}$

Wir würfeln, ohne das Ergebnis sehen zu können, und erhalten die Information, dass wir keine Sechs gewürfelt haben. Mit welcher Wahrscheinlichkeit ist das Ergebnis dann eine gerade Zahl?

Sei $B = \{1, 2, 3, 4, 5\}$ das Ereignis „keine Sechs“ und $A = \{2, 4, 6\}$ das Ereignis „gerade Zahl“. Wir interessieren uns dann für die W. von A unter der *Bedingung*, dass B eintritt. Um diese W. zu bestimmen, orientieren wir uns an relativen Häufigkeiten. Wenn wir das Zufallsexperiment n -mal wiederholen und nur noch diejenigen Durchführungen betrachten, bei denen B eintritt, so ist die *bedingte* relative Häufigkeit von A

$$\frac{H_n(A \cap B)}{H_n(B)} = \frac{\frac{1}{n} H_n(A \cap B)}{\frac{1}{n} H_n(B)} \approx \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{2, 4\})}{\mathbb{P}(\{6\}^c)} = \frac{\frac{3}{25} + \frac{4}{25}}{1 - \frac{10}{25}} = \frac{7}{15};$$

die *bedingte* W. von A sollte also

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{7}{15}$$

sein. □

Wir definieren allgemein:

Definition 3.2 (Bedingte Wahrscheinlichkeit).

Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und ist $B \in \mathcal{A}$ ein Ereignis mit $\mathbb{P}(B) > 0$, so heißt für jedes Ereignis $A \in \mathcal{A}$

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

(elementare) *bedingte Wahrscheinlichkeit von A unter B / gegeben B .*

Anschaulich beschreibt $\mathbb{P}(A | B)$ die Chance des Eintretens von A , wenn schon feststeht, dass B eintritt.

Beispiel 3.3 (Kugelziehung). Aus einer Urne mit r roten und s schwarzen Kugeln wird $2 \times$ nacheinander gezogen. Wie groß ist die W., dass in der 2. Ziehung „rot“ gezogen wird, wenn in der 1. Ziehung „schwarz“ gezogen wird?

Ziehen mit Zurücklegen:

Wir wählen $\Omega := \{1, \dots, r+s\}^2$ und $\mathbb{P} := \mathcal{U}_\Omega$ (wegen Symmetrie auf der Ebene der Kugeln), wobei $1, \dots, r$ die roten Kugeln und $r+1, \dots, r+s$ die schwarzen Kugeln seien, und $A := \{\omega \in \Omega : \omega_2 \leq r\}$, $B := \{\omega \in \Omega : \omega_1 > r\}$. Dann gilt $\mathbb{P}(B) = \frac{s}{r+s}$ und $\mathbb{P}(A \cap B) = \frac{rs}{(r+s)^2}$ und damit

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{rs}{(r+s)^2}}{\frac{s}{r+s}} = \frac{r}{r+s}$$

(wie anschaulich zu erwarten).

Ziehen ohne Zurücklegen:

Wir wählen $\Omega := \{(\omega_1, \omega_2) \in \{1, \dots, r+s\}^2 : \omega_1 \neq \omega_2\}$ und $\mathbb{P} := \mathcal{U}_\Omega$ (wegen Symmetrie auf der Ebene der Kugeln), wobei $1, \dots, r$ die roten Kugeln und $r+1, \dots, r+s$ die schwarzen Kugeln seien, und $A := \{\omega \in \Omega : \omega_2 \leq r\}$, $B := \{\omega \in \Omega : \omega_1 > r\}$. Dann gilt $\mathbb{P}(B) = \frac{s}{r+s}$ und $\mathbb{P}(A \cap B) = \frac{rs}{(r+s)(r+s-1)}$ und damit

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{rs}{(r+s)(r+s-1)}}{\frac{s}{r+s}} = \frac{r}{r+s-1}$$

(wie anschaulich zu erwarten). □

In diesem Beispiel liegt ein zweistufiges Experiment vor, bei dem sich das Ereignis B auf die erste Stufe und das Ereignis A auf die zweite Stufe bezieht. Dies ist keinesfalls notwendig – wir können $\mathbb{P}(A|B)$ für beliebige Ereignisse $A, B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ betrachten, ohne dass irgendein besonderer Zusammenhang zwischen den Ereignissen bestehen muss.

Lemma 3.4. Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und ist $B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ fest, so ist die Abbildung

$$\mathbb{P}(\cdot | B) : \mathcal{A} \rightarrow \mathbb{R} \quad A \mapsto \mathbb{P}(A|B)$$

ein W.maß auf \mathcal{A} mit $\mathbb{P}(B|B) = 1$ und $\mathbb{P}(B^c|B) = 0$.

Beweis. Man prüft nach, dass $\mathbb{P}(\cdot | B)$ nicht-negativ, normiert und σ -additiv ist. □

Bemerkung. Damit bleiben alle Rechenregeln auch für bedingte W.en gültig.

Ist $\Omega \neq \emptyset$ beliebig, so heißt eine Familie $(B_i)_{i \in I}$ von nicht-leeren Teilmengen von Ω *Zerlegung* oder *Partition* von Ω , falls gilt:

- $\bigcup_{i \in I} B_i = \Omega$.
- Für alle $i, j \in I$ mit $i \neq j$ gilt $B_i \cap B_j = \emptyset$.

Eine Zerlegung $(B_i)_{i \in I}$ von Ω heißt *abzählbar*, falls die Indexmenge I abzählbar ist. Wir werden im Folgenden – zu einem gegebenen W.raum $(\Omega, \mathcal{A}, \mathbb{P})$ – nur Zerlegungen mit $B_i \in \mathcal{A}$ für alle $i \in I$ betrachten; wir schreiben dafür kurz $(B_i)_{i \in I} \subseteq \mathcal{A}$.

Satz 3.5 (Satz von der totalen W.). *Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und $(B_i)_{i \in I} \subseteq \mathcal{A}$ eine abzählbare Zerlegung von Ω mit $\mathbb{P}(B_i) > 0$ für alle $i \in I$, so gilt für jedes $A \in \mathcal{A}$*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i).$$

Beweis.

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap \bigcup_{i \in I} B_i) \stackrel{\text{Distr.gegesetz}}{=} \mathbb{P}\left(\bigcup_{i \in I} (A \cap B_i)\right) \\ &\stackrel{\sigma\text{-Additivität}}{=} \sum_{i \in I} \mathbb{P}(A \cap B_i) \stackrel{\text{Definition}}{=} \sum_{i \in I} \mathbb{P}(A|B_i) \mathbb{P}(B_i). \end{aligned}$$

□

Satz 3.6 (Formel von Bayes). *Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und $(B_i)_{i \in I} \subseteq \mathcal{A}$ eine abzählbare Zerlegung von Ω mit $\mathbb{P}(B_i) > 0$ für alle $i \in I$, so gilt für alle $i \in I$ und alle $A \in \mathcal{A}$ mit $\mathbb{P}(A) > 0$*

$$\mathbb{P}(B_i | A) = \frac{\mathbb{P}(A | B_i) \mathbb{P}(B_i)}{\sum_{j \in I} \mathbb{P}(A | B_j) \mathbb{P}(B_j)}.$$

Beweis.

$$\mathbb{P}(B_i | A) \stackrel{\text{Definition}}{=} \frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)} \stackrel{\text{Definition}}{=} \frac{\mathbb{P}(A | B_i) \mathbb{P}(B_i)}{\mathbb{P}(A)} \stackrel{\text{totale W.}}{=} \frac{\mathbb{P}(A | B_i) \mathbb{P}(B_i)}{\sum_{j \in I} \mathbb{P}(A | B_j) \mathbb{P}(B_j)}.$$

□

Satz 3.7 (Multiplikationsregel).

Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und sind $A_1, \dots, A_n \in \mathcal{A}$ mit $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$, so gilt

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

Beweis. Wegen $\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0$ gilt $\mathbb{P}(A_1 \cap \dots \cap A_i) > 0$ für alle $i = 1, \dots, n-1$ (wegen der Monotonie von \mathbb{P}). Damit ist die rechte Seite gleich

$$\mathbb{P}(A_1) \cdot \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \cdot \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} \cdot \dots \cdot \frac{\mathbb{P}(A_1 \cap \dots \cap A_{n-1} \cap A_n)}{\mathbb{P}(A_1 \cap \dots \cap A_{n-1})},$$

woraus sich durch Kürzen die linke Seite ergibt.

□

Die Multiplikationsregel erlaubt es, in einem gegebenen W.raum W.en als Produkte von bedingten W.en zu berechnen:

Beispiel 3.8 (Skat). Mit welcher W. erhält beim Skat jeder der 3 Spieler genau 1 Buben?

Lösung: Wir zerlegen das Gesamtexperiment (gedanklich) in drei Teilexperimente, wobei im i -ten Schritt die Karten für den i -ten Spieler ausgewählt werden. Es bezeichne B_i das Ereignis, dass der i -te Spieler genau 1 Buben erhält. Dann ist anschaulich klar, dass

$$\mathbb{P}(B_1) = \frac{\binom{28}{9}\binom{4}{1}}{\binom{32}{10}}, \quad \mathbb{P}(B_2 | B_1) = \frac{\binom{19}{9}\binom{3}{1}}{\binom{22}{10}}, \quad \mathbb{P}(B_3 | B_1 \cap B_2) = \frac{\binom{10}{9}\binom{2}{1}}{\binom{12}{10}},$$

gilt, so dass mit der Multiplikationsregel

$$\mathbb{P}(B_1 \cap B_2 \cap B_3) = \mathbb{P}(B_1) \mathbb{P}(B_2 | B_1) \mathbb{P}(B_3 | B_1 \cap B_2) \approx 5,56\%$$

folgt. □

Hier sind wir anders als in Beispiel 3.3 nicht von den W.en für das Gesamtexperiment ausgegangen, sondern wir haben mit Hilfe unserer Anschauung die (bedingten) W.en für die Teilexperimente bestimmt.

In der Tat lässt sich die *Multiplikationsregel* auch zur *Konstruktion* von W.räumen zur Beschreibung mehrstufiger Zufallsexperimente verwenden; sie ist in diesem Zshg. auch als *Pfadregel* bekannt. Wir beginnen mit zwei Beispielen:

Beispiel 3.9 (Ziegenproblem). Es gibt 3 Türen, hinter denen der Moderator 1 Auto sowie 2 Ziegen versteckt hat. Der Spieler wählt eine Tür. Der Moderator öffnet eine andere Tür, hinter der sich eine Ziege befindet, und bietet dem Spieler an, die Tür zu wechseln. Soll der Spieler darauf eingehen?

Wir wollen annehmen, dass

- der Moderator das Auto „rein zufällig“ versteckt (um es dem Spieler besonders schwer zu machen),
- der Spieler (aus Symmetriegründen) immer die 1. Tür wählt,
- der Moderator (sofern die Möglichkeit besteht) „rein zufällig“ entscheidet, welche Tür er öffnet.

Wir betrachten die Ereignisse

$$A_i : \text{„das Auto steht hinter Tür Nr. } i\text{“} \quad (i = 1, 2, 3),$$

und

$$T_j : \text{„der Moderator öffnet Tür Nr. } j\text{“} \quad (j = 2, 3).$$

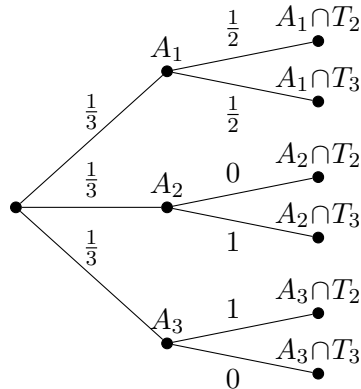
Laut Problemstellung gilt dann

$$\mathbb{P}(A_i) = \frac{1}{3} \quad \text{für } i = 1, 2, 3$$

und

$$\mathbb{P}(T_j | A_i) = \begin{cases} \frac{1}{2} & \text{falls } i = 1 \text{ und } j > 1, \\ 1 & \text{falls } i > 1 \text{ und } j \neq i, \\ 0 & \text{falls } i > 1 \text{ und } j = i, \end{cases} \quad \text{für } i = 1, 2, 3 \text{ und } j = 2, 3.$$

Diese Informationen lassen sich schön in einem *Wahrscheinlichkeitsbaum* darstellen:



Die gesuchte (bedingte) W. ist nun (für $j = 2, 3$) durch

$$\mathbb{P}(A_1 | T_j) \underset{\text{Bayes}}{=} \frac{\mathbb{P}(T_j | A_1) \mathbb{P}(A_1)}{\sum_{i=1}^3 \mathbb{P}(T_j | A_i) \mathbb{P}(A_i)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{1}{3}$$

gegeben. Im Mittel lohnt es sich also, zu wechseln! □

Die Formel von Bayes wird häufig bei der Untersuchung von Ursache-Wirkungs-Beziehungen verwendet:

Beispiel 3.10 (Medizinischer Test). Eine Krankheit kommt bei 1 % der Bevölkerung vor. Um zu untersuchen, ob eine Person von dieser Krankheit betroffen ist, wird ein Schnelltest verwendet:

- Ist die Person krank, so ist der Test m.W. 98 % positiv. (Sensitivität 98%)
- Ist die Person gesund, so ist der Test m.W. 4 % positiv. (Spezifität 96%)

Der Schnelltest wird im Rahmen einer Routine-Untersuchung bei einer Person durchgeführt.

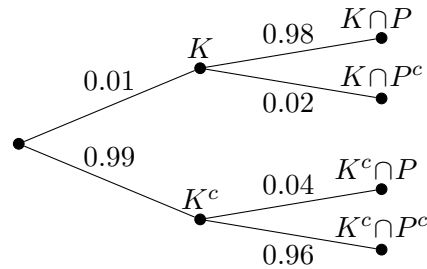
Wir führen die Ereignisse K : „die Person ist krank“ und P : „der Test ist positiv“ ein. Es gilt dann

$$\mathbb{P}(K) = 0.01, \quad \mathbb{P}(P | K) = 0.98, \quad \mathbb{P}(P | K^c) = 0.04,$$

also auch (nach der Formel für die Gegen-W.)

$$\mathbb{P}(K^c) = 0.99, \quad \mathbb{P}(P^c | K) = 0.02, \quad \mathbb{P}(P^c | K^c) = 0.96.$$

Diese Informationen lassen sich wieder in einem *Wahrscheinlichkeitsbaum* darstellen:



1. Mit welcher W. ist das Testergebnis positiv?

$$\mathbb{P}(P) \underset{\text{totale W.}}{=} \mathbb{P}(P | K)\mathbb{P}(K) + \mathbb{P}(P | K^c)\mathbb{P}(K^c) = 0.98 \cdot 0.01 + 0.04 \cdot 0.99 = 0.0494.$$

2. Mit welcher W. liegt die Krankheit vor, wenn das Testergebnis positiv ist?

$$\mathbb{P}(K | P) \underset{\text{Bayes}}{=} \frac{\mathbb{P}(P | K)\mathbb{P}(K)}{\mathbb{P}(P | K)\mathbb{P}(K) + \mathbb{P}(P | K^c)\mathbb{P}(K^c)} = \frac{0.98 \cdot 0.01}{0.98 \cdot 0.01 + 0.04 \cdot 0.99} \approx 0.1984.$$

oder (wenn wir berücksichtigen, was wir schon berechnet haben)

$$\mathbb{P}(K | P) \underset{\text{Def.}}{=} \frac{\mathbb{P}(P \cap K)}{\mathbb{P}(P)} \underset{\text{Def.}}{=} \frac{\mathbb{P}(P | K)\mathbb{P}(K)}{\mathbb{P}(P)} = \frac{0.98 \cdot 0.01}{0.0494} \approx 0.1984.$$

3. Mit welcher W. liegt die Krankheit vor, wenn das Testergebnis negativ ist?

$$\mathbb{P}(K | P^c) \underset{\text{Bayes}}{=} \frac{\mathbb{P}(P^c | K)\mathbb{P}(K)}{\mathbb{P}(P^c | K)\mathbb{P}(K) + \mathbb{P}(P^c | K^c)\mathbb{P}(K^c)} = \frac{0.02 \cdot 0.01}{0.02 \cdot 0.01 + 0.96 \cdot 0.99} \approx 0.0002104.$$

oder (wenn wir berücksichtigen, was wir schon berechnet haben)

$$\mathbb{P}(K | P^c) \underset{\text{Def.}}{=} \frac{\mathbb{P}(P^c \cap K)}{\mathbb{P}(P^c)} \underset{\text{Def.}}{=} \frac{\mathbb{P}(P^c | K)\mathbb{P}(K)}{\mathbb{P}(P^c)} \underset{\text{Gegen-W.}}{=} \frac{0.02 \cdot 0.01}{1 - 0.0494} \approx 0.0002104.$$

Fazit: Bei einem negativen Testergebnis können wir also „sehr sicher“ davon ausgehen, dass die Person nicht von der Krankheit betroffen ist. Bei einem positiven Testergebnis ist die Situation nicht so eindeutig; hier sollte die Person eingehender untersucht werden. Dies ist aber nur bei ca. 5 % der getesteten Personen erforderlich, so dass der Schnelltest insgesamt als „gut“ einzustufen ist. \square

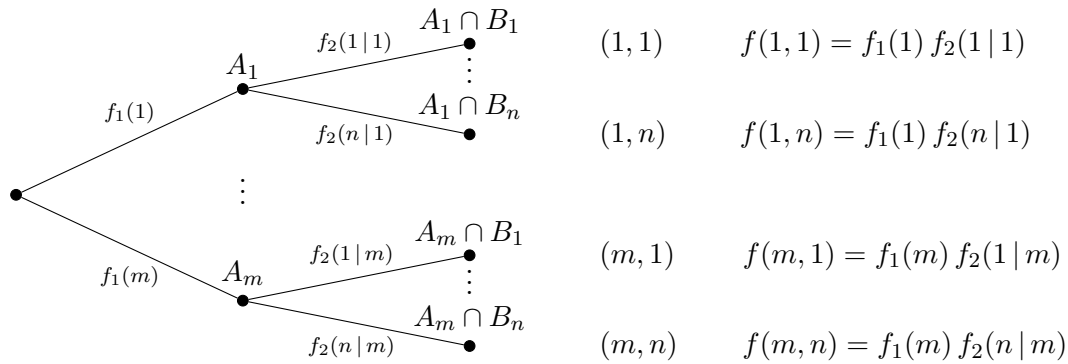
Im Allgemeinen sind Rückschlüsse von bedingten W.en auf kausale Zusammenhänge allerdings unzulässig.

Streng genommen stellt sich natürlich die Frage, ob es in Beispiel 3.9 und Beispiel 3.10 überhaupt einen W.raum $(\Omega, \mathcal{A}, \mathbb{P})$ gibt, der zu der angegebenen Beschreibung passt.

Angenommen, wir haben ein zweistufiges Zufallsexperiment, bei dem

- in der ersten Stufe genau eines der Ereignisse A_1, \dots, A_m auftritt, und zwar mit der W. $f_1(\omega_1)$ ($\omega_1 = 1, \dots, m$),
- in der zweiten Stufe genau eines der Ereignisse B_1, \dots, B_n auftritt, und zwar mit der bedingten W. $f_2(\omega_2 | \omega_1)$ ($\omega_1 = 1, \dots, m, \omega_2 = 1, \dots, n$).

Dann lässt sich das Zufallsexperiment durch die Grundmenge $\Omega = \{1, \dots, m\} \times \{1, \dots, n\}$ und das W.maß \mathbb{P} zu den Elementar-W.en $f(\omega) = f_1(\omega_1) f_2(\omega_2 | \omega_1)$ ($\omega = (\omega_1, \omega_2) \in \Omega$) beschreiben. Diese Konstruktion lässt sich wieder mit Hilfe von *Wahrscheinlichkeitsbäumen* veranschaulichen, wobei wir dieses Mal zusätzlich die möglichen Ergebnisse und die zugehörigen Elementar-W.en angeben:



Genauer besagt die *Pfadregel* hier, dass man die Elementar-W.en $f(\omega)$ für das Gesamtexperiment dadurch erhält, dass man die Gewichte entlang des entsprechenden Pfades durch den W.baum aufmultipliziert.

Setzt man $A_i := \{\omega \in \Omega | \omega_1 = i\}$ ($i = 1, \dots, m$), $B_j := \{\omega \in \Omega | \omega_2 = j\}$ ($j = 1, \dots, n$), so gilt im Fall $f_1(i) > 0$

$$\mathbb{P}(B_j | A_i) = \frac{\mathbb{P}(A_i \cap B_j)}{\mathbb{P}(A_i)} = \frac{f_1(i) f_2(j|i)}{\sum_{k=1}^n f_1(i) f_2(k|i)} = \frac{f_1(i) f_2(j|i)}{f_1(i)} = f_2(j|i),$$

d. h. die bei der Konstruktion eingesetzten bedingten W.en lassen sich in der Tat als bedingte W.en im Gesamtexperiment wiederfinden!

Ähnliches gilt natürlich für mehrstufige Zufallsexperimente mit mehr als zwei Stufen; hier sind in der k -ten Stufe die bedingten W.en $f_k(\omega_k | \omega_1, \dots, \omega_{k-1})$ vorzugeben.

Merkregel:

Sind bei einem mehrstufigen Zufallsexperiment, dessen Stufen sich durch abzählbare Mengen beschreiben lassen, für die erste Stufe die W.en sowie für die übrigen Stufen die bedingten W.en gegeben die Ergebnisse in den vorhergehenden Stufen gegeben, so existiert genau ein W.maß auf dem Gesamttraum, das zu dieser Beschreibung passt, und dieses ist durch die *Pfadregel* gegeben.

Ausblick.

Die obige Konstruktion bietet sich u. a. an, wenn man zufallsabhängige Entwicklungen im Zeitablauf modellieren will; dies ist für viele Anwendungen – auch in der Informatik und in den Ingenieurwissenschaften – von Interesse. Ist $\Omega = \Omega_1 \times \dots \times \Omega_n$ der Grundraum, so lassen sich die *Projektionen* $X_i : \Omega \rightarrow \Omega_i$ mit $X_i(\omega_1, \dots, \omega_n) = \omega_i$ ($i = 1, \dots, n$) häufig als Zustände des zugrunde liegenden Systems zu den einzelnen Zeitpunkten interpretieren. Die Werte $f_k(\omega_k | \omega_1, \dots, \omega_{k-1})$ werden hier auch als *Übergangs-W.en* bezeichnet, da sie die W.en für die Übergänge zwischen den Zuständen beschreiben.

Spezialfälle:

- Gilt für die Übergangs-W.en $f_k(\omega_k | \omega_1, \dots, \omega_{k-1}) = f_k(\omega_k | \omega_{k-1})$ für alle $k = 2, \dots, n$ und alle $(\omega_1, \dots, \omega_k) \in \Omega_1 \times \dots \times \Omega_k$ (d. h. hängen die Übergangs-W.en immer nur vom aktuellen Zustand ab), so heißt die Folge $(X_i)_{i=1, \dots, n}$ *Markoff-Kette*.

Gilt zudem $\Omega_1 = \dots = \Omega_n$ und $f_2 = \dots = f_n$, so heißt sie *homogene Markoff-Kette*.

- Gilt für die Übergangs-W.en $f_k(\omega_k | \omega_1, \dots, \omega_{k-1}) = f_k(\omega_k)$ für alle $k = 2, \dots, n$ und alle $(\omega_1, \dots, \omega_k) \in \Omega_1 \times \dots \times \Omega_k$ (d. h. hängen die Übergangs-W.en nicht von den bisherigen Zuständen ab), so heißen X_1, \dots, X_n *unabhängig*.

Gilt zudem $\Omega_1 = \dots = \Omega_n$ und $f_1 = \dots = f_n$, so heißen X_1, \dots, X_n *unabhängig und identisch verteilt*.

Mit dem Begriff der *Unabhängigkeit* werden wir uns in den nächsten beiden Abschnitten genauer befassen.

3.2 Unabhängigkeit von Ereignissen

Gilt $\mathbb{P}(A | B) = \mathbb{P}(A)$, so hat das Eintreten von B keinen Einfluss auf die W. von A . Man sagt dann auch, A sei von B „stochastisch unabhängig“. Beachte, dass

$$\mathbb{P}(A | B) = \mathbb{P}(A) \quad \Leftrightarrow \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Die zweite Gleichung bleibt auch im Fall $\mathbb{P}(B) = 0$ sinnvoll und zeigt darüber hinaus, dass die Eigenschaft „stochastisch unabhängig“ symmetrisch in A und B ist.

Dies führt zur folgenden Definition:

Definition 3.11 (Unabhängigkeit von Ereignissen). Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum.

- (i) Zwei Ereignisse $A, B \in \mathcal{A}$ heißen (*stochastisch*) *unabhängig unter \mathbb{P}* , falls

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

gilt.

- (ii) Eine Familie $(A_i)_{i \in I}$ von Ereignissen in \mathcal{A} heißt (*stochastisch*) *unabhängig unter \mathbb{P}* , falls für alle $m \in \mathbb{N}$ und alle paarweise verschiedenen Indizes $i_1, \dots, i_m \in I$

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_m}) = \mathbb{P}(A_{i_1}) \cdot \dots \cdot \mathbb{P}(A_{i_m})$$

gilt.

Statt zu sagen, dass die Familie $(A_i)_{i \in I}$ unabhängig unter \mathbb{P} ist, sagen wir etwas kürzer auch, dass die A_i ($i \in I$) unabhängig unter \mathbb{P} sind.

Anschaulich bedeutet Unabhängigkeit, dass wir aus dem Eintreten eines der Ereignisses (auf der Ebene der W.en) nichts Neues über das Eintreten der anderen Ereignisse lernen können.

Beispiel 3.12 (Kugelziehung; Fortsetzung von Beispiel 3.3). Aus einer Urne mit r roten und s schwarzen Kugeln wird $2 \times$ nacheinander gezogen; es interessieren die Ereignisse A : „2. Kugel rot“ und B : „1. Kugel schwarz“.

Ziehen mit Zurücklegen:

$$\mathbb{P}(A | B) = \frac{r}{r+s} = \frac{r}{r+s} = \mathbb{P}(A) \Rightarrow A \text{ und } B \text{ sind „stoch. unabhängig“}$$

Ziehen ohne Zurücklegen:

$$\mathbb{P}(A | B) = \frac{r}{r+s-1} > \frac{r}{r+s} = \mathbb{P}(A) \Rightarrow A \text{ und } B \text{ sind „stoch. abhängig“}$$

□

Bemerkungen 3.13.

- (a) Um die Unabhängigkeit von $n > 2$ Ereignissen $A_1, \dots, A_n \in \mathcal{A}$ nachzuweisen, ist es nicht ausreichend, $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$ für alle $i \neq j$ zu zeigen.

Gegenbeispiel: 2-facher Würfelwurf (mit einem fairen Würfel),

A : „1. Augenzahl gerade“,

B : „2. Augenzahl gerade“,

C : „Augensumme gerade“.

- (b) Um die Unabhängigkeit von $n > 2$ Ereignissen $A_1, \dots, A_n \in \mathcal{A}$ nachzuweisen, ist es nicht ausreichend, $\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \cdot \dots \cdot \mathbb{P}(A_n)$ zu zeigen.

Gegenbeispiel: 3-facher Münzwurf (mit einer fairen Münze),

A : „mind. 2 mal Kopf“,

B : „1. Ergebnis Kopf“,

C : „2. Ergebnis = 3. Ergebnis“.

- (c) Um z. B. die Unabhängigkeit von n Ereignissen nachzuweisen, sind im Allgemeinen $2^n - n - 1$ Gleichungen zu überprüfen. [...]
- (d) Teilfamilien von unabhängigen Familien von Ereignissen sind wieder unabhängig. (Um dies zu erreichen, ist die Unabhängigkeit einer Familie von Ereignissen gerade wie in Definition 3.11 (ii) definiert!)
- (e) Eine unendliche Familie von Ereignissen ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist.

3.3 Unabhängigkeit von Zufallsgrößen

Grob gesprochen heißt eine Familie von Zufallsgrößen $X_i : \Omega \rightarrow \mathcal{X}_i$ ($i \in I$) (*stochastisch unabhängig unter* \mathbb{P}), wenn für jede Wahl von Mengen $B_i \subseteq \mathcal{X}_i$ ($i \in I$) die Urbildmengen $X_i^{-1}(B_i)$ unabhängig unter \mathbb{P} sind. Die formale Definition ist wieder etwas schwieriger, weil wir im Allgemeinen nur Mengen aus den jeweiligen σ -Algebren betrachten dürfen.

Definition 3.14 (Unabhängigkeit von Zufallsgrößen). Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum.

- (i) Zwei Zufallsgrößen $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$, $Y : (\Omega, \mathcal{A}) \rightarrow (\mathcal{Y}, \mathcal{C})$ heißen (*stochastisch unabhängig unter* \mathbb{P}), wenn für beliebige Mengen $B \in \mathcal{B}$ und $C \in \mathcal{C}$ die Ereignisse $X^{-1}(B)$ und $Y^{-1}(C)$ unabhängig unter \mathbb{P} sind, also

$$\mathbb{P}(\{X \in B\} \cap \{Y \in C\}) = \mathbb{P}(X \in B) \mathbb{P}(Y \in C)$$

gilt.

- (ii) Eine Familie von Zufallsgrößen $X_i : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}_i, \mathcal{B}_i)$ ($i \in I$) heißt (*stochastisch unabhängig unter* \mathbb{P}), wenn für jede Wahl von Mengen $B_i \in \mathcal{B}_i$ ($i \in I$) die Ereignisse $X_i^{-1}(B_i)$ ($i \in I$) unabhängig unter \mathbb{P} sind, also für alle $m \in \mathbb{N}$ und alle paarweise verschiedenen Indizes $i_1, \dots, i_m \in I$

$$\mathbb{P}(\{X_{i_1} \in B_{i_1}\} \cap \dots \cap \{X_{i_m} \in B_{i_m}\}) = \mathbb{P}(X_{i_1} \in B_{i_1}) \cdot \dots \cdot \mathbb{P}(X_{i_m} \in B_{i_m})$$

gilt.

Statt zu sagen, dass die Familie $(X_i)_{i \in I}$ unabhängig unter \mathbb{P} ist, sagen wir etwas kürzer auch, dass die X_i ($i \in I$) unabhängig unter \mathbb{P} sind.

Anschaulich bedeutet Unabhängigkeit, dass wir aus den Werten einer der Zufallsgrößen (auf der Ebene der W.en) nichts Neues über die Werte der anderen Zufallsgrößen lernen können.

Bezeichnung. Wir schreiben statt $\mathbb{P}(\{X \in B\} \cap \{Y \in C\})$ auch $\mathbb{P}(X \in B, Y \in C)$, statt $\mathbb{P}(\{X = x\} \cap \{Y = y\})$ auch $\mathbb{P}(X = x, Y = y)$ usw.

Beispiel 3.15 (Kugelziehung; Fortsetzung von Beispiel 3.3). Aus einer Urne mit r roten und s schwarzen Kugeln wird $2 \times$ nacheinander gezogen; es interessieren die Farben X_1 und X_2 der gezogenen Kugeln.

Ziehen mit Zurücklegen:

Für alle $B_1, B_2 \subseteq \{\text{rot, sw.}\}$ gilt $\mathbb{P}(X_1 \in B_1, X_2 \in B_2) = \mathbb{P}(X_1 \in B_1) \mathbb{P}(X_2 \in B_2)$. Hat B_1 oder B_2 Mächtigkeit 0 oder 2, so ist dies klar. Die übrigen Fälle prüft man nach:

$$\mathbb{P}(X_1 = \text{rot}, X_2 = \text{rot}) = \frac{r^2}{(r+s)^2} = \frac{r}{r+s} \frac{r}{r+s} = \mathbb{P}(X_1 = \text{rot}) \mathbb{P}(X_2 = \text{rot}).$$

$$\mathbb{P}(X_1 = \text{rot}, X_2 = \text{sw.}) = \frac{rs}{(r+s)^2} = \frac{r}{r+s} \frac{s}{r+s} = \mathbb{P}(X_1 = \text{rot}) \mathbb{P}(X_2 = \text{sw.}).$$

$$\mathbb{P}(X_1 = \text{sw.}, X_2 = \text{rot}) = \frac{sr}{(r+s)^2} = \frac{s}{r+s} \frac{r}{r+s} = \mathbb{P}(X_1 = \text{sw.}) \mathbb{P}(X_2 = \text{rot}).$$

$$\mathbb{P}(X_1 = \text{sw.}, X_2 = \text{sw.}) = \frac{s^2}{(r+s)^2} = \frac{s}{r+s} \frac{s}{r+s} = \mathbb{P}(X_1 = \text{sw.}) \mathbb{P}(X_2 = \text{sw.}).$$

$\leadsto X_1$ und X_2 sind unabhängig.

Ziehen ohne Zurücklegen:

Hier gilt

$$\mathbb{P}(X_1 = \text{rot}, X_2 = \text{rot}) = \frac{r(r-1)}{(r+s)(r+s-1)} \neq \frac{r}{r+s} \frac{r}{r+s} = \mathbb{P}(X_1 = \text{rot}) \mathbb{P}(X_2 = \text{rot}).$$

$\leadsto X_1$ und X_2 sind abhängig (d. h. nicht unabhängig).

□

Bemerkungen 3.16.

- (a) Es sind $n > 2$ Zufallsgrößen X_1, \dots, X_n (wobei $X_i : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}_i, \mathcal{B}_i)$, $i = 1, \dots, n$) genau dann unabhängig, wenn für beliebige Mengen $B_i \in \mathcal{B}_i$, $i = 1, \dots, n$

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdot \dots \cdot \mathbb{P}(X_n \in B_n)$$

gilt. (Anders als in Bemerkung 3.13(b) genügt es hier also, alle n -fachen Schnitte zu untersuchen. Dies liegt daran, dass wir hier $B_i = \mathcal{X}_i$ setzen können.)

- (b) Teilfamilien von unabhängigen Familien von Zufallsgrößen sind wieder unabhängig.
(c) Eine unendliche Familie von Zufallsgrößen ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist.
(d) Um zu zeigen, dass eine Familie von reellwertigen Zufallsgrößen mit Werten unabhängig ist, genügt es zu zeigen, dass für jede Wahl von Intervallen $B_i \subseteq \mathbb{R}$, $i \in I$, die Ereignisse $X_i^{-1}(B_i)$ unabhängig sind.

Im Folgenden werden wir auf die Angabe der σ -Algebren meistens verzichten, vgl. dazu die Bemerkungen im Anschluss an Lemma 2.20.

Das folgende Lemma zeigt, wie sich die Unabhängigkeit

(a) für „diskrete Zufallsgrößen“ mit Hilfe von Elementar-W.'en

(b) für „stetige Zufallsgrößen“ mit Hilfe von Riemann-Dichten charakterisieren lässt:

Lemma 3.17.

(a) Sind X_1, \dots, X_n Zufallsgrößen mit Werten in abzählbaren Mengen $\mathcal{X}_1, \dots, \mathcal{X}_n$, so sind X_1, \dots, X_n unabhängig genau dann, wenn für alle $x_1 \in \mathcal{X}_1, \dots, x_n \in \mathcal{X}_n$

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdots \mathbb{P}(X_n = x_n)$$

gilt.

(b) Sind X_1, \dots, X_n reellwertige Zufallsgrößen mit den Riemann-Dichten f_1, \dots, f_n , so sind X_1, \dots, X_n unabhängig genau dann, wenn (X_1, \dots, X_n) die Riemann-Dichte

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

besitzt.

Beweis. Um Schreibarbeit zu sparen, beschränken wir uns auf den Fall $n = 2$.

(a) \Rightarrow ist klar, und \Leftarrow folgt durch Summation über Elementar-W.en:

$$\begin{aligned} \mathbb{P}(X_1 \in A_1, X_2 \in A_2) &= \sum_{x_1 \in A_1, x_2 \in A_2} \mathbb{P}(X_1 = x_1, X_2 = x_2) \\ &= \left(\sum_{x_1 \in A_1} \mathbb{P}(X_1 = x_1) \right) \left(\sum_{x_2 \in A_2} \mathbb{P}(X_2 = x_2) \right) = \mathbb{P}(X_1 \in A_1) \mathbb{P}(X_2 \in A_2) \end{aligned}$$

(b) \Rightarrow : Für alle Intervalle $I_1, I_2 \subseteq \mathbb{R}$ gilt

$$\begin{aligned} \mathbb{P}((X_1, X_2) \in I_1 \times I_2) &= \mathbb{P}(X_1 \in I_1, X_2 \in I_2) = \mathbb{P}(X_1 \in I_1) \mathbb{P}(X_2 \in I_2) \\ &= \int_{I_1} f_1(x_1) dx_1 \cdot \int_{I_2} f_2(x_2) dx_2 = \int_{I_1 \times I_2} f_1(x_1) f_2(x_2) d(x_1, x_2). \end{aligned}$$

Dies bedeutet, dass $f_1(x_1) f_2(x_2)$ eine Dichte von (X_1, X_2) ist.

\Leftarrow : Ist $f_1(x_1) f_2(x_2)$ eine Dichte von (X_1, X_2) , so gilt für alle Intervalle $I_1, I_2 \subseteq \mathbb{R}$

$$\begin{aligned} \mathbb{P}(X_1 \in I_1, X_2 \in I_2) &= \mathbb{P}((X_1, X_2) \in I_1 \times I_2) = \int_{I_1 \times I_2} f_1(x_1) f_2(x_2) d(x_1, x_2) \\ &= \int_{I_1} f_1(x_1) dx_1 \cdot \int_{I_2} f_2(x_2) dx_2 = \mathbb{P}(X_1 \in I_1) \mathbb{P}(X_2 \in I_2). \end{aligned}$$

Nach Bemerkung 3.16 (d) folgt daraus die Unabhängigkeit von X_1 und X_2 .

□

Warnung: Riemann-Dichten sind nicht (ganz) eindeutig bestimmt; man kann sie z. B. auf endlichen Mengen abändern, ohne das W.maß zu verändern. Man kann daher in (b) die Riemann-Dichte f von (X_1, \dots, X_n) so abändern, dass sie nicht mehr (vollständig) die angegebene Produkt-Gestalt besitzt.

Beispiel 3.18. Die Verteilung eines Zufallsvektors $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ sei durch die *n-dimensionale Standardnormalverteilung* gegeben, d. h. durch das stetige W.maß auf \mathbb{R}^n mit der *n-dimensionalen Riemann-Dichte*

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2}.$$

Dann gilt

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} \right) = \prod_{i=1}^n \varphi(x_i),$$

wobei φ die Riemann-Dichte der 1-dimensionalen Standardnormalverteilung bezeichnet. Nach Lemma 2.23 gilt für die Randdichten

$$f_{X_i}(x_i) = \int \cdots \int \cdots \int f(x_1, \dots, x_n) dx_n \cdots dx_{i+1} dx_{i-1} \cdots dx_1 = \varphi(x_i),$$

$i = 1, \dots, n$. Damit stimmen die Funktionen $f(x_1, \dots, x_n)$ und $f_{X_1}(x_1) \cdots f_{X_n}(x_n)$ überein, so dass X_1, \dots, X_n nach Lemma 3.17 (b) unabhängig sind. Außerdem sind X_1, \dots, X_n jeweils 1-dimensional standardnormalverteilt. \square

Bemerkung 3.19. Lemma 3.17 zeigt, dass bei unabhängigen Zufallsgrößen X_1, \dots, X_n mit diskreten / stetigen Verteilungen die *gemeinsame Verteilung* (d. h. die Verteilung von (X_1, \dots, X_n)) durch die *Randverteilungen* (d. h. die Verteilungen von X_1, \dots, X_n) eindeutig bestimmt ist, da die gemeinsame Dichte

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

durch die sog. *Produktdichte* der Randdichten gegeben ist.⁶ Allgemein gilt:

Merkregel:

Sind X_1, \dots, X_n unabhängige Zufallsgrößen, so ist die gemeinsame Verteilung durch die Randverteilungen eindeutig bestimmt.

Bei abhängigen Zufallsgrößen X_1, \dots, X_n gilt dies in aller Regel nicht, wie einfache Gegenbeispiele zeigen. Aus diesem Grund ist „Unabhängigkeit“ eine beliebte Annahme bei der stoch. Modellierung (selbst in Situationen, wo diese Annahme diskussionswürdig erscheint). \square

⁶ In diesem Zusammenhang sei daran erinnert, dass für eine Zufallsgröße X mit einer *diskreten* Verteilung auf einer abzählbaren Menge \mathcal{X} die Zähldichte von X gerade durch die Werte $\mathbb{P}(X = x)$, $x \in \mathcal{X}$, gegeben ist. Schreibt man $f_X(x)$ statt $\mathbb{P}(X = x)$, so sind die Formeln in Lemma 3.17 (a) und (b) formal identisch.

Produktdichten sind insbesondere auch dann nützlich, wenn man W.räume für Zufallsexperimente sucht, die sich aus mehreren „unabhängigen“ Telexperimenten zusammensetzen, deren Versuchsausgänge sich nicht gegenseitig beeinflussen:

Bemerkung 3.20 (Produktmodell). Gegeben seien n Zufallsexperimente, die sich jeweils

1. Fall: durch eine abzählbare Menge Ω_i und eine Zähldichte $f_i(\omega_i)$
2. Fall: durch die Menge $\Omega_i = \mathbb{R}$ und eine 1-dimensionale Riemann-Dichte $f_i(\omega_i)$

beschreiben lassen. Werden diese Zufallsexperimente „unabhängig voneinander“ durchgeführt, so lässt sich das Gesamtexperiment

1. Fall: durch die abzählbare Menge $\Omega := \Omega_1 \times \cdots \times \Omega_n$ und die Zähldichte $f(\omega_1, \dots, \omega_n) = f_1(\omega_1) \cdots f_n(\omega_n)$
2. Fall: durch die Menge $\Omega = \mathbb{R}^n$ und die n -dimensionale Riemann-Dichte $f(\omega_1, \dots, \omega_n) = f_1(\omega_1) \cdots f_n(\omega_n)$

beschreiben. Die Einzelergebnisse sind dann formal durch die Projektionen $X_i : \Omega \rightarrow \Omega_i$ mit $X_i(\omega_1, \dots, \omega_n) = \omega_i$ gegeben, und dies sind nach Lemma 3.17 (bzw. genauer mit ähnlicher Begründung wie in Beispiel 3.18) unabhängige Zufallsgrößen, da die gemeinsame Verteilung, also die Verteilung von $X = (X_1, \dots, X_n) = \text{id}_\Omega$, in beiden Fällen durch die *Produktdichte* $f(\omega_1, \dots, \omega_n) = f_1(\omega_1) \cdots f_n(\omega_n)$ gegeben ist.

Besonders wichtig ist der Spezialfall, dass dasselbe Zufallsexperiment n -mal „unabhängig voneinander“ durchgeführt wird; hier gilt: Beschreiben $\tilde{\Omega}$ und \tilde{f} das einfache Zufallsexperiment, so lässt sich die n -malige Durchführung durch $\Omega := \tilde{\Omega}^n$ und $f(\omega_1, \dots, \omega_n) := \tilde{f}(\omega_1) \cdots \tilde{f}(\omega_n)$ beschreiben. \square

Dieses *Produktmodell* haben wir (zumindest im diskreten Fall) schon oft verwendet, etwa in Beispiel 3 in Kapitel 0 oder bei der Herleitung der Binomialverteilung, wo wir die Wahl der (Produkt-)Dichte allerdings noch *ad hoc* mittels relativer Häufigkeiten begründen mussten.

Das Prinzip von der „getrennten Verarbeitung“.

Der folgende Satz zeigt, dass Unabhängigkeit bei „getrennter Verarbeitung“ erhalten bleibt:

Satz 3.21. *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum. Sind $X_i : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}_i, \mathcal{B}_i)$ ($i \in I$) unabhängige Zufallsgrößen und ist für jedes $i \in I$ eine messbare Abbildung $h_i : (\mathcal{X}_i, \mathcal{B}_i) \rightarrow (\mathcal{Y}_i, \mathcal{C}_i)$ gegeben, so sind auch $Y_i := h_i(X_i) : (\Omega, \mathcal{A}) \rightarrow (\mathcal{Y}_i, \mathcal{C}_i)$ ($i \in I$) unabhängige Zufallsgrößen.*

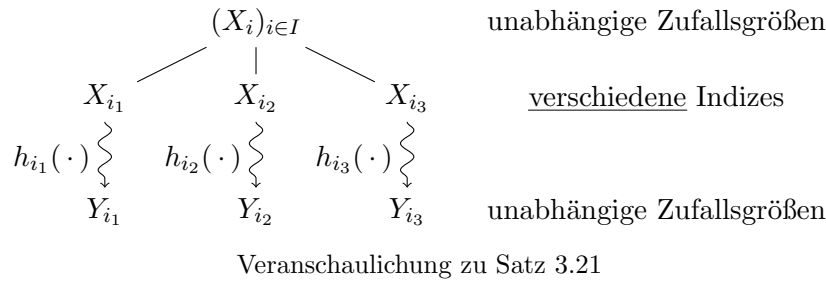
Beweis. Sind $C_i \in \mathcal{C}_i$ ($i \in I$) beliebig und dazu $B_i := h_i^{-1}(C_i) \in \mathcal{B}_i$ ($i \in I$), so gilt für alle $i \in I$

$$Y_i \in C_i \Leftrightarrow h_i(X_i) \in C_i \Leftrightarrow X_i \in h_i^{-1}(C_i) \Leftrightarrow X_i \in B_i$$

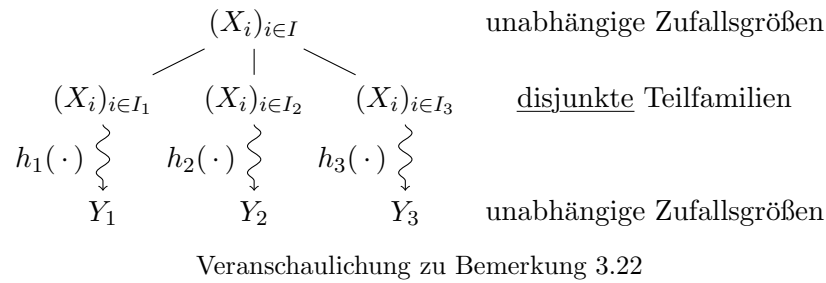
und damit für jede endliche Teilmenge $J \subseteq I$

$$\mathbb{P}\left(\bigcap_{i \in J} \{Y_i \in C_i\}\right) = \mathbb{P}\left(\bigcap_{i \in J} \{X_i \in B_i\}\right) \stackrel{\text{Unabh.}}{=} \prod_{i \in J} \mathbb{P}(\{X_i \in B_i\}) = \prod_{i \in J} \mathbb{P}(\{Y_i \in C_i\}).$$

□



Bemerkung 3.22. Satz 3.21 lässt sich wie folgt verallgemeinern: Zerlegt man eine unabhängige Familie von Zufallsgrößen $(X_i)_{i \in I}$ in mehrere Teilfamilien $(X_i)_{i \in I_1}, \dots, (X_i)_{i \in I_m}$ (d. h. jedes Element $i \in I$ muss in genau einer Menge I_m vorkommen) und wendet man auf jede dieser Teilfamilien eine (messbare) Abbildung h_1, \dots, h_m an, so sind auch die Zufallsgrößen $h_1((X_i)_{i \in I_1}), \dots, h_m((X_i)_{i \in I_m})$ unabhängig.



Satz 3.23 (Existenz von unabhängigen Zufallsgrößen mit vorgegebenen Verteilungen).
Es sei zu jedem $n \in \mathbb{N}$ ein W.maß P_n auf (\mathbb{R}, \mathbb{B}) gegeben. Dann existieren ein W.raum $(\Omega, \mathcal{A}, \mathbb{P})$ und eine unabhängige Folge von Zufallsgrößen $(X_n)_{n \in \mathbb{N}}$, so dass $\mathbb{P}_{X_n} = P_n$ für alle $n \in \mathbb{N}$ gilt.

Der Satz bleibt auch für allgemeinere Räume als (\mathbb{R}, \mathbb{B}) richtig, allerdings ist er dann nicht mehr so einfach zu beweisen.

*Beweis.*⁷ Wir wählen $(\Omega, \mathcal{A}, \mathbb{P}) := ([0, 1[, \mathbb{B}_{[0,1[}, \mathcal{U}_{[0,1[})$. Für $\omega \in \Omega$ und $i \in \mathbb{N}$ sei $Y_i(\omega)$ die i -te Nachkommastelle in der Dezimalbruchdarstellung von ω . (Es kann vorkommen, dass ω zwei verschiedene Darstellungen besitzt, etwa $\omega = \frac{1}{2} = 0.4999\dots = 0.5000\dots$; in diesem Fall wählen wir die Darstellung auf $\dots 000\dots$.) Dann sind Y_1, Y_2, Y_3, \dots unabhängige Zufallsgrößen, die jeweils gleichverteilt auf der Menge $\{0, \dots, 9\}$ sind, denn: Für alle $n \in \mathbb{N}$ und alle $y_1, \dots, y_n \in \{0, \dots, 9\}$ gilt mit $y := \sum_{i=1}^n y_i 10^{-i}$

$$\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n) = \mathbb{P}([y, y + 10^{-n}[) = 10^{-n}.$$

Damit folgt einerseits für alle $n \in \mathbb{N}$ und alle $y_n \in \{0, \dots, 9\}$

$$\mathbb{P}(Y_n = y_n) \stackrel{\text{Lemma 1.19}}{=} \sum_{y_1, \dots, y_{n-1} \in \{0, \dots, 9\}} \mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n) = 10^{n-1} 10^{-n} = 10^{-1},$$

d. h. jede Zufallsgröße Y_n ist gleichverteilt auf der Menge $\{0, \dots, 9\}$, und andererseits für alle $n \in \mathbb{N}$ und alle $y_1, \dots, y_n \in \{0, \dots, 9\}$

$$\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n) = 10^{-n} = 10^{-1} \cdot \dots \cdot 10^{-1} = \mathbb{P}(Y_1 = y_1) \cdot \dots \cdot \mathbb{P}(Y_n = y_n),$$

d. h. die Zufallsgrößen Y_1, Y_2, Y_3, \dots sind (nach Lemma 3.17 (a) und Bemerkung 3.16 (c)) unabhängig.

Wir ordnen die Zufallsgrößen nun in der Form

$$\begin{array}{ccccccc} Y_1 & Y_3 & Y_6 & Y_{10} & \dots & & \\ Y_2 & Y_5 & Y_9 & & & & \\ Y_4 & Y_8 & & \ddots & & & \\ Y_7 & & \ddots & & & & \\ \vdots & & & & & & \end{array}$$

an und definieren U_n als diejenige Zahl im Intervall $[0, 1]$, deren Dezimalziffern durch die n -te Zeile gegeben sind. Dann sind die Zufallsgrößen U_n nach dem Prinzip von der getrennten Verarbeitung unabhängig. Ferner ist jede Zufallsgröße U_n $\mathcal{U}_{(0,1)}$ -verteilt: Sind nämlich $Y_{n,1}, Y_{n,2}, Y_{n,3}, \dots$ die Ziffern in der n -ten Zeile, so gilt für alle $m \in \mathbb{N}$ für alle $y \in \{0, \dots, 10^m - 1\}$ mit der Dezimaldarstellung $y = \sum_{i=0}^{m-1} y_i 10^i$

$$\mathbb{P}(U_n \in [y 10^{-m}, (y+1) 10^{-m}[) = \mathbb{P}(Y_{n,1} = y_{m-1}, \dots, Y_{n,m} = y_1) = 10^{-m}$$

und damit für alle $x \in \{0, \dots, 10^m - 1\}$ mit der Dezimaldarstellung $x = \sum_{i=0}^{m-1} x_i 10^i$

$$\mathbb{P}(U_n < x 10^{-m}) = \sum_{y \in \mathbb{N}_0 : y < x} \mathbb{P}(U_n \in [y 10^{-m}, (y+1) 10^{-m}[) = x 10^{-m}.$$

Aus Monotoniegründen folgt daher $\mathbb{P}(U_n \leq x) = x$ für alle $x \in (0, 1)$ bzw. $\mathbb{P}_{U_n} = \mathcal{U}_{(0,1)}$, da die Verteilung von U_n durch die Verteilungsfunktion von U_n bestimmt ist.

⁷ Dieser Beweis ist in der Vorlesung übersprungen worden.

Wir setzen nun $X_n := G_n(U_n)$, $n \in \mathbb{N}$, wobei für jedes $n \in \mathbb{N}$ G_n die Pseudo-Inverse der Verteilungsfunktion zu P_n sei (vgl. Satz 2.25). Dann sind die Zufallsgrößen X_n nach dem Prinzip von der getrennten Verarbeitung wieder unabhängig, und nach Satz 2.25 gilt $\mathbb{P}_{X_n} = P_n$ für alle $n \in \mathbb{N}$. \square

Verteilungen von Summen unabhängiger Zufallsgrößen.⁸

Wie in Bemerkung 3.19 angemerkt, sind im Falle der Unabhängigkeit der Zufallsgrößen X und Y $\mathbb{P}_{(X,Y)}$ und damit \mathbb{P}_{X+Y} eindeutig durch \mathbb{P}_X und \mathbb{P}_Y bestimmt. Genauer gilt:

Lemma 3.24 (Summen unabhängiger Zufallsgrößen).

(a) Sind X, Y unabhängige Zufallsgrößen mit Werten in \mathbb{Z} und Zähldichten f_X, f_Y , so hat $Z := X + Y$ die Zähldichte

$$f_Z(n) = \sum_{k=-\infty}^{+\infty} f_X(k) f_Y(n-k) = \sum_{l=-\infty}^{+\infty} f_X(n-l) f_Y(l), \quad n \in \mathbb{Z}.$$

(b) Sind X, Y unabhängige Zufallsgrößen mit Werten in \mathbb{R} mit Riemann-Dichten f_X, f_Y , so hat $Z := X + Y$ die Riemann-Dichte

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy, \quad z \in \mathbb{R}.$$

Beweis. Es genügt jeweils, die erste Formel zu zeigen; die zweite Formel folgt analog. Da X, Y unabhängig sind, besitzt (X, Y) die Produktdichte $f_X(x) f_Y(y)$. Damit folgt in Teil (a)

$$\mathbb{P}(X + Y = n) = \sum_{k,l \in \mathbb{Z} : k+l=n} f_X(k) f_Y(l) \stackrel{l=n-k}{=} \sum_{k \in \mathbb{Z}} f_X(k) f_Y(n-k)$$

für alle $n \in \mathbb{Z}$ und in Teil (b)

$$\begin{aligned} \mathbb{P}(X + Y \leq z) &= \int_{\{(x,y) \in \mathbb{R}^2 : x+y \leq z\}} f_X(x) f_Y(y) d(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) dy dx \\ &\stackrel{y=t-x}{=} \int_{-\infty}^{+\infty} \int_{-\infty}^z f_X(x) f_Y(t-x) dt dx \stackrel{\text{Reihenfolge vertauschen}}{=} \int_{-\infty}^z \int_{-\infty}^{+\infty} f_X(x) f_Y(t-x) dx dt \end{aligned}$$

für alle $z \in \mathbb{R}$. Dies ist aber gerade die Verteilungsfunktion von $X_1 + X_2$, aus der man durch Differenziation nach z die Riemann-Dichte von $X_1 + X_2$ erhält. \square

⁸ Dieser Abschnitt ist in der Vorlesung übersprungen worden.

Beispiele 3.25 (Summen unabhängiger Zufallsgrößen).

(a) Für alle $n, m \in \mathbb{N}$ und alle $p \in [0, 1]$ gilt:

$$X \sim \mathcal{B}(n, p), Y \sim \mathcal{B}(m, p) \text{ unabhängig} \Rightarrow X + Y \sim \mathcal{B}(n + m, p).$$

(b) Für alle $a, b \in \mathbb{R}$ und alle $\sigma, \tau \in (0, \infty)$ gilt:

$$X \sim \mathcal{N}(a, \sigma^2), Y \sim \mathcal{N}(b, \tau^2) \text{ unabhängig} \Rightarrow X + Y \sim \mathcal{N}(a + b, \sigma^2 + \tau^2).$$

Beweis.

(a) Für alle $k \in \mathbb{N}_0$ gilt

$$\begin{aligned} f_Z(k) &= \sum_{j=0}^k f_X(j) f_Y(k-j) = \sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j} (1-p)^{m-k+j} \\ &= \left(\sum_{j=0}^k \binom{n}{j} \binom{m}{k-j} \right) p^k (1-p)^{n+m-k} = \binom{n+m}{k} p^k (1-p)^{n+m-k}. \end{aligned}$$

Dabei folgt der letzte Schritt aus der kombinatorischen Interpretation des Binomialkoeffizienten: Eine k -elementige Menge von $\{1, \dots, n+m\}$ entsteht dadurch, dass man j Elemente aus $\{1, \dots, n\}$ und $k-j$ Elemente aus $\{n+1, \dots, n+m\}$ auswählt, wobei $j \in \{0, \dots, k\}$.

Alternativ kann man sich dieses Ergebnis mit Hilfe der anschaulichen Bedeutung der Binomialverteilung klar machen: Ist ein Bernoulli-Experiment mit Erfolgs-W. p gegeben, $X \sim \mathcal{B}(n, p)$ die Anzahl der „Erfolge“ bei n Versuchen und $Y \sim \mathcal{B}(m, p)$ die Anzahl der „Erfolge“ bei m Versuchen, wobei alle Versuche unabhängig sind, so ist $X + Y \sim \mathcal{B}(n + m, p)$ die Anzahl der „Erfolge“ bei $n + m$ Versuchen.

(b) Wir betrachten zunächst den Fall $a = 0, b = 0$. Hier gilt für alle $y, z \in \mathbb{R}$

$$\begin{aligned} \frac{(z-y)^2}{2\sigma^2} + \frac{y^2}{2\tau^2} &= \frac{y^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} - \frac{2\frac{\tau^2}{\sigma^2+\tau^2}zy}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} + \frac{\left(\frac{\tau^2}{\sigma^2+\tau^2}\right)^2 z^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} + \frac{z^2}{2\sigma^2} - \frac{\left(\frac{\tau^2}{\sigma^2+\tau^2}\right)^2 z^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} \\ &= \frac{\left(y - \frac{\tau^2}{\sigma^2+\tau^2}z\right)^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} + \frac{z^2}{2(\sigma^2 + \tau^2)} \end{aligned}$$

und damit für alle $z \in \mathbb{R}$

$$\begin{aligned} f_Z(z) &= \int_{\mathbb{R}} f_X(z-y) f_Y(y) dy = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-y)^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{y^2}{2\tau^2}\right) dy \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \exp\left(-\frac{z^2}{2(\sigma^2 + \tau^2)}\right) \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}} \exp\left(-\frac{\left(y - \frac{\tau^2}{\sigma^2+\tau^2}z\right)^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}\right) dy}_{=1} \end{aligned}$$

(da unter dem Integral eine Normalverteilungsdichte steht), also $Z \sim \mathcal{N}(0, \sigma^2 + \tau^2)$.

Sind nun $a, b \in \mathbb{R}$ beliebig, so können wir (nach Beispiel 2.22 (b)) Darstellungen $X = \tilde{X} + a$ und $Y = \tilde{Y} + b$ mit $\tilde{X} \sim \mathcal{N}(0, \sigma^2)$, $\tilde{Y} \sim \mathcal{N}(0, \tau^2)$ und \tilde{X}, \tilde{Y} unabhängig (Satz 3.21!) finden, und damit folgt mit dem bereits Gezeigten (und nochmals Beispiel 2.22 (b)) $\tilde{X} + \tilde{Y} \sim \mathcal{N}(0, \sigma^2 + \tau^2)$ bzw. $X + Y = (\tilde{X} + \tilde{Y}) + (a + b) \sim \mathcal{N}(a + b, \sigma^2 + \tau^2)$. □

Warnung. Dass die Verteilung der Summe zweier unabhängiger Zufallsgrößen wieder eine bekannte Verteilung ist, ist eher die Ausnahme.

Zusammenfassung:

- Mit Hilfe von bedingten W.en lassen sich W.en unter Teilinformationen über den Ausgang eines Zufallsexperimentes beschreiben.
- Bedingte W.en lassen sich bei mehrstufigen Zufallsexperimenten zur Konstruktion von W.räumen verwenden.
- Zwei Ereignisse $A, B \in \mathcal{A}$ heißen unabhängig, wenn $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Anschaulich bedeutet dies, dass man aus dem Eintreten des einen Ereignisses keine Rückschlüsse auf das Eintreten des anderen Ereignisses ziehen kann.
- Zwei Zufallsgrößen $X : \Omega \rightarrow \mathcal{X}$, $Y : \Omega \rightarrow \mathcal{Y}$ heißen unabhängig, wenn $\mathbb{P}(X \in B, Y \in C) = \mathbb{P}(X \in B)\mathbb{P}(Y \in C)$ für alle Mengen B, C aus den jeweiligen σ -Algebren \mathcal{B}, \mathcal{C} auf der Bildseite. Anschaulich bedeutet dies, dass man aus dem Wert der einen Zufallsgröße keine Rückschlüsse auf den Wert der anderen Zufallsgröße ziehen kann.
- Die Unabhängigkeit von mehr als zwei Ereignissen bzw. Zufallsgrößen ist etwas aufwendiger zu definieren, um sicherzustellen, dass Unabhängigkeit bei Übergang zu Teilfamilien erhalten bleibt.
- Die Unabhängigkeit von Zufallsgrößen X_1, \dots, X_n mit diskreten bzw. stetigen Verteilungen kann am einfachsten über die gemeinsame Dichte (falls existent!) untersucht werden:

X_1, \dots, X_n unabhängig mit den Dichten $f_1(x_1), \dots, f_n(x_n)$.

$\Leftrightarrow (X_1, \dots, X_n)$ besitzt die „Produktdichte“ $f_1(x_1) \cdots f_n(x_n)$.

- Braucht man unabhängige Zufallsgrößen X_1, \dots, X_n mit vorgegebenen diskreten bzw. stetigen Verteilungen, so kann man die Projektionen auf dem jeweiligen Produktraum versehen mit der *Produktdichte* betrachten.

4 Erwartungswert und Varianz

Es seien stets $(\Omega, \mathcal{A}, \mathbb{P})$ ein W.raum und $X : \Omega \rightarrow \mathbb{R}$ eine reellwertige Zufallsgröße.

Wir wollen nun einige *Kennzahlen* zur Beschreibung von X einführen, den *Mittelwert* / *Erwartungswert* als „Lagemaß“ und die *mittlere quadratische Abweichung* / *Varianz* als „Streuungsmaß“.

Bemerkung. Hat eine Zufallsgröße X die Verteilung P , so schreiben wir kurz $X \sim P$.

4.1 Erwartungswert

Wir beginnen (anders als sonst) mit einer Zusammenstellung der wichtigsten Formeln zur Berechnung von Erwartungswerten; Motivationen, Definitionen, Bemerkungen und Beweiseideen sowie weitere Beispiele werden in Kapitel 4.2 und 4.3 nachgetragen.

Berechnung von Erwartungswerten von Zufallsgrößen.

Ist $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße mit einer diskreten Verteilung und $\mathcal{X} \subset \mathbb{R}$ eine abzählbare Menge mit $\mathbb{P}(X \in \mathcal{X}) = 1$, so gilt

$$\mathbb{E}X = \sum_{x \in \mathcal{X}} x \mathbb{P}(X = x), \quad \text{falls} \quad \mathbb{E}|X| = \sum_{x \in \mathcal{X}} |x| \mathbb{P}(X = x) < \infty.$$

Ist $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße mit einer stetigen Verteilung mit der Riemann-Dichte f , so gilt

$$\mathbb{E}X = \int_{\mathbb{R}} x f(x) dx \quad \text{falls} \quad \mathbb{E}|X| = \int_{\mathbb{R}} |x| f(x) dx < \infty.$$

Berechnung von Erwartungswerten von Funktionen von Zufallsgrößen.

Ist $Y : \Omega \rightarrow \mathbb{R}^d$ eine Zufallsgröße mit einer diskreten Verteilung, $\mathcal{Y} \subset \mathbb{R}^d$ eine abzählbare Menge mit $\mathbb{P}(Y \in \mathcal{Y}) = 1$ und $h : \mathbb{R}^d \rightarrow \mathbb{R}$ eine Funktion, so gilt

$$\mathbb{E}(h \circ Y) = \sum_{y \in \mathcal{Y}} h(y) \mathbb{P}(Y = y), \quad \text{falls} \quad \mathbb{E}|h \circ Y| = \sum_{y \in \mathcal{Y}} |h(y)| \mathbb{P}(Y = y) < \infty.$$

Ist $Y : \Omega \rightarrow \mathbb{R}^d$ eine Zufallsgröße mit einer stetigen Verteilung mit der Riemann-Dichte f und $h : \mathbb{R}^d \rightarrow \mathbb{R}$ eine Funktion, so dass $h \cdot f_Y$ „lokal Riemann-integrierbar“ ist, so gilt

$$\mathbb{E}(h \circ Y) = \int_{\mathbb{R}^d} h(y) f(y) dy \quad \text{falls} \quad \mathbb{E}|h \circ Y| = \int_{\mathbb{R}^d} |h(y)| f(y) dy < \infty.$$

Diese Formeln sind auch als *Transformationsformeln* bekannt, da sie eine Beziehung zwischen dem Erwartungswert von $h \circ Y$ und der Verteilung von Y herstellen.

Im diskreten Fall könnten wir statt $\mathbb{P}(X = x)$ und $\mathbb{P}(Y = y)$ auch $f_X(x)$ und $f_Y(y)$ schreiben, wobei f_X und f_Y die Zähldichte von X bzw. Y bezeichnen. Es gilt dann:

Merkregel:

Die Erwartungswertformeln für Zufallsgrößen mit stetigen Verteilungen ergeben sich formal aus den Erwartungswertformeln für Zufallsgrößen mit diskreten Verteilungen, indem man Summen durch Integrale ersetzt.

Wir können die Erwartungswertformeln verwenden, um für eine binomial- bzw. normalverteilte Zufallsgröße X jeweils $\mathbb{E}(X)$ und $\mathbb{E}(X^2)$ zu berechnen:

Beispiel (Berechnung von $\mathbb{E}X$ und $\mathbb{E}X^2$ für $X \sim \mathcal{B}(n, p)$). Sei X eine Zufallsgröße mit $X \sim \mathcal{B}(n, p)$. Dann nimmt X m.W. 1 Werte in der *endlichen* Menge $\{0, \dots, n\}$ an, so dass die folgenden Erwartungswerte existieren. Es gilt

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k \mathbb{P}(X = k) \\ &= \sum_{k=1}^n \underbrace{k \binom{n}{k}}_{=n \binom{n-1}{k-1}} p^k (1-p)^{n-k} \\ &\stackrel{j=k-1}{=} np \underbrace{\sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}}_{=1 \left(\begin{array}{l} \text{Summe der Elementar-W.en} \\ \text{der } \mathcal{B}(n-1, p)\text{-Verteilung} \end{array} \right)} = np \end{aligned}$$

und (nach der Transformationsformel angewendet mit $Y := X$ und $h(x) := x^2$)

$$\begin{aligned} \mathbb{E}(X^2) &= \sum_{k=0}^n k^2 \mathbb{P}(X = k) \\ &= \sum_{k=0}^n k \mathbb{P}(X = k) + \sum_{k=0}^n k(k-1) \mathbb{P}(X = k) \\ &= np + \sum_{k=2}^n \underbrace{k(k-1) \binom{n}{k}}_{=n(n-1) \binom{n-2}{k-2}} p^k (1-p)^{n-k} \\ &\stackrel{j=k-2}{=} np + n(n-1)p^2 \underbrace{\sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j}}_{=1 \left(\begin{array}{l} \text{Summe der Elementar-W.en} \\ \text{der } \mathcal{B}(n-2, p)\text{-Verteilung} \end{array} \right)} = n^2 p^2 + np(1-p). \end{aligned}$$

Im Fall $n = 1$ ist die zweite große Summe dabei als Null zu lesen. □

Beispiel (Berechnung von $\mathbb{E}X$ und $\mathbb{E}X^2$ für $X \sim \mathcal{N}(0, 1)$). Sei X eine Zufallsgröße mit $X \sim \mathcal{N}(0, 1)$. Dann gilt (mit $\varphi :=$ Riemann-Dichte der Standardnormalverteilung) $\mathbb{E}|X|^k = \int |x|^k \varphi(x) dx < \infty$ für alle $k \in \mathbb{N}$ (nach Analysis, da $\varphi(x)$ für $x \rightarrow \pm\infty$ schneller fällt, als jede Potenz von $|x|$ wächst), so dass $\mathbb{E}X^k$ für alle $k \in \mathbb{N}$ existiert. Weiter gilt wegen $\varphi'(x) = -x\varphi(x)$

$$\mathbb{E}X = \int_{\mathbb{R}} x\varphi(x) dx = \left[-\varphi(x) \right]_{x=-\infty}^{x=+\infty} = 0$$

und (nach der Transformationsformel angewendet mit $Y := X$ und $h(x) = x^2$)

$$\mathbb{E}X^2 = \int_{\mathbb{R}} x^2 \varphi(x) dx \stackrel{\text{partielle Integration}}{=} \left[-x\varphi(x) \right]_{x=-\infty}^{x=+\infty} + \int_{\mathbb{R}} \varphi(x) dx = 0 + 1 = 1$$

(da φ als Wahrscheinlichkeitsdichte das Integral 1 besitzt). □

Für Erwartungswerte gelten die folgenden Rechenregeln:

Satz (Rechenregeln für Erwartungswerte).

Es seien X und Y Zufallsgrößen, so dass $\mathbb{E}(X)$ und $\mathbb{E}(Y)$ existieren, und $a, b, c \in \mathbb{R}$. Dann gilt:

- (a) $\mathbb{E}(aX + bY)$ existiert, und $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$. (Linearität)
- (b) Aus $X \leq Y$ folgt $\mathbb{E}(X) \leq \mathbb{E}(Y)$. (Monotonie)
- (c) $\mathbb{E}(c) = c$.
- (d) Sind X, Y unabhängig, so existiert $\mathbb{E}(XY)$, und es gilt $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. (Multiplikationssatz)

Formal führen wir den Erwartungswert (EW) nun zuerst für diskrete Zufallsgrößen und anschließend für allgemeine Zufallsgrößen ein. Dabei werden wir den Erwartungswert (etwas allgemeiner als bisher) auch für Zufallsgrößen erklären, die weder eine diskrete noch eine stetige Verteilung besitzen. Außerdem werden wir uns natürlich (zumindest andeutungsweise) überlegen, warum die obigen Erwartungsformeln „richtig“ sind.

4.2 Der Erwartungswert von diskreten Zufallsgrößen

Definition 4.1. Ist $X(\Omega) := \{X(\omega) : \omega \in \Omega\}$ eine abzählbare Teilmenge von \mathbb{R} , so bezeichnen wir X als *diskrete Zufallsgröße*.

In diesem Fall können wir \mathbb{P}_X als diskretes W.maß auf $X(\Omega)$ ansehen.

Zur Motivation des Erwartungswertes sei angenommen, dass wir das zugrunde liegende Zufallsexperiment – unabhängig voneinander – n -mal durchführen und dabei die Beobachtungen $x_1, \dots, x_n \in \mathcal{X}$ erhalten. Dann gilt für den Mittelwert

$$\frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{x \in \mathcal{X}} x H_n(\{X=x\}) = \sum_{x \in \mathcal{X}} x h_n(\{X=x\}) \approx \sum_{x \in \mathcal{X}} x \mathbb{P}(\{X=x\}) =: \mathbb{E}(X)$$

(wobei wir unterstellen, dass die auftretenden Summen sinnvoll definiert werden können). Dieser Mittelwert lässt sich z. B. zur Bestimmung von fairen Einsätzen bei Glücksspielen heranziehen, wo der „faire“ Einsatz durch die mittlere Auszahlung gegeben ist:

Beispiel 4.2 (Fairer Einsatz bei Glücksspielen).

- (a) Ein fairer Würfel wird geworfen; der Spieler erhält die Augenzahl als Auszahlung (in Euro). Wie lautet der „faire“ Einsatz?

Sei X die gewürfelte Augenzahl. Dann gilt $X \sim \mathcal{U}_{\{1,2,3,4,5,6\}}$. Damit ergibt sich für die mittlere Auszahlung

$$\mathbb{E}(X) = \sum_{i=1}^6 i \mathbb{P}(\{X=i\}) = \sum_{i=1}^6 i \frac{1}{6} = 3,5,$$

so dass der „faire“ Einsatz 3,5 Euro beträgt.

- (b) Eine faire Münze wird geworfen, bis das erste Mal „Kopf“ fällt. Geschieht dies im i -ten Versuch, so erhält der Spieler i Euro. Wie lautet der „faire“ Einsatz?

Sei Y die Anzahl der benötigten Versuche. Dann gilt $Y \sim \mathcal{G}(\frac{1}{2})$. Damit folgt für die mittlere Auszahlung

$$\mathbb{E}(Y) = \sum_{i=1}^{\infty} i \mathbb{P}(Y=i) = \sum_{i=1}^{\infty} i 2^{-i} = \frac{\frac{1}{2}}{(1 - \frac{1}{2})^2} = 2,$$

so dass der „faire“ Einsatz 2 Euro beträgt.

- (c) Eine faire Münze wird geworfen, bis das erste Mal „Kopf“ fällt. Geschieht dies im i -ten Versuch, so erhält der Spieler 2^i Euro. Wie lautet der „faire“ Einsatz?

Sei Y die Anzahl der benötigten Versuche. Dann gilt $Y \sim \mathcal{G}(\frac{1}{2})$. Damit folgt für die mittlere Auszahlung

$$\mathbb{E}(2^Y) = \sum_{i=1}^{\infty} 2^i \mathbb{P}(2^Y = 2^i) = \sum_{i=1}^{\infty} 2^i \mathbb{P}(Y=i) = \sum_{i=1}^{\infty} 2^i 2^{-i} = \sum_{i=1}^{\infty} 1 = \infty,$$

d. h. es existiert kein „fairer“ Einsatz!

- (d) Eine faire Münze wird geworfen, bis das erste Mal „Kopf“ fällt. Geschieht dies im i -ten Versuch, so erhält der Spieler 2^i Euro, falls i gerade ist, und zahlt 2^i Euro, falls i ungerade ist. (Ein Einsatz ist nicht vorgesehen.) Ist das Spiel „fair“?

Das Spiel ist „fair“, wenn der mittlere Gewinn = 0 ist. Mit Y wie in (b) und (c) ergibt sich für den mittleren Gewinn

$$\mathbb{E}((-2)^Y) = \sum_{i=1}^{\infty} (-2)^i \mathbb{P}((-2)^Y = (-2)^i) = \sum_{i=1}^{\infty} (-1)^i 2^i 2^{-i} = \sum_{i=1}^{\infty} (-1)^i = ???,$$

d. h. es existiert kein mittlerer Gewinn!

Wie die Beispiele (c) und (d) verdeutlichen, muss man bei der Einführung von Mittelwerten ein wenig aufpassen. Die formale Definition des Erwartungswertes ist deshalb etwas komplizierter:

Definition 4.3. Sei $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsgröße. Falls

$$\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < \infty, \quad (*)$$

so sagen wir, dass der Erwartungswert von X *existiert*, und nennen

$$\mathbb{E}X := \mathbb{E}(X) := \sum_{x \in X(\Omega)} x \mathbb{P}(X = x) \quad (**)$$

Erwartungswert von X .

Bemerkungen 4.4.

- (a) Ist (*) erfüllt, so ist die Reihe in (**) absolut konvergent, so dass $\mathbb{E}(X)$ durch (**) sinnvoll definiert (also insbesondere unabhängig von der Summationsreihenfolge) ist.
- (b) Ist X beschränkt, so existiert der Erwartungswert von X immer.
- (c) Ist $X \geq 0$, so sei $\mathbb{E}X := \sum_{x \in X(\Omega)} x \mathbb{P}(X = x)$, selbst wenn die Reihe divergiert! (Wir bezeichnen $\mathbb{E}X$ dann als *Erwartungswert im weiteren Sinne*.)
- (d) Mit der Konvention aus Teil (c) gilt: Der Erwartungswert von X existiert genau dann, wenn $\mathbb{E}|X| < \infty$.

denn: Mit $Z := |X|$ gilt

$$\begin{aligned} \mathbb{E}|X| &\stackrel{\text{Bem. 4.4 (c)}}{=} \sum_{z \in Z(\Omega)} z \mathbb{P}(|X| = z) \\ &\stackrel{\text{Additivität}}{=} \sum_{z \in Z(\Omega)} z \sum_{x \in X(\Omega): |x|=z} \mathbb{P}(X = x) \\ &\stackrel{\text{Umordnen}}{=} \sum_{x \in X(\Omega)} \sum_{z \in Z(\Omega): z=|x|} z \mathbb{P}(X = x) = \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x). \end{aligned}$$

Dabei ist Umordnen zulässig, da alle Summanden ≥ 0 sind.

- (e) Da in den auftretenden Summen die Terme mit $\mathbb{P}(X = x) = 0$ keine Rolle spielen, hängt der Erwartungswert von X nur von der Verteilung von X ab. Man spricht daher auch vom Erwartungswert einer Verteilung. Insbesondere müssen wir also – ähnlich wie in Bemerkung 1.21 – nur \mathbb{P}_X kennen, um $\mathbb{E}(X)$ zu berechnen.

Oft interessiert man sich nicht für den Erwartungswert von Y , sondern von einer Funktion von Y , etwa $\mathbb{E}(Y^2)$. Hier ist der folgende Satz nützlich, nach dem man $\mathbb{E}(h \circ Y)$ ausgehend von \mathbb{P}_Y berechnen kann, ohne den Umweg über $\mathbb{P}_{h \circ Y}$ gehen zu müssen:

Satz 4.5 (Transformationsformel für Zufallsgrößen mit diskreten Verteilungen).

Ist $Y : \Omega \rightarrow \mathbb{R}^d$ eine diskrete Zufallsgröße und $h : \mathbb{R}^d \rightarrow \mathbb{R}$ eine Abbildung, so ist $h \circ Y : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsgröße, und es gilt

$$\mathbb{E}|h \circ Y| = \sum_{y \in Y(\Omega)} |h(y)| \mathbb{P}(Y = y).$$

Also existiert $\mathbb{E}(h \circ Y)$ genau dann, wenn die rechte Seite endlich ist, und in diesem Fall gilt

$$\mathbb{E}(h \circ Y) = \sum_{y \in Y(\Omega)} h(y) \mathbb{P}(Y = y).$$

*Beweis.*⁹ Mit $Y(\Omega)$ ist auch $(h \circ Y)(\Omega)$ abzählbar, d. h. $h \circ Y$ ist eine diskrete Zufallsgröße. Weiter gilt mit $Z := h \circ Y$

$$\begin{aligned} \mathbb{E}|h \circ Y| &\stackrel{\text{Bem. 4.4 (d)}}{=} \sum_{z \in Z(\Omega)} |z| \mathbb{P}(h \circ Y = z) \\ &\stackrel{\sigma\text{-Additivität}}{=} \sum_{z \in Z(\Omega)} |z| \sum_{y \in Y(\Omega): h(y)=z} \mathbb{P}(Y = y) \\ &\stackrel{\text{Umordnen}}{=} \sum_{y \in Y(\Omega)} \sum_{z \in Z(\Omega): z=h(y)} |z| \mathbb{P}(Y = y) = \sum_{y \in Y(\Omega)} |h(y)| \mathbb{P}(Y = y) \end{aligned}$$

und damit die erste Gleichung, wobei Umordnen zulässig ist, da alle Summanden ≥ 0 sind. Sind die Summen endlich, so folgt die zweite Gleichung aus der gleichen Rechnung ohne Betragsstriche, wobei Umordnen wegen absoluter Konvergenz zulässig ist. \square

⁹ Dieser Beweis ist in der Vorlesung übersprungen worden.

Satz 4.6 (Rechenregeln für Erwartungswerte).

Die Rechenregeln für Erwartungswerte (vgl. den Satz in Kapitel 4.1) gelten für diskrete Zufallsgrößen.

Beweis. Mit X und Y ist auch (X, Y) eine diskrete Zufallsgröße.

(a) Mit der Wahl $h(x, y) := ax + by$ in Satz 4.5 folgt

$$\begin{aligned}
\mathbb{E}|aX + bY| &= \sum_{x \in X(\Omega), y \in Y(\Omega)} |ax + by| \mathbb{P}(X = x, Y = y) \\
&\leq \sum_{x \in X(\Omega), y \in Y(\Omega)} (|a||x| + |b||y|) \mathbb{P}(X = x, Y = y) \\
&= |a| \sum_{x \in X(\Omega)} |x| \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y) + |b| \sum_{y \in Y(\Omega)} |y| \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y) \\
&= |a| \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) + |b| \sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) \\
&= |a| \mathbb{E}|X| + |b| \mathbb{E}|Y| < \infty.
\end{aligned}$$

Dabei kann im 3. Schritt umgeordnet worden, weil alle Summanden ≥ 0 sind, und im 4. Schritt ist die Formel für die Randverteilung (Lemma 1.19) eingegangen. Also existiert $\mathbb{E}(aX + bY)$. Dieselbe Rechnung ohne Betragsstriche (und ohne Dreiecksungleichung) liefert dann $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$, wobei Umordnen dieses Mal wegen absoluter Konvergenz zulässig ist.

(b) Wegen der Linearität des Erwartungswertes genügt es, $Y - X \geq 0 \Rightarrow \mathbb{E}(Y - X) \geq 0$ zu zeigen, und dies ist offensichtlich.

(c) Dies ist klar.

(d) Mit der Wahl $h(x, y) := xy$ in Satz 4.5 folgt

$$\begin{aligned}
\mathbb{E}|XY| &= \sum_{x \in X(\Omega), y \in Y(\Omega)} |xy| \mathbb{P}(X = x, Y = y) \\
(\text{Unabh.}) &= \sum_{x \in X(\Omega), y \in Y(\Omega)} |x||y| \mathbb{P}(X = x) \mathbb{P}(Y = y) \\
&= \left(\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) \right) \left(\sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) \right) \\
&= \mathbb{E}|X| \mathbb{E}|Y| < \infty.
\end{aligned}$$

Dabei ist im 3. Schritt die Unabhängigkeit von X und Y eingegangen, und im 4. Schritt kann wieder umgeordnet werden, weil alle Summanden ≥ 0 sind. Also existiert $\mathbb{E}(XY)$. Dieselbe Rechnung ohne Betragsstriche liefert dann $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, wobei Umordnen dieses Mal wegen absoluter Konvergenz zulässig ist. \square

Beispiele 4.7.

(a) (Indikatorfunktion)

Ist $X = \mathbf{1}_A$, so gilt $\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.

(b) (Binomialverteilung)

Ist $X \sim \mathcal{B}(n, p)$, so gilt $\mathbb{E}(X) = np$.

(c) (hypergeometrische Verteilung)

Ist $X \sim \mathcal{H}(n, r, s)$, so gilt $\mathbb{E}(X) = \frac{nr}{r+s}$.

(d) (Poisson-Verteilung)

Ist $X \sim \mathcal{P}(\lambda)$, so gilt $\mathbb{E}(X) = \lambda$.

(e) (geometrische Verteilung)

Ist $X \sim \mathcal{G}(p)$, so gilt $\mathbb{E}(X) = 1/p$.

Bemerkung: Teil (b) + (c) zeigt man am elegantesten dadurch, dass man X als Summe von Indikatorfunktionen darstellt und Satz 4.6 (a) verwendet:

- Teil (b): Wir betrachten das Modell

$$\Omega = \{0, 1\}^n, \quad f(\omega) = \prod_{i=1}^n p^{\omega_i} (1-p)^{1-\omega_i}.$$

Bezeichnet X_i den Indikator, ob die i -te Komponente = 1 ist, so gilt $\mathbb{E}(X_i) = p$ (Übung!), und $X := \sum_{i=1}^n X_i$ ist $\mathcal{B}(n, p)$ -verteilt mit

$$\mathbb{E}(X) \underset{\text{EW linear}}{=} \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n p = np.$$

- Teil (c): Wir betrachten das Modell

$$\Omega = \{\omega \in \{1, \dots, r+s\}^n : \omega_i \neq \omega_j \text{ für } i \neq j\}, \quad f(\omega) = 1/|\Omega|.$$

Bezeichnet X_i den Indikator, ob die i -te Komponente $\leq r$ ist, so gilt $\mathbb{E}(X_i) = \frac{r}{r+s}$ (Übung!), und $X := \sum_{i=1}^n X_i$ ist $\mathcal{H}(n, r, s)$ -verteilt mit

$$\mathbb{E}(X) \underset{\text{EW linear}}{=} \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n \frac{r}{r+s} = \frac{nr}{r+s}.$$

4.3 Der Erwartungswert von allgemeinen Zufallsgrößen

Wir führen nun den Erwartungswert für allgemeine Zufallsgrößen ein, wobei wir die nahe liegende Idee verwenden, allgemeine Zufallsgrößen durch diskrete Zufallsgrößen zu approximieren.

Sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsgröße, und sei für jedes $n \in \mathbb{N}$

$$X_{(n)} := \frac{1}{n} [nX]$$

wobei $[\cdot]$ die Gaußklammer-Funktion bezeichnet. Wir runden also auf das nächstkleinere ganzzahlige Vielfache von $\frac{1}{n}$ ab. Dann ist $X_{(n)}$ eine diskrete Zufallsgröße mit Werten in der (abzählbaren) Menge $\frac{1}{n}\mathbb{Z} := \{\frac{k}{n} : k \in \mathbb{Z}\}$, und es gilt

$$X_{(n)} \leq X \leq X_{(n)} + \frac{1}{n}.$$

Damit wissen wir schon, was wir unter $\mathbb{E}(X_{(n)})$ verstehen wollen. Es ist nun nahe liegend,

$$\mathbb{E}(X) := \lim_{n \rightarrow \infty} \mathbb{E}(X_{(n)}) \quad (*)$$

zu setzen. Dazu überlegt man sich Folgendes: Wenn $\mathbb{E}(X_{(n)})$ für irgendein $n \in \mathbb{N}$ existiert, so existiert $\mathbb{E}(X_{(n)})$ schon für alle $n \in \mathbb{N}$, und der Grenzwert in $(*)$ existiert. Darüber hinaus muss man sich überlegen, dass für diskrete Zufallsgrößen die „neue“ Definition des Erwartungswertes mit der „alten“ Definition des Erwartungswertes konsistent ist, so dass es nicht zu Definitionskonflikten kommt. Anschließend kann man vereinbaren:

Definition 4.8. Existiert in der obigen Situation $\mathbb{E}(X_{(n)})$ für irgendein $n \in \mathbb{N}$ und damit für alle $n \in \mathbb{N}$, so sagen wir, dass der Erwartungswert von X *existiert*, und nennen

$$\mathbb{E}X := \mathbb{E}(X) := \lim_{n \rightarrow \infty} \mathbb{E}(X_{(n)})$$

Erwartungswert von X .

Bemerkungen 4.9.

- (a) Ist X m. W. 1 beschränkt (d. h. existiert $K \in \mathbb{R}$ mit $\mathbb{P}(|X| \leq K) = 1$), so existiert der Erwartungswert von X immer.
- (b) Auch hier lassen sich wieder Erwartungswerte i. w. S. einführen, indem man für nicht-negative Zufallsgrößen X immer $\mathbb{E}(X) := \lim_{n \rightarrow \infty} \mathbb{E}(X_{(n)})$ setzt. Es gilt dann wieder: $\mathbb{E}(X)$ existiert genau dann, wenn $\mathbb{E}|X| < \infty$.
- (c) Auch Bemerkung 4.4 (e) gilt entsprechend. Insbesondere müssen wir also – ähnlich wie in Bem. 1.21 – nur \mathbb{P}_X kennen, um $\mathbb{E}(X)$ zu berechnen.

Die Berechnung von Erwartungswerten mit Hilfe von Definition 4.8 wäre sehr mühselig. Die meisten Zufallsgrößen, die in Anwendungen auftreten, besitzen allerdings diskrete oder stetige Verteilungen. Für Zufallsgrößen mit diskreten Verteilungen gelten die Formeln aus Kapitel 4.1 entsprechend. Bei Zufallsgrößen mit stetigen Verteilungen hilft der folgende Satz weiter:

Satz 4.10 (Transformationsformel für Zufallsgrößen mit stetigen Verteilungen).

Es seien $Y : \Omega \rightarrow \mathbb{R}^d$ eine Zufallsgröße mit einer stetigen Verteilung mit der Dichte f und $h : \mathbb{R}^d \rightarrow \mathbb{R}$ eine messbare Abbildung, so dass das Produkt $h \cdot f$ lokal Riemann-integrierbar (d. h. Riemann-integrierbar über beschränkten d -dimensionalen Intervallen) ist. Dann gilt

$$\mathbb{E}|h \circ Y| = \int_{\mathbb{R}^d} |h(y)| f(y) dy.$$

Also existiert $\mathbb{E}(h \circ Y)$ genau dann, wenn die rechte Seite endlich ist, und in diesem Fall gilt

$$\mathbb{E}(h \circ Y) = \int_{\mathbb{R}^d} h(y) f(y) dy.$$

*Beweisskizze.*¹⁰ Wir betrachten hier nur den Fall, dass $d = 1$ gilt und die Abbildung h stückweise monoton ist. Dann sind die Mengen $\{y : h_{(n)}(y) = z/n\}$ endliche Vereinigungen von Intervallen, und es folgt

$$\begin{aligned} \mathbb{E}|h \circ Y| &\approx \mathbb{E}|h_{(n)} \circ Y| \stackrel{\text{Satz 4.5}}{=} \sum_{z \in \mathbb{Z}} |z/n| \mathbb{P}(h_{(n)}(Y) = z/n) \stackrel{\text{Dichte}}{=} \\ &\sum_{z \in \mathbb{Z}} \int_{\{y: h_{(n)}(y)=z/n\}} |z/n| f(y) dy \stackrel{\text{Umordnen}}{=} \int_{\mathbb{R}} |h_{(n)}(y)| f(y) dy \approx \int_{\mathbb{R}} |h(y)| f(y) dy. \end{aligned}$$

und damit die erste Gleichung, wobei Umordnen zulässig ist, da der Integrand ≥ 0 ist. Sind die Ausdrücke endlich, so folgt die zweite Gleichung aus der gleichen Rechnung ohne Betragsstriche, wobei Umordnen nun wegen absoluter Konvergenz (des uneigentlichen Riemann-Integrals) zulässig ist. \square

Wählen wir in Satz 4.10 für eine reellwertige Zufallsgröße X mit einer stetigen Verteilung $Y := X$ und $h(x) := x$, so erhalten wir gerade die Erwartungswertformel für $\mathbb{E}(X)$ in Kapitel 4.1 zurück.

¹⁰ Dieser Beweis ist in der Vorlesung übersprungen worden.

Bemerkung (Berechnung von Erwartungswerten bei allgemeinen W.mäßen).

Ist X eine *nicht-negative* Zufallsgröße mit der Verteilungsfunktion F_X , so kann man den Erwartungswert von X (i. w. S.) auch mit der Formel

$$\mathbb{E}X = \int_0^\infty x \mathbb{P}(X > x) dx = \int_0^\infty x(1 - F_X(x)) dx$$

bestimmen. (Beachte, dass die Verteilungsfunktion einer reellwertigen Zufallsgröße immer existiert, wohingegen die Existenz der Dichte(funktion) einer reellwertigen Zufallsgröße eine Zusatzvoraussetzung darstellt!)

Ist X eine reellwertige Zufallsgröße mit einer diskreten oder stetigen Verteilung, so ist es aber i. d. R. einfacher, den Erwartungswert von X über die diskrete Dichte f_X bzw. die Riemann-Dichte f_X zu berechnen. Dieser Zugang¹¹ ist auch dann noch anwendbar, wenn \mathbb{P}_X ein „Mischtyp“ einer diskreten und einer stetigen Verteilung ist (vgl. dazu das dritte Beispiel auf Seite 40); hier besitzt der diskrete Anteil eine diskrete Dichte f_1 und der stetige Anteil eine Riemann-Dichte f_2 (wobei $\sum_{x \in \Omega_0} f_1(x) + \int_{\mathbb{R}} f_2(x) dx \stackrel{!}{=} 1$), und für „gutartige“ Funktionen $h : \mathbb{R} \rightarrow \mathbb{R}$ gilt (im Falle der Existenz)

$$\mathbb{E}(h(X)) = \sum_{x \in \Omega_0} h(x) f_1(x) + \int_{\mathbb{R}} h(x) f_2(x) dx.$$

□

Satz 4.11 (Rechenregeln für Erwartungswerte).

Die Rechenregeln für Erwartungswerte (vgl. den Satz in Kapitel 4.1) gelten für allgemeine Zufallsgrößen.

Beweisidee. Auch dies lässt sich durch diskrete Approximation und Zurückführung auf die entsprechenden Eigenschaften des diskreten Erwartungswertes beweisen. □

Beispiele 4.12.

(a) (Standardnormalverteilung)

Ist $Z \sim \mathcal{N}(0, 1)$, so gilt $\mathbb{E}(Z) = 0$.

(b) (Normalverteilung)

Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so gilt $\mathbb{E}(X) = \mu$. (Beachte $Z \sim \mathcal{N}(0, 1) \Rightarrow \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$ (vgl. Beispiel 2.22 (b)) und benutze Satz 4.11 (a).)

(c) (Exponentialverteilung)

Ist $X \sim \mathcal{E}(\lambda)$, so gilt $\mathbb{E}(X) = 1/\lambda$.

(d) (Cauchy-Verteilung)

Hat X die Riemann-Dichte $f(x) = \frac{1}{\pi(1+x^2)}$, so existiert $\mathbb{E}(X)$ nicht.

¹¹ Diese Anmerkung ist in der Vorlesung übersprungen worden.

Neben dem Erwartungswert gibt es weitere Kennzahlen, die den „mittleren“ Wert einer Zufallsgröße beschreiben:

Bemerkung 4.13 (Median). $x \in \mathbb{R}$ heißt *Median* von X , falls $\mathbb{P}(X \leq x) \geq \frac{1}{2}$ und $\mathbb{P}(X \geq x) \geq \frac{1}{2}$. Der Median existiert immer, ist aber nicht immer eindeutig bestimmt. (Betrachte die Stelle, an der die Verteilungsfunktion von X den Wert $\frac{1}{2}$ „überschreitet“.) Im Gegensatz zum Erwartungswert ist der Median „robust gegenüber Ausreißern“.

4.4 Varianz und Kovarianz

Oft interessieren nicht nur der Mittelwert, sondern auch die Abweichungen vom Mittelwert:

Beispiel 4.14 (Vergleich zweier Investitionen). Ein Investor hat die Auswahl zwischen zwei Investitionen:

• **Investition A:**

Kosten: 1000 €	Erlöse: m.W. $\frac{1}{2}$	700 €
	m.W. $\frac{1}{2}$	1700 €

• **Investition B:**

Kosten: 1000 €	Erlöse: m.W. $\frac{1}{2}$	900 €
	m.W. $\frac{1}{2}$	1400 €

Obwohl Investition A im Mittel zu einem größeren Erlös führt, wird ein (risikoscheuer) Investor zumindest darüber nachdenken, ob er nicht Investition B vorziehen soll. \square

Wir wollen daher Kennzahlen einführen, die die Abweichung vom Mittelwert beschreiben. Wir beginnen mit einer technischen Bemerkung, die die Existenz der nachfolgend auftretenden Erwartungswerte (in der Definition der Varianz bzw. der Kovarianz) sichert:

Bemerkung 4.15. Seien X und Y Zufallsgrößen, so dass $\mathbb{E}X^2$ und $\mathbb{E}Y^2$ endlich sind, und $a, b \in \mathbb{R}$. Dann sind auch $\mathbb{E}|X|$, $\mathbb{E}(X+b)^2$, $\mathbb{E}(aX)^2$, $\mathbb{E}(X+Y)^2$ und $\mathbb{E}|XY|$ endlich.

Beweis. Dies folgt aus den Ungleichungen $|x| \leq x^2+1$, $(x+b)^2 \leq 2x^2+2b^2$, $(ax)^2 \leq a^2x^2$, $(x+y)^2 \leq 2x^2+2y^2$ und $|xy| \leq x^2+y^2$ (für $x, y \in \mathbb{R}$) sowie den Eigenschaften des Erwartungswertes. \square

Definition 4.16 (Varianz). Für eine Zufallsgröße X mit $\mathbb{E}X^2 < \infty$ heißt

$$\text{Var}(X) := \mathbb{E}\left((X - \mathbb{E}X)^2\right)$$

Varianz von X und

$$\sigma(X) := \sqrt{\text{Var}(X)}$$

Standardabweichung von X .

Bemerkungen 4.17.

- (a) Wegen $\mathbb{E}X^2 < \infty$ ist $\text{Var } X$ nach Bemerkung 4.15 wohldefiniert und endlich.
- (b) Die Varianz von X hängt nur von der Verteilung von X ab.
- (c) Die Varianz von X ist ein Maß für die Streuung der Werte von X .

Satz 4.18 (Eigenschaften der Varianz). *Seien X und Y Zufallsgrößen mit $\mathbb{E}X^2 < \infty$ und $\mathbb{E}Y^2 < \infty$ und $a, b \in \mathbb{R}$. Dann gilt:*

- (a) $\text{Var}(X) \geq 0$, wobei $\text{Var}(X) = 0 \Leftrightarrow \exists c \in \mathbb{R} : \mathbb{P}(X = c) = 1$.
- (b) $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. (Verschiebungssatz)
- (c) $\text{Var}(X + b) = \text{Var}(X)$.
- (d) $\text{Var}(aX) = a^2 \text{Var}(X)$.
- (e) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$.

Warnung:

Im Allgemeinen gilt also $\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y)$.

Beweis von Satz 4.18. Wir schreiben (wie allgemein üblich) $\mathbb{E}W^2 := \mathbb{E}(W^2)$.

- (a) Es gilt $(X - \mathbb{E}X)^2 \geq 0$ und damit wegen der Monotonie des Erwartungswertes $\text{Var}(X) \geq 0$. Um die Charakterisierung der Gleichheit zu zeigen, benutzen wir, dass für jede Zufallsgröße $Z \geq 0$

$$\mathbb{E}(Z) = 0 \Leftrightarrow \mathbb{P}(Z = 0) = 1$$

gilt. (\Leftarrow ist klar, und \Rightarrow folgt für diskrete Zufallsgrößen aus der Definition des EW und für allgemeine Zufallsgrößen per Approximation.) Damit gilt

$$\text{Var}(X) = 0 \Leftrightarrow \mathbb{E}(X - \mathbb{E}X)^2 = 0 \Leftrightarrow \mathbb{P}(X = \mathbb{E}X) = 1 \Leftrightarrow (\exists c \in \mathbb{R}) \mathbb{P}(X = c) = 1.$$

- (b) $\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2 - 2\mathbb{E}(X)X + (\mathbb{E}(X))^2) \stackrel{\text{EW linear}}{=} \mathbb{E}X^2 - 2\mathbb{E}(X)\mathbb{E}(X) + (\mathbb{E}(X))^2 = \mathbb{E}X^2 - (\mathbb{E}(X))^2$.
- (c) $\text{Var}(X + b) = \mathbb{E}((X + b) - \mathbb{E}(X + b))^2 \stackrel{\text{EW linear}}{=} \mathbb{E}(X - \mathbb{E}X)^2 = \text{Var}(X)$.
- (d) $\text{Var}(aX) = \mathbb{E}(aX - \mathbb{E}(aX))^2 \stackrel{\text{EW linear}}{=} a^2\mathbb{E}(X - \mathbb{E}X)^2 = a^2\text{Var}(X)$.
- (e) $\text{Var}(X + Y) = \mathbb{E}((X + Y) - \mathbb{E}(X + Y))^2 \stackrel{\text{EW linear}}{=} \mathbb{E}((X - \mathbb{E}X) + (Y - \mathbb{E}Y))^2 = \mathbb{E}((X - \mathbb{E}X)^2 + (Y - \mathbb{E}Y)^2 + 2(X - \mathbb{E}X)(Y - \mathbb{E}Y)) \stackrel{\text{EW linear}}{=} \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y))$.

□

Definition 4.19 (Kovarianz). Sind X und Y Zufallsgrößen mit $\mathbb{E}X^2 < \infty$ und $\mathbb{E}Y^2 < \infty$, so heißt

$$\text{Cov}(X, Y) := \mathbb{E}\left((X - \mathbb{E}X)(Y - \mathbb{E}Y)\right)$$

Kovarianz von X und Y und, falls zusätzlich $\text{Var}(X) > 0$ und $\text{Var}(Y) > 0$ gilt,

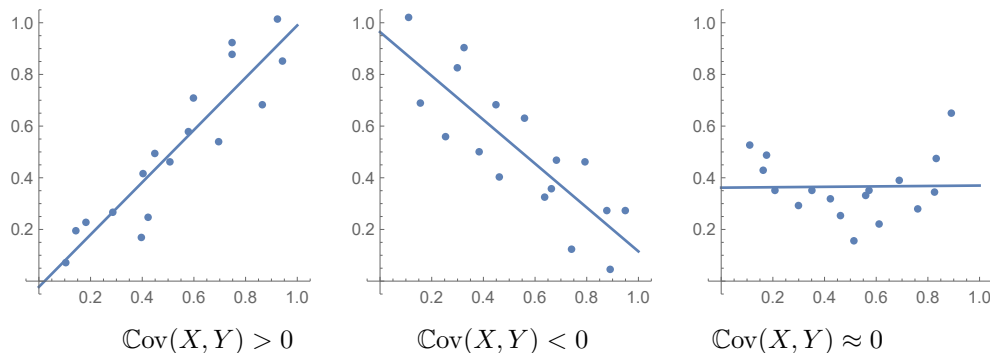
$$\varrho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var } X} \cdot \sqrt{\text{Var } Y}}$$

Korrelationskoeffizient von X und Y .

Ferner heißen X, Y $\left\{ \begin{array}{l} \text{positiv korreliert} \\ \text{unkorreliert} \\ \text{negativ korreliert} \end{array} \right\}$, falls $\left\{ \begin{array}{l} \text{Cov}(X, Y) > 0 \\ \text{Cov}(X, Y) = 0 \\ \text{Cov}(X, Y) < 0 \end{array} \right\}$ gilt.

Bemerkungen 4.20.

- (a) Wegen $\mathbb{E}X^2 < \infty$ und $\mathbb{E}Y^2 < \infty$ ist $\text{Cov}(X, Y)$ nach Bemerkung 4.15 wohldefiniert und endlich.
- (b) Die Kovarianz von X und Y hängt nur von der gemeinsamen Verteilung $\mathbb{P}_{(X, Y)}$ von X und Y ab.
- (c) Die Kovarianz von X und Y ist ein Maß für den linearen Zusammenhang der Werte von X und Y .



Beispiele zur Kovarianz. Wir betrachten jeweils die Gleichverteilung auf den dargestellten Punktwolken.

Die Eigenschaften der Kovarianz sind weitgehend analog zu denen der Varianz:¹²

Satz 4.21 (Eigenschaften der Kovarianz). Seien X, Y und Z Zufallsgrößen mit $\mathbb{E}X^2 < \infty$, $\mathbb{E}Y^2 < \infty$ und $\mathbb{E}Z^2 < \infty$ und $a, b, c, d \in \mathbb{R}$. Dann gilt:

- (a) $\text{Cov}(X, X) = \text{Var}(X)$.
- (b) $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. (Verschiebungssatz)

¹² Dieser Satz ist in der Vorlesung mit Ausnahme von Teil (b) übersprungen worden.

$$(c) \operatorname{Cov}(X + c, Y + d) = \operatorname{Cov}(X, Y).$$

$$(d) \operatorname{Cov}(aX, Y) = a \operatorname{Cov}(X, Y). \\ \operatorname{Cov}(X, bY) = b \operatorname{Cov}(X, Y).$$

$$(e) \operatorname{Cov}(X + Y, Z) = \operatorname{Cov}(X, Z) + \operatorname{Cov}(Y, Z). \\ \operatorname{Cov}(X, Y + Z) = \operatorname{Cov}(X, Y) + \operatorname{Cov}(X, Z).$$

$$(f) \operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X). \quad (\text{Symmetrie})$$

$$(g) |\operatorname{Cov}(X, Y)| \leq \sqrt{\operatorname{Var}(X)} \sqrt{\operatorname{Var}(Y)}. \quad (\text{Cauchy-Schwarz-Ungleichung}) \\ \text{Also gilt } |\rho(X, Y)| \leq 1, \text{ falls } \operatorname{Var}(X) > 0, \operatorname{Var}(Y) > 0.$$

Beweis. (a) – (f): Übung! – (g): Nach Teil (a), (d), (e), (f) ist $(V, W) \mapsto \operatorname{Cov}(V, W)$ eine positiv-semidefinite symmetrische Bilinearform auf dem Vektorraum aller Zufallsgrößen U mit $\mathbb{E}(U^2) < \infty$. Damit folgt die Behauptung wie für das euklidische Skalarprodukt in der Linearen Algebra. \square

Bemerkung (Berechnung der Kovarianz). Nach Satz 4.21 (b) genügt es, zur Berechnung von $\operatorname{Cov}(X, Y)$ $\mathbb{E}(XY)$, $\mathbb{E}(X)$ und $\mathbb{E}(Y)$ zu berechnen; dabei gilt im diskreten Fall

$$\mathbb{E}(XY) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} xy \mathbb{P}(X = x, Y = y)$$

(wobei \mathcal{X}, \mathcal{Y} abzählbar mit $\mathbb{P}((X, Y) \in \mathcal{X} \times \mathcal{Y}) = 1$) und im stetigen Fall

$$\mathbb{E}(XY) = \int_{\mathbb{R}^2} xy f(x, y) d(x, y).$$

(wobei f die gemeinsame Riemann-Dichte von X und Y bezeichnet).

Lemma 4.22 (Unabhängigkeit und Unkorreliertheit). *Seien X und Y Zufallsgrößen mit $\mathbb{E}X^2 < \infty$ und $\mathbb{E}Y^2 < \infty$. Sind X und Y unabhängig, so sind X und Y unkorreliert.*

Beweis. Nach dem Multiplikationssatz für unabhängige Zufallsgrößen (Satz 4.11 (d)) gilt $\operatorname{Cov}(X, Y) \stackrel{\text{Vers.satz}}{=} \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$. \square

Warnung:

Die Umkehrung von Lemma 4.22 ist im Allgemeinen falsch.

Bemerkung 4.23. Sind X und Y Zufallsgrößen mit $\mathbb{E}X^2 < \infty$ und $\mathbb{E}Y^2 < \infty$, so gilt nach Satz 4.18 (e)

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \operatorname{Cov}(X, Y).$$

Sind X und Y unkorreliert, so gilt damit

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y).$$

Allgemeiner gilt: Sind X_1, \dots, X_n Zufallsgrößen mit $\mathbb{E}X_i^2 < \infty$ für alle $i = 1, \dots, n$, so gilt

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}\text{ar}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathbb{C}\text{ov}(X_i, X_j).$$

Sind X_1, \dots, X_n paarweise unkorreliert, so gilt damit

$$\mathbb{V}\text{ar}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}\text{ar}(X_i).$$

Bemerkung 4.24 (Standardisierung). Ist X eine Zufallsgröße mit $\mathbb{E}X^2 < \infty$ und $\mathbb{V}\text{ar}(X) > 0$, so hat $X^* := (X - \mathbb{E}X)/\sqrt{\mathbb{V}\text{ar}(X)}$ Erwartungswert 0 und Varianz 1. X^* heißt *Standardisierung* von X .

Beispiele 4.25.

(a) (Indikatorfunktion)

Ist $X = \mathbf{1}_A$, so gilt $\mathbb{V}\text{ar}(\mathbf{1}_A) = \mathbb{P}(A)(1 - \mathbb{P}(A))$.

(b) (Binomialverteilung)

Ist $X \sim \mathcal{B}(n, p)$, so gilt $\mathbb{V}\text{ar}(X) = np(1 - p)$.

(c) (hypergeometrische Verteilung)

Ist $X \sim \mathcal{H}(n, r, s)$, so gilt $\mathbb{V}\text{ar}(X) = \frac{nrs(r + s - n)}{(r + s)(r + s)(r + s - 1)}$.

(d) (Poisson-Verteilung)

Ist $X \sim \mathcal{P}(\lambda)$, so gilt $\mathbb{V}\text{ar}(X) = \lambda$.

(e) (geometrische Verteilung)

Ist $X \sim \mathcal{G}(p)$, so gilt $\mathbb{E}(X) = (1 - p)/p^2$.

(f) (Standardnormalverteilung)

Ist $X \sim \mathcal{N}(0, 1)$, so gilt $\mathbb{V}\text{ar}(X) = 1$.

(g) (Normalverteilung)

Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so gilt $\mathbb{V}\text{ar}(X) = \sigma^2$.

(Beachte $Z \sim \mathcal{N}(0, 1) \Rightarrow \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$ sowie Satz 4.18 (c) + (d).)

(h) (Exponentialverteilung)

Ist $X \sim \mathcal{E}(\lambda)$, so gilt $\mathbb{V}\text{ar}(X) = 1/\lambda^2$.

Bemerkung: Teil (b) + (c) zeigt man am elegantesten dadurch, dass man X als Summe von Indikatorfunktionen darstellt und Bemerkung 4.23 verwendet:

- Teil (b): Wir betrachten wieder das Modell

$$\Omega = \{0, 1\}^n, \quad f(\omega) = \prod_{i=1}^n p^{\omega_i} (1 - p)^{1 - \omega_i}.$$

Bezeichnet X_i den Indikator, ob die i -te Komponente = 1 ist, so ist $X := \sum_{i=1}^n X_i$ $\mathcal{B}(n, p)$ -verteilt mit

$$\text{Var}(X) \underset{\text{Unabh.}}{=} \sum_{i=1}^n \text{Var}(X_i) = np(1-p).$$

- Teil (c): Wir betrachten wieder das Modell

$$\Omega = \{\omega \in \{1, \dots, r+s\}^n : \omega_i \neq \omega_j \text{ für } i \neq j\}, \quad f(\omega) = 1/|\Omega|.$$

Bezeichnet X_i den Indikator, ob die i -te Komponente $\leq r$ ist, so ist $X := \sum_{i=1}^n X_i$ $\mathcal{H}(n, r, s)$ -verteilt mit

$$\text{Var}(X) = \sum_{i=1, \dots, n} \text{Var}(X_i) + \sum_{\substack{i, j=1, \dots, n \\ i \neq j}} \text{Cov}(X_i, X_j) = \frac{nrs}{(r+s)^2} - \frac{n(n-1)rs}{(r+s)^2(r+s-1)}.$$

Zusammenfassung:

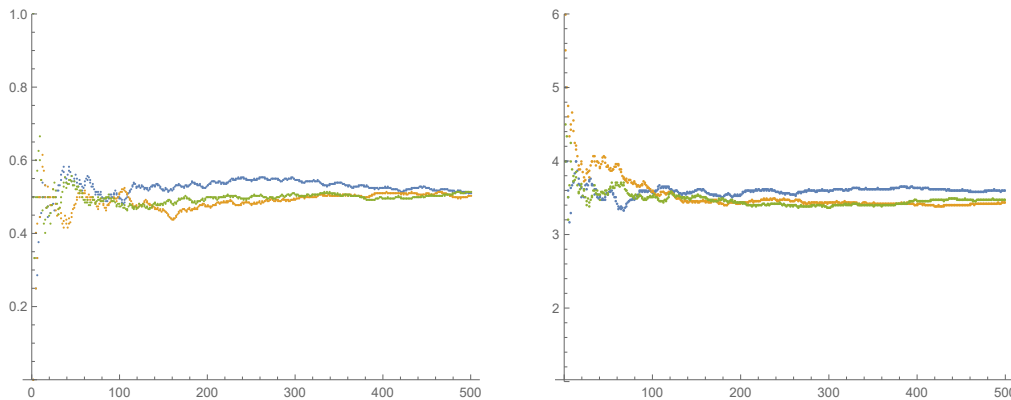
- Wichtige Kennzahlen zur Beschreibung von Zufallsgrößen sind der Erwartungswert (als Lagemaß) und die Varianz (als Streuungsmaß).
- Für eine Zufallsgröße X mit einer diskreten Verteilung gilt (im Falle der Existenz) $\mathbb{E}X = \sum x \mathbb{P}(X=x)$, wobei die Summe über die Menge $\{x \in \mathbb{R} : \mathbb{P}(X=x) > 0\}$ zu bilden ist.
- Für eine Zufallsgröße X mit einer stetigen Verteilung gilt (im Falle der Existenz) $\mathbb{E}X = \int_{-\infty}^{+\infty} x f(x) dx$, wobei f die Riemann-Dichte von X ist.
- Allgemeiner lassen sich mit den „Transformationsformeln“ Erwartungswerte von Funktionen von Zufallsgrößen berechnen.
- Es gilt $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}X)^2) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.
- Es gilt $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.
- Sind X und Y unabhängig, so gilt $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$.
- Es gilt $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$.
- Sind X und Y unabhängig, so auch unkorreliert.
- Wir haben Erwartungswerte und Varianzen einiger wichtiger Verteilungen bestimmt.

5 Grenzwertsätze für Summen von u.i.v. Zufallsgrößen

5.1 Das Gesetz der großen Zahlen

Motivation

Ein Zufallsexperiment wird mehrmals („unabhängig voneinander“) wiederholt, wobei jede Durchführung zu einer reellwertigen Beobachtung führt. Die Erfahrung lehrt uns, dass bei einer zunehmenden Anzahl von Versuchswiederholungen eine Stabilisierung des (empirischen) Mittelwertes der Beobachtungen erfolgt:



Entwicklung des Mittelwertes beim Wurf einer fairen Münze (Kopf \triangleq 1, Zahl \triangleq 0)
bzw. beim Wurf eines fairen Würfels (Augenzahl) für jeweils 3 Versuchsreihen

Wir wollen nun zeigen, dass dieser Erfahrungstatsache im Rahmen unseres mathematischen Modells ein *Grenzwertsatz* für Summen von unabhängigen, identisch verteilten (u.i.v.) Zufallsgrößen entspricht. Seien dazu X_1, X_2, X_3, \dots u.i.v. reellwertige Zufallsgrößen mit Erwartungswert μ und Varianz σ^2 auf einem W.raum $(\Omega, \mathcal{A}, \mathbb{P})$, vgl. Satz 3.23, und seien S_1, S_2, S_3, \dots die zugehörigen Summen. Dann gilt nach den Rechenregeln für Erwartungswert und Varianz

$$\mathbb{E}(S_n) \underset{\text{EW linear}}{=} \sum_{i=1}^n \mathbb{E}(X_i) = n\mu \quad \text{und} \quad \text{Var}(S_n) \underset{\text{Unabhängigkeit}}{=} \sum_{i=1}^n \text{Var}(X_i) = n\sigma^2$$

und damit

$$\mathbb{E}(S_n/n) = \mu \quad \text{und} \quad \text{Var}(S_n/n) = \sigma^2/n.$$

Man könnte also vermuten, dass sich die Zufallsgrößen S_n/n für $n \rightarrow \infty$ um ihren Mittelwert μ „stabilisieren“. Das *Gesetz der großen Zahlen* wird diese Vermutung – allerdings nur in einem eingeschränkten Sinne – bestätigen.

Bezeichnung. Ist X eine reellwertige Zufallsgröße und ist $r > 0$ eine reelle Zahl, so schreiben wir $X \in \mathcal{L}^r$, falls $\mathbb{E}|X|^r < \infty$.

Satz 5.1 (Markoff-Ungleichung). *Sei X eine nicht-negative Zufallsgröße. Dann gilt für alle $x > 0$*

$$\mathbb{P}(X \geq x) \leq \mathbb{E}(X)/x.$$

Beweis.

$$\begin{aligned} \mathbb{P}(X \geq x) &\stackrel{\text{EW Indikatorfunktion}}{=} \mathbb{E}(\mathbf{1}_{\{X \geq x\}}) \stackrel{\text{EW monoton}}{\leq} \mathbb{E}\left(\frac{X}{x} \mathbf{1}_{\{X \geq x\}}\right) \\ &\stackrel{\text{EW linear}}{=} \frac{\mathbb{E}(X \mathbf{1}_{\{X \geq x\}})}{x} \stackrel{\text{EW monoton}}{\leq} \frac{\mathbb{E}(X)}{x}. \quad \square \end{aligned}$$

Satz 5.2 (Tschebyscheff-Ungleichung). *Sei Y eine reellwertige Zufallsgröße mit $\mathbb{E}Y^2 < \infty$. Dann gilt für alle $t > 0$*

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq t) \leq \text{Var}(Y)/t^2.$$

Beweis. Da die Funktion $y \mapsto y^2$ auf $[0, \infty[$ streng monoton wachsend ist, gilt

$$\mathbb{P}(|Y - \mathbb{E}Y| \geq t) = \mathbb{P}(|Y - \mathbb{E}Y|^2 \geq t^2) \stackrel{\text{Satz 5.1}}{\leq} \frac{\mathbb{E}|Y - \mathbb{E}Y|^2}{t^2} = \frac{\text{Var } Y}{t^2}. \quad \square$$

Bemerkungen.

- (a) Es gibt Fälle, in denen in der Markoff/Tschebyscheff-Ungleichung die Gleichheit gilt; die Ungleichung lässt sich im Allgemeinen also nicht verbessern.
- (b) Es gibt Fälle, in denen die Markoff/Tschebyscheff-Ungleichung sehr ungenau bzw. völlig unbrauchbar ist, etwa wenn die rechte Seite ≥ 1 ist.

Satz 5.3 (Schwaches Gesetz der großen Zahlen). *Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsgrößen in \mathcal{L}^2 mit Erwartungswert μ und Varianz σ^2 . Dann gilt für jedes $\varepsilon > 0$*

$$\mathbb{P}(|(\frac{1}{n} \sum_{i=1}^n X_i) - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad \forall n \in \mathbb{N},$$

insbesondere also

$$\lim_{n \rightarrow \infty} \mathbb{P}(|(\frac{1}{n} \sum_{i=1}^n X_i) - \mu| \geq \varepsilon) = 0. \quad (5.1)$$

Beweis. Nach Bemerkung 4.15 liegt mit X_1, \dots, X_n auch $\frac{1}{n} \sum_{i=1}^n X_i$ in \mathcal{L}^2 . Damit folgt für alle $\varepsilon > 0$ und alle $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}(|(\frac{1}{n} \sum_{i=1}^n X_i) - \mu| \geq \varepsilon) &\stackrel{\text{Rechenregeln EW}}{=} \mathbb{P}(|(\frac{1}{n} \sum_{i=1}^n X_i) - \mathbb{E}(\frac{1}{n} \sum_{i=1}^n X_i)| \geq \varepsilon) \\ &\stackrel{\text{Tschebyscheff}}{\leq} \frac{\text{Var}(\frac{1}{n} \sum_{i=1}^n X_i)}{\varepsilon^2} \stackrel{\text{Rechenregeln Varianz Unabhängigkeit}}{=} \frac{\sum_{i=1}^n \text{Var}(X_i)}{n^2 \varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \end{aligned}$$

und damit die Behauptung. \square

Satz 5.4 (Starkes Gesetz der großen Zahlen). *Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsgrößen in \mathcal{L}^2 mit Erwartungswert μ und Varianz σ^2 . Dann gilt*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \mu\right) = 1. \quad (5.2)$$

*Beweis.*¹³ Setze $Z_n := \sum_{i=1}^n (X_i - \mathbb{E}X_i)$, $n \in \mathbb{N}$. Dann ist zu zeigen: $\mathbb{P}(\lim_{n \rightarrow \infty} Z_n/n = 0) = 1$.

Sei $k \in \mathbb{N}$ beliebig. Betrachte zunächst die Teilfolge $(Z_{n^2}/n^2)_{n \in \mathbb{N}}$. Dann gilt nach Tschebyscheff (ähnlich wie im Beweis von Satz 5.3)

$$\sum_{n=1}^{\infty} \mathbb{P}(|Z_{n^2}/n^2| > \frac{1}{2k}) \leq \sum_{n=1}^{\infty} \frac{4k^2 \text{Var}(Z_{n^2})}{n^4} = \sum_{n=1}^{\infty} \frac{4k^2 n^2 \sigma^2}{n^4} = \sum_{n=1}^{\infty} \frac{4k^2 \sigma^2}{n^2} < \infty$$

(da $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$) und damit mit $A_n := \{|Z_{n^2}/n^2| > \frac{1}{2k}\}$

$$\begin{aligned} \mathbb{P}(\underbrace{A_n \text{ unendlich oft}}_{=:A}) &= \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\right) \\ &\stackrel{\text{Stetigkeit von oben}}{=} \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty} A_n\right) \stackrel{\sigma\text{-Subadditivität}}{\leq} \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n) = 0. \end{aligned}$$

Betrachte nun die Folge $((Z_n - Z_{m(n)})/n)_{n \in \mathbb{N}}$, wobei für jedes $n \in \mathbb{N}$ $m(n)$ die größte Quadratzahl $\leq n$ bezeichnet. Dann gilt $n < (\sqrt{m(n)} + 1)^2 = m(n) + 2\sqrt{m(n)} + 1$ bzw. $n - m(n) \leq 2\sqrt{m(n)} \leq 2\sqrt{n}$. Sei $u > 0$ beliebig. Dann folgt nach Tschebyscheff (nochmals ähnlich wie im Beweis von Satz 5.3)

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}(|(Z_n - Z_{m(n)})/n| > \frac{1}{2k}) &\leq \sum_{n=1}^{\infty} \frac{4k^2 \text{Var}(Z_n - Z_{m(n)})}{n^2} \\ &= \sum_{n=1}^{\infty} \frac{4k^2 (n - m(n)) \sigma^2}{n^2} \leq \sum_{n=1}^{\infty} \frac{8k^2 \sigma^2}{n^{3/2}} < \infty \end{aligned}$$

(da $\sum_{n=1}^{\infty} \frac{1}{n^{3/2}} < \infty$) und damit mit $B_n := \{|(Z_n - Z_{m(n)})/n| > \frac{1}{2k}\}$

$$\begin{aligned} \mathbb{P}(\underbrace{B_n \text{ unendlich oft}}_{=:B}) &= \mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} B_n\right) \\ &\stackrel{\text{Stetigkeit von oben}}{=} \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcup_{n=m}^{\infty} B_n\right) \stackrel{\sigma\text{-Subadditivität}}{\leq} \lim_{m \rightarrow \infty} \sum_{n=m}^{\infty} \mathbb{P}(B_n) = 0. \end{aligned}$$

¹³ Dieser schwierigere Beweis ist in der Vorlesung übersprungen worden.

Nun gilt mit $\mathbb{P}(A) = 1$ und $\mathbb{P}(B) = 1$ auch $\mathbb{P}(A \cap B) = 1$ (*Übung!*), und für alle $\omega \in A \cap B$ existiert ein $N(\omega)$ mit $|Z_n(\omega)/n^2| \leq \frac{1}{2k} \quad \forall n \geq N(\omega)$ und $|Z_n(\omega) - Z_{m(n)}(\omega)|/n \leq \frac{1}{2k} \quad \forall n \geq N(\omega)$ und damit auch

$$\left| \frac{Z_n(\omega)}{n} \right| \leq \frac{m(n)}{n} \left| \frac{Z_{m(n)}(\omega)}{m(n)} \right| + \left| \frac{Z_n(\omega) - Z_{m(n)}(\omega)}{n} \right| \leq \frac{1}{k} \quad \forall n \geq N^2(\omega).$$

Also gilt für jedes $k \in \mathbb{N}$

$$\mathbb{P}(\{|Z_n/n| > \frac{1}{k} \text{ unendlich oft}\}) = 0$$

und damit (mittels σ -Subadditivität)

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} \{|Z_n/n| > \frac{1}{k} \text{ unendlich oft}\}\right) = 0,$$

woraus durch Übergang zur Gegen-W. $\mathbb{P}(\lim_{n \rightarrow \infty} Z_n/n = 0) = 1$ folgt. \square

Bemerkungen 5.5.

- (a) Das WLLN besagt, dass die Wahrscheinlichkeit dafür, dass $\frac{1}{n} \sum_{i=1}^n X_i$ von μ um mindestens ε abweicht, für jedes $\varepsilon > 0$ gegen Null konvergiert. Dies ist eine schwache Konvergenzaussage, die auch als *stochastische Konvergenz* bekannt ist:

Definition. Eine Folge $(X_n)_{n \in \mathbb{N}}$ von reellwertigen Zufallsgrößen konvergiert *stochastisch* gegen eine reellwertige Zufallsgröße X [kurz: $X_n \xrightarrow{p} X$], wenn für jedes $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

Damit kann man (5.1) auch wie folgt formulieren:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu.$$

Das WLLN besagt nicht, dass $\frac{1}{n} \sum_{i=1}^n X_i$ im Sinne der Analysis gegen μ konvergiert!

- (b) Das SLLN besagt, dass $\frac{1}{n} \sum_{i=1}^n X_i$ im Sinne der Analysis gegen μ konvergiert, *sofern* wir uns auf eine (geeignete) Menge $\Omega^* \subseteq \Omega$ mit $\mathbb{P}(\Omega^*) = 1$ zurückziehen bzw. eine (geeignete) Ausnahmemenge m. W. 0 von der Betrachtung ausschließen. Dies ist eine starke Konvergenzaussage, die auch als *fast sichere Konvergenz* bekannt ist:

Definition. Eine Folge $(X_n)_{n \in \mathbb{N}}$ von reellwertigen Zufallsgrößen konvergiert *fast sicher* gegen eine reellwertige Zufallsgröße X [kurz: $X_n \xrightarrow{f.s.} X$], wenn gilt:

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

Damit kann man (5.2) auch wie folgt formulieren:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{f.s.} \mu.$$

- (c) Man kann zeigen: Aus fast sicherer Konvergenz folgt stochastische Konvergenz; die Umkehrung ist im Allgemeinen falsch. Die Aussage (5.2) ist also stärker als die Aussage (5.1), wie die Namensgebung (starkes Gesetz bzw. schwaches Gesetz) vermuten lässt.
- (d) Die Aussagen (5.1) und (5.2) bleiben unter der schwächeren Voraussetzung gültig, dass die u.i.v. Zufallsgrößen X_n (nur) in \mathcal{L}^1 liegen. Auch die u.i.v.-Voraussetzung lässt sich noch abschwächen.

Besonders wichtig ist der folgende Spezialfall von Satz 5.3 und Satz 5.4, der auch als *Bernoulli'sches Gesetz der großen Zahlen* bekannt ist: Sind X_1, X_2, X_3, \dots u.i.v. Zufallsgrößen mit Werten in $(\mathcal{X}, \mathcal{B})$ und ist $B \in \mathcal{B}$ beliebig, so gilt

$$\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k \in B\}} \xrightarrow{p} \mathbb{P}(X_1 \in B)$$

bzw.

$$\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k \in B\}} \xrightarrow{f.s.} \mathbb{P}(X_1 \in B),$$

d. h. die (theoretischen) relativen Häufigkeiten von B konvergieren stochastisch bzw. fast sicher gegen die W. von B . Beachte dazu, dass die Zufallsgrößen $Y_i := \mathbf{1}_{\{X_i \in B\}}$ nach dem Prinzip von der getrennten Verarbeitung (Satz 3.21) wieder u.i.v. mit $\mathbb{E}(Y_1) = \mathbb{E}(\mathbf{1}_{\{X_1 \in B\}}) = \mathbb{P}(X_1 \in B)$ und $\text{Var}(Y_1) = \text{Var}(\mathbf{1}_{\{X_1 \in B\}}) = \mathbb{P}(X_1 \in B)(1 - \mathbb{P}(X_1 \in B))$ sind.

5.2 Der zentrale Grenzwertsatz

Motivation

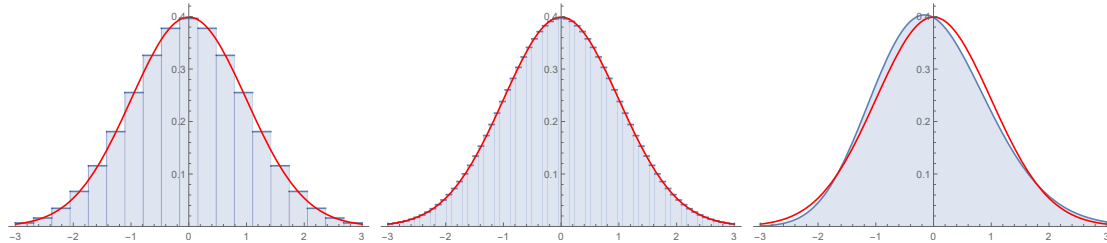
Es seien X_1, X_2, X_3, \dots u.i.v. reellwertige Zufallsgrößen mit Erwartungswert μ und Varianz σ^2 und S_1, S_2, S_3, \dots die zugehörigen Summen. Wir wollen die *Abweichungen* der Zufallsgrößen S_n von ihren Erwartungswerten $\mathbb{E}(S_n) = n\mu$ untersuchen. Nach dem schwachen Gesetz der großen Zahlen gilt $S_n/n \xrightarrow{p} \mu$; Abweichungen von der Größenordnung n treten für großes $n \in \mathbb{N}$ also „typischerweise“ nicht auf. Es stellt sich nun die Frage, von welcher Größenordnung „typische“ Abweichungen sind bzw. ob es sogar eine Skalierung gibt, so dass sich die *Verteilung* dieser Abweichungen „stabilisiert“.

Wegen $\mathbb{E}(S_n) = n\mu$ und $\text{Var}(S_n) = n\sigma^2$ gilt nach Bemerkung 4.24

$$\mathbb{E}((S_n - n\mu)/\sqrt{n\sigma^2}) = 0 \quad \text{und} \quad \text{Var}((S_n - n\mu)/\sqrt{n\sigma^2}) = 1,$$

wobei $\sigma^2 > 0$ vorauszusetzen ist. Man könnte also hoffen, dass sich die Verteilung der *standardisierten* Summen $S_n^* := (S_n - n\mu)/\sqrt{n\sigma^2}$ (vgl. Bemerkung 4.24) für $n \rightarrow \infty$ einer Grenzverteilung „annähert“. Dass dies in der Tat richtig ist, zeigt der *zentrale Grenzwertsatz*.

Inspizieren wir die Dichten der standardisierten Summen S_n^* für u.i.v. Zufallsgrößen X_n mit diskreten Verteilungen (z. B. der Bernoulli-Verteilung) oder stetigen Verteilungen (z. B. der Exponentialverteilung), so scheint die Dichte der Standardnormalverteilung schon für moderate Werte von n (z. B. $n = 40$) eine gute Approximation zu liefern:



Verteilung der standardisierten Summe S_{40}^* für $X_i \sim \mathcal{B}(1, 0.5)$ („Münzwurf“),
 $X_i \sim \mathcal{U}(\{1, 2, 3, 4, 5, 6\})$ („Würfelwurf“) bzw. $X_i \sim \mathcal{E}(1)$ („Wartezeit“)

Problem. Im Allgemeinen brauchen die Zufallsgrößen S_n^* keine Dichten zu besitzen.

Lösung. Wir definieren Verteilungskonvergenz über Verteilungsfunktionen, wobei allerdings noch eine Zusatzbedingung ins Spiel kommt ...

Definition 5.6 (Verteilungskonvergenz). Es seien X, X_1, X_2, X_3, \dots reellwertige Zufallsgrößen mit Verteilungsfunktionen F, F_1, F_2, F_3, \dots . Dann konvergiert X_n in Verteilung gegen X [kurz: $X_n \xrightarrow{d} X$ oder $X_n \Rightarrow X$], wenn für jede Stetigkeitsstelle $x \in \mathbb{R}$ von F

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

gilt.

Bemerkungen 5.7 (Bemerkungen zur Verteilungskonvergenz).

- (a) Für $X := 0$ und $X_n := 1/n$ gilt $X_n \rightarrow X$, obwohl $F_n(0) = 0 \not\rightarrow 1 = F(0)$ gilt. Dieses Beispiel zeigt, dass es unzumutbar wäre, die Konvergenz der Verteilungsfunktionen an jeder Stelle $x \in \mathbb{R}$ zu fordern.
- (b) Aus $X_n \xrightarrow{p} X$ folgt $X_n \xrightarrow{d} X$; die Umkehrung ist im Allgemeinen falsch. (Sie ist allerdings richtig, wenn X fast sicher konstant ist.)
- (c) Es gilt $X_n \xrightarrow{d} X$ genau dann, wenn für alle $-\infty \leq a \leq b \leq +\infty$ mit $\mathbb{P}(X = a) = 0 = \mathbb{P}(X = b)$ $\mathbb{P}(a < X_n \leq b) \rightarrow \mathbb{P}(a < X \leq b)$ gilt. (Zudem dürfen wir hier $<$ durch \leq oder \leq durch $<$ ersetzen.)

Satz 5.8 (Zentraler Grenzwertsatz). Es sei $(X_n)_{n \in \mathbb{N}}$ eine Folge unabhängiger, identisch verteilter Zufallsgrößen in \mathcal{L}^2 mit Erwartungswert μ und Varianz $\sigma^2 \in (0, \infty)$. Dann gilt für die Summen $S_n := \sum_{i=1}^n X_i$

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z,$$

wobei Z eine standardnormalverteilte Zufallsgröße bezeichnet.

Es ist bemerkenswert, dass stets die Normalverteilung herauskommt, unabhängig davon, mit welcher Verteilung für die X_n (mit Erwartungswert μ und Varianz $\sigma^2 \in (0, \infty)$) man startet. Diese Eigenschaft wird auch als *Universalität der Normalverteilung* bezeichnet.

Ähnlich wie im Gesetz der großen Zahlen lassen sich die Voraussetzungen auch hier noch (etwas) abschwächen: Addieren sich viele kleine unabhängige Zufallseffekte zu einem Gesamteffekt, so tritt näherungsweise die Normalverteilung auf. Dies erklärt (zumindest teilweise) das häufige Auftreten der Normalverteilung in Anwendungen.

Auf den Beweis des zentralen Grenzwertsatzes müssen wir im Rahmen dieser Vorlesung leider verzichten.

Da die Normalverteilung (als W.maß mit einer Riemann-Dichte) eine stetige Verteilungsfunktion besitzt, gilt in der Situation von Satz 5.8

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(a < \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq b\right) = \mathbb{P}(a < Z \leq b) = \int_a^b \varphi(x) dx = \Phi(b) - \Phi(a)$$

für alle $-\infty \leq a \leq b \leq +\infty$, wobei φ und Φ die Dichte(funktion) bzw. die Verteilungsfunktion der Standardnormalverteilung bezeichnen. (Auch hier können wir $<$ durch \leq oder \leq durch $<$ ersetzen.)

Bemerkungen:

- Φ lässt sich nicht „elementar“ angeben, ist aber in Tabellen vertafelt bzw. vom Computer berechenbar.
- Es gilt $\Phi(0) = \frac{1}{2}$ und $\Phi(-x) = 1 - \Phi(x)$ für alle $x > 0$.
- Für $a = -\infty$ bzw. $b = +\infty$ ist $\Phi(-\infty) := 0$ bzw. $\Phi(+\infty) := 1$ zu setzen.

Normalapproximation der Verteilung von Summen

Aufgrund von Satz 5.8 ist es nahe liegend, für eine Summe $S_n = \sum_{i=1}^n X_i$ wie in Satz 5.8 und $a, b \in \mathbb{R} \cup \{\pm\infty\}$ mit $-\infty \leq a \leq b \leq +\infty$ die *Normalapproximation*

$$\mathbb{P}(a \leq S_n \leq b) = \mathbb{P}\left(\frac{a-n\mu}{\sqrt{n\sigma^2}} \leq S_n^* \leq \frac{b-n\mu}{\sqrt{n\sigma^2}}\right) \approx \int_{(a-n\mu)/\sqrt{n\sigma^2}}^{(b-n\mu)/\sqrt{n\sigma^2}} \varphi(x) dx = \Phi\left(\frac{b-n\mu}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{a-n\mu}{\sqrt{n\sigma^2}}\right)$$

zu verwenden. Es stellt sich die Frage, wie groß n mindestens sein muss, um eine brauchbare Näherung für die gesuchte W. zu erhalten. Die Antwort hängt natürlich von der benötigten Genauigkeit (die sich aus der jeweiligen Anwendung ergibt) sowie auch von der Verteilung der X_i ab. Für „einfache“ Anwendungen wird in der Literatur die folgende Regel vorgeschlagen:

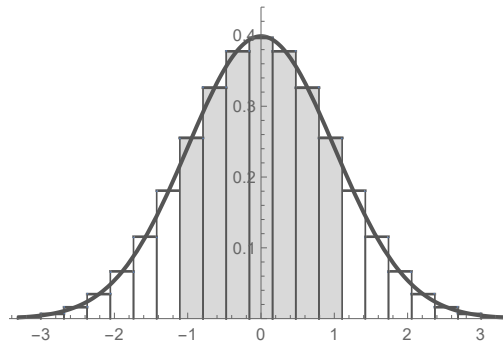
Faustregel:

Die Normalapproximation ist brauchbar, wenn $n \geq 30$ gilt.

Liegen *ganzzahlige* Zufallsgrößen X_1, X_2, X_3, \dots vor, so lässt sich die Approximation für *ganzzahlige* Werte $a, b \in \mathbb{Z} \cup \{\pm\infty\}$ oft noch verbessern, indem man die *Normalapproximation mit Stetigkeitskorrektur*

$$\mathbb{P}(a \leq S_n \leq b) = \mathbb{P}\left(\frac{a-n\mu}{\sqrt{n\sigma^2}} \leq S_n^* \leq \frac{b-n\mu}{\sqrt{n\sigma^2}}\right) \approx \int_{(a-n\mu-\frac{1}{2})/\sqrt{n\sigma^2}}^{(b-n\mu+\frac{1}{2})/\sqrt{n\sigma^2}} \varphi(x) dx = \Phi\left(\frac{b-n\mu+\frac{1}{2}}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{a-n\mu-\frac{1}{2}}{\sqrt{n\sigma^2}}\right)$$

verwendet. Dies kann man dadurch plausibel machen, dass man die Elementar-W.en von S_n^* durch ein „Histogramm“ darstellt und feststellt, dass $\mathbb{P}(a \leq S_n \leq b)$ der Summe der Rechteckflächen zwischen der Stelle $\frac{a-n\mu-(1/2)}{\sqrt{n\sigma^2}}$ und der Stelle $\frac{b-n\mu+(1/2)}{\sqrt{n\sigma^2}}$ entspricht:



Veranschaulichung der Verteilung von S_n^* für $S_n \sim \mathcal{B}(n, p)$ mit $n=40, p=0.5$;
hier gilt $\frac{a-n\mu-(1/2)}{\sqrt{n\sigma^2}} \approx -1.11$, $\frac{a-n\mu}{\sqrt{n\sigma^2}} \approx -0.95$, $\frac{b-n\mu}{\sqrt{n\sigma^2}} \approx +0.95$, $\frac{b-n\mu+(1/2)}{\sqrt{n\sigma^2}} \approx +1.11$.

Beispiel 5.9 (Anwendungsbeispiele zur Normalapproximation).

- (a) Wie groß ist die W., dass beim 50maligen Wurf eines fairen Würfels die Summe der Augenzahlen zwischen 160 und 190 liegt?
- (b) Man bestimme ein möglichst kleines Intervall, in dem die Summe der Augenzahlen beim 50maligen Wurf eines fairen Würfels mit einer W. von mindestens 90 % enthalten ist.

Lösung.

(a) Sei X_i die Augenzahl im i -ten Wurf. Dann sind die X_i u.i.v. Zufallsgrößen mit $\mathbb{P}_{X_i} = \mathcal{U}_{\{1,2,3,4,5,6\}}$, $\mu := \mathbb{E}X_i = \frac{6+1}{2} = \frac{7}{2}$ und $\sigma^2 := \text{Var } X_i = \frac{6^2-1}{12} = \frac{35}{12}$. Setze $S_n := \sum_{i=1}^{50} X_i$ und $n := 50$. Dann ist die Normalapproximation anwendbar, da $n \geq 30$, und wir erhalten ohne Stetigkeitskorrektur

$$\begin{aligned} \mathbb{P}(160 \leq S_n \leq 190) &\approx \Phi\left(\frac{190 - n\mu}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{160 - n\mu}{\sqrt{n\sigma^2}}\right) \\ &= \Phi(+1.24) - \Phi(-1.24) \stackrel{\text{Tabelle}}{\approx} 0.8925 - (1 - 0.8925) = 0.7850 \end{aligned}$$

und mit Stetigkeitskorrektur

$$\begin{aligned}\mathbb{P}(160 \leq S_n \leq 190) &\approx \Phi\left(\frac{190 - n\mu + \frac{1}{2}}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{160 - n\mu - \frac{1}{2}}{\sqrt{n\sigma^2}}\right) \\ &= \Phi(+1.28) - \Phi(-1.28) \stackrel{\text{Tabelle}}{\approx} 0.8997 - (1 - 0.8997) = 0.7994.\end{aligned}$$

Die gesuchte W. beträgt also etwa 78,50 % bzw. 79,94 %. Der exakte W. (\rightsquigarrow Computer) beträgt übrigens (gerundet) 80,02 %, so dass die Approximation mit Stetigkeitskorrektur deutlich genauer ist.

(b) Wir verwenden die Bezeichnungen aus Teil (a). Da S_n^* näherungsweise normalverteilt ist, ist es näherungsweise optimal, das Intervall symmetrisch um $\mathbb{E}(S_n) = 175$ anzuordnen. Gesucht ist also ein möglichst kleines $t > 0$ mit

$$\mathbb{P}(175 - t \leq S_n \leq 175 + t) \stackrel{!}{\geq} 0.9.$$

Normalapproximation (anwendbar, da $n \geq 30$) liefert ¹⁴

$$\Phi\left(\frac{t}{\sqrt{n\sigma^2}}\right) - \Phi\left(-\frac{t}{\sqrt{n\sigma^2}}\right) \stackrel{!}{\gtrsim} 0.9.$$

bzw. (wegen $\Phi(-x) = 1 - \Phi(x)$)

$$\Phi\left(\frac{t}{\sqrt{n\sigma^2}}\right) \stackrel{!}{\gtrsim} 0.95.$$

Mittels Tabelle ergibt sich daher

$$\frac{t}{\sqrt{n\sigma^2}} \gtrsim 1.65 \quad \text{bzw.} \quad t \gtrsim 1.65 \sqrt{n\sigma^2} \approx 19.93.$$

Um die erforderliche Mindestgenauigkeit von 90 % einzuhalten, runden wir sicherheits halber auf und wählen $t := 20$ bzw. $I := [175 - 20, 175 + 20] = [155, 195]$. \square

Zum Abschluss wollen wir noch auf einen Spezialfall, den Satz von de Moivre–Laplace, eingehen, der sowohl historisch als auch didaktisch von Interesse ist (u. a. weil dieser Satz häufig in der Schule behandelt wird):

Satz 5.10 (Satz von de Moivre–Laplace). *Es seien X_1, X_2, X_3, \dots unabhängige Zufallsgrößen, die Bernoulli-verteilt zum Parameter $p \in (0, 1)$ sind. Dann gilt für die Summen $S_n := \sum_{i=1}^n X_i$*

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z,$$

wobei Z eine standardnormalverteilte Zufallsgröße bezeichnet.

¹⁴ Hier haben wir die Stetigkeitskorrektur aus Gründen der Einfachheit weggelassen. Mit Stetigkeitskorrektur kämen wir auf $t \approx 19.43$, was aber hier nach Aufrundung zum selben Ergebnis führen würde.

Normalapproximation der Binomialverteilung.

Ähnlich wie oben verwenden wir Satz 5.10 in Anwendungen, um für eine Zufallsgröße $S_n \sim \mathcal{B}(n, p)$ und $a, b \in \mathbb{Z} \cup \{\pm\infty\}$ mit $-\infty \leq a \leq b \leq +\infty$ die Normalapproximation

$$\mathbb{P}(a \leq S_n \leq b) \approx \int_{(a-np)/\sqrt{npq}}^{(b-np)/\sqrt{npq}} \varphi(x) dx = \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right) \quad (*)$$

bzw. – besser noch – die Normalapproximation *mit Stetigkeitskorrektur*

$$\mathbb{P}(a \leq S_n \leq b) \approx \int_{(a-np-\frac{1}{2})/\sqrt{npq}}^{(b-np+\frac{1}{2})/\sqrt{npq}} \varphi(x) dx = \Phi\left(\frac{b-np+\frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np-\frac{1}{2}}{\sqrt{npq}}\right) \quad (**)$$

zu erhalten, wobei (wie auch im Folgenden) $q := 1 - p$ gesetzt sei.

Auch hier stellt sich die Frage, wie groß n mindestens sein muss, um eine brauchbare Näherung für die gesuchte Wahrscheinlichkeit zu erhalten. Hierzu wird in der Literatur die folgende Regel vorgeschlagen (die mit der vorhergehenden Regel nicht konsistent ist, die aber speziell für die Binomialverteilung vorzuziehen ist):

Faustregel:

Die Normalapproximation (*) bzw. (**) für die Binomialverteilung ist brauchbar, wenn $npq \geq 9$ gilt.

Wenn p oder q sehr nahe bei Null liegt, ist es u.U. sinnvoll, statt auf die Normalapproximation auf die Poisson-Approximation der Binomialverteilung (vgl. Satz 1.27) zurückzugreifen.

Zusammenfassung:

- Die Grenzwertsätze der Stochastik beschreiben (in ihrer einfachsten Fassung) das Verhalten von Summen unabhängiger, identisch verteilter Zufallsgrößen
- Die arithmetischen Mittel von u.i.v. Zufallsgrößen in \mathcal{L}^1 stabilisieren sich um den Erwartungswert – in dem Sinne, dass die W. für eine Abweichung um mindestens ε für jedes $\varepsilon > 0$ gegen Null konvergiert.
- Die Verteilung der Summe von u.i.v. Zufallsgrößen in \mathcal{L}^2 wird für $n \rightarrow \infty$ durch die Normalverteilung beschrieben; dabei sind die Fluktuationen um den Erwartungswert bei einer n -fachen Summe von der Größenordnung \sqrt{n} .

6 Statistik

Die *Statistik* befasst sich allgemein mit der Erhebung, Aufbereitung und Auswertung von (durch Beobachtungen oder durch Experimente gewonnenen) Daten. Man unterscheidet:

- *beschreibende (deskriptive) Statistik*: vor allem Aufbereitung von Daten, etwa durch Berechnung von Kennzahlen oder Erstellung von Graphiken
- *schließende (induktive) Statistik*: vor allem Auswertung von Daten; Rückschlüsse von den Daten auf die zugrunde liegenden „Gesetzmäßigkeiten“

In diesem Kapitel soll eine kurze Einführung in die schließende Statistik gegeben werden. Die schließende Statistik wird auch als *mathematische Statistik* bezeichnet, da mathematische Methoden dabei eine wesentliche Rolle spielen.

6.1 Einführende Beispiele

Zur Motivation betrachten wir zwei einführende Beispiele:

Beispiel 6.1 (Verbogene Münze).

Wir betrachten das folgende Glücksspiel: Eine verbogene Münze wird geworfen.

- „Kopf“ \rightarrow Spieler gewinnt 2€
- „Zahl“ \rightarrow Spieler verliert 1€

Ist die Teilnahme an diesem Glücksspiel zu empfehlen?

Um die Frage zu beantworten, versuchen wir, uns einen Eindruck von der W. von „Kopf“ zu verschaffen. Es gelingt uns, die Münze „auszuleihen“ und 100 mal zu werfen; dabei beobachten wir eine Folge $ZKZZKZKKZZZKZZK \dots$, die (genau) 35 mal „Kopf“ enthält. Was können wir daraus lernen? \square

Beispiel 6.2 (Schraubenlängen).

In einem Betrieb werden Schrauben mit einer Soll-Länge von 5 cm produziert. Es soll überprüft werden, ob die Maschinen richtig eingestellt sind.

Zu diesem Zweck werden aus dem laufenden Produktionsprozess 100 Schrauben entnommen und vermessen; man erhält 100 Messwerte mit einem arithmetischen Mittel von 4,86 cm. Was kann man daraus lernen? \square

Es liegen also Daten x_1, \dots, x_n vor, die wir bei unserer Entscheidung berücksichtigen sollten. Wir wollen stets davon ausgehen, dass diese Daten durch die Realisierungen von Zufallsgrößen X_1, \dots, X_n gegeben sind, wobei wir die Verteilung von (X_1, \dots, X_n) nicht oder zumindest nicht vollständig kennen. Das Ziel der Statistik besteht dann darin, eine Aussage über diese Verteilung zu treffen.

Wir wollen zunächst eine formale Beschreibung der Situation angeben:

Beispiel (Verbogene Münze – Fortsetzung von Beispiel 6.1).

Wir beobachten die Folge der 100 Münzwurfergebnisse. Diese Folge kann als Element $x = (x_1, \dots, x_n)$ der Menge $\mathcal{X} := \{0, 1\}^n$ angesehen werden, wobei x_i das Ergebnis im i -ten Wurf angibt, $1 \triangleq \text{Kopf}$, $0 \triangleq \text{Zahl}$ und $n := 100$. Bei der Wahl des W.maßes stoßen wir auf das Problem, dass dieses nicht vollständig bekannt ist – ist allerdings $p \in [0, 1]$ die (unbekannte) W. von „Kopf“ beim 1-fachen Münzwurf, so lässt sich der n -fache Münzwurf durch das W.maß \mathbb{P}_p auf \mathcal{X} zur *Produktdichte*

$$f_p(x) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum_{i=1}^n x_i} \quad (x \in \{0, 1\}^n)$$

beschreiben. Dabei nehmen wir an, dass die einzelnen Beobachtungen X_1, \dots, X_n *unabhängig* und *identisch verteilt* sind, vgl. Bemerkung 3.20. Wir haben es also mit einer ganzen Familie $(\mathbb{P}_p)_{p \in [0,1]}$ von in Frage kommenden W.maßen zu tun!

Bemerkung. Bei der Wahl des Modells sind wir davon ausgegangen, dass die einzelnen Ergebnisse bei den 100 Münzwürfen notiert werden. Wir hätten auch davon ausgehen können, dass die absolute Häufigkeit von „Kopf“ bei den 100 Münzwürfen notiert wird. In diesem Fall hätten wir als Grundmenge $\mathcal{X} := \{0, \dots, n\}$ mit $x \triangleq$ absolute Häufigkeit von „Kopf“ und $n := 100$ und als W.maße die Binomialverteilungen $\mathbb{P}_p := \mathcal{B}(n, p)$ mit den Dichten $f_p(x) = \binom{n}{x} p^x (1-p)^{n-x}$ ($x \in \{0, \dots, n\}$) wählen können. – Wie schon in der Stochastik ist es also auch hier wichtig, dass man sich zunächst klar macht, was genau beobachtet wird. \square

Beispiel (Schraubenlängen Fortsetzung von Beispiel 6.2).

Wir beobachten die Folge der 100 Messwerte. Diese Folge ist ein Element $x = (x_1, \dots, x_n)$ der Menge $\mathcal{X} = \mathbb{R}^n$, wobei x_i die Länge der i -ten Schraube (in cm) angibt und $n := 100$. Wir wollen annehmen, dass die Schraubenlängen – etwa infolge zufälliger Schwankungen im Produktionsprozess – zufällig sind und – stark vereinfachend – als Realisierungen unabhängiger, $\mathcal{N}(\mu, \sigma^2)$ -verteilter Zufallsgrößen angesehen werden können. (Dabei sehen wir über das Problem, dass Schraubenlängen stets positiv sind, während normalverteilte Zufallsgrößen auch negative Werte annehmen, großzügig hinweg.) Als W.maß können wir dann das W.maß $\mathbb{P}_{\mu, \sigma^2}$ auf \mathbb{R}^n mit der *Produktdichte*

$$f_{\mu, \sigma^2}(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2 / (2\sigma^2)} \quad (x \in \mathbb{R}^n)$$

wählen. Dabei sind $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ zwei unbekannte Parameter. Auch hier haben wir es also mit einer ganzen Familie $(\mathbb{P}_{\mu, \sigma^2})_{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+}$ von in Frage kommenden W.maßen zu tun!

Bemerkung. Bei der Wahl des Modells nehmen wir an, dass sowohl der Erwartungswert als auch die Varianz der Beobachtungen unbekannt sind. Treffen wir hingegen die (diskussionswürdige!) Annahme, dass die Varianz $\sigma_0^2 > 0$ bekannt ist, so können wir als W.maß das W.maß \mathbb{P}_μ auf \mathbb{R}^n mit der Produktdichte

$$f_\mu(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i-\mu)^2/(2\sigma_0^2)} \quad (x \in \mathbb{R}^n)$$

wählen. Hier besteht der unbekannte Parameter nur noch aus dem Erwartungswert μ ; die Varianz σ_0^2 ist ja bekannt! \square

6.2 Grundbegriffe der Statistik

In der *Wahrscheinlichkeitstheorie* geht man davon aus, dass das W.maß \mathbb{P} bekannt ist, und berechnet dann Wahrscheinlichkeiten, induzierte Verteilungen, Erwartungswerte usw.

In der *Mathematischen Statistik* geht man hingegen davon aus, dass das W.maß \mathbb{P} unbekannt ist (oder zumindest nicht vollständig bekannt ist), und befasst sich mit dem Problem, wie man anhand von Beobachtungen Rückschlüsse auf das unbekannte W.maß \mathbb{P} ziehen kann.

Definition 6.3 (Statistisches Modell, Standardmodell).

- (a) Ein *statistisches Modell* ist ein Tripel $(\mathcal{X}, \mathcal{B}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ bestehend aus einer Menge $\mathcal{X} \neq \emptyset$, einer σ -Algebra \mathcal{B} über \mathcal{X} sowie einer Familie $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ von W.mäßen auf $(\mathcal{X}, \mathcal{B})$.
- (b) Ein statistisches Modell heißt *Standardmodell*, wenn *entweder*

\mathcal{X} abzählbar, $\mathcal{B} = \mathfrak{P}(\mathcal{X})$, $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ = Familie von diskreten W.mäßen auf \mathcal{X}
mit Zähldichten $(f_\vartheta)_{\vartheta \in \Theta}$

(„diskretes Modell“) *oder*

$\mathcal{X} \subseteq \mathbb{R}^n$, $\mathcal{B} = \mathbb{B}^n|_{\mathcal{X}}$, $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ = Familie von stetigen W.mäßen auf \mathcal{X}
mit Riemann-Dichten $(f_\vartheta)_{\vartheta \in \Theta}$

(„stetiges Modell“) gilt.

Bemerkungen 6.4.

- Anschaulich ist \mathcal{X} die Menge der möglichen Beobachtungen, \mathcal{B} das System der Ereignisse, denen sich Wahrscheinlichkeiten zuordnen lassen, und $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ die Familie der in Frage kommenden W.mäße auf $(\mathcal{X}, \mathcal{B})$.
- \mathcal{X} heißt *Stichprobenraum*.
- Θ heißt *Parameterraum*.

- Wir werden im Folgenden vor allem Standardmodelle betrachten. Zur Spezifikation eines solchen Standardmodells reicht es, den Stichprobenraum \mathcal{X} , den Parameterraum Θ sowie die Familie $(f_\vartheta)_{\vartheta \in \Theta}$ der Dichten anzugeben. (Insbesondere können wir als auf die Angabe der σ -Algebra \mathcal{B} verzichten!)
- Zur Berechnung von Wahrscheinlichkeiten und Erwartungswerten erweist es sich als nützlich, die Zufallsgröße $X : \mathcal{X} \rightarrow \mathcal{X}$ mit $X(x) := x$ einzuführen, die gerade unsere Beobachtung beschreibt.
- Oft liegen n gleichartige, sich nicht gegenseitig beeinflussende Beobachtungen vor. Lässt sich dann eine einzelne Beobachtung durch die Menge $\tilde{\mathcal{X}}$ und die Dichten $(\tilde{f}_\vartheta)_{\vartheta \in \Theta}$ beschreiben, so lässt sich die Gesamtbeobachtung durch den *Produkttraum* $\mathcal{X} := \tilde{\mathcal{X}}^n$ und die *Produktdichten* $(f_\vartheta)_{\vartheta \in \Theta}$ mit

$$f_\vartheta(x) := \prod_{i=1}^n \tilde{f}_\vartheta(x_i) \quad (x = (x_1, \dots, x_n) \in \mathcal{X})$$

beschreiben.¹⁵ Die Zahl n wird auch als *Stichprobenumfang* bezeichnet.

Die Komponenten X_1, \dots, X_n der Zufallsgröße X sind dann unter dem W.maß \mathbb{P}_ϑ zur Dichte f_ϑ unabhängige und identisch verteilte Zufallsgrößen mit der Verteilung $\tilde{\mathbb{P}}_\vartheta$ zur Dichte \tilde{f}_ϑ , vgl. dazu Bemerkung 3.20.

- Die Bezeichnung $(\mathcal{X}, \mathcal{B})$ statt (Ω, \mathcal{A}) ist Konvention – von der Idee her arbeiten wir bereits auf der „Bildseite“ der Zufallsgröße, die unsere Beobachtung beschreibt.

Wir wollen noch die folgende Vereinbarung treffen, die von nun an stillschweigend vorausgesetzt wird:

Generalvoraussetzung 6.5. Es gilt $|\Theta| > 1$ sowie $\vartheta_1 \neq \vartheta_2 \Rightarrow \mathbb{P}_{\vartheta_1} \neq \mathbb{P}_{\vartheta_2}$.

Die Annahme $|\Theta| > 1$ schließt den Fall aus, dass es nur ein in Frage kommendes W.maß gibt; hier hätten wir (als Statistiker) nichts zu tun. Die Annahme $\vartheta_1 \neq \vartheta_2 \Rightarrow \mathbb{P}_{\vartheta_1} \neq \mathbb{P}_{\vartheta_2}$ schließt den Fall aus, dass zwei Parameter zum selben W.maß führen; hier hätten wir keine Chance, anhand der Beobachtungen zwischen ϑ_1 und ϑ_2 zu unterscheiden.

Beispiel (Verbogene Münze – Fortsetzung von Beispiel 6.1). In Beispiel 6.1 können wir als statistisches Modell – je nachdem, was wir annehmen, was genau beobachtet wird – entweder das Standardmodell

$$\mathcal{X} = \{0, 1\}^n, \quad f_p(x) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \quad (x \in \{0, 1\}^n), \quad \vartheta \triangleq p, \quad \Theta = [0, 1] \quad (6.1)$$

oder das Standardmodell

$$\mathcal{X} = \{0, \dots, n\}, \quad f_p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (x \in \mathcal{X}), \quad \vartheta \triangleq p, \quad \Theta = [0, 1] \quad (6.2)$$

wählen. Dabei ist $n \in \mathbb{N}$ bekannt und $p \in [0, 1]$ unbekannt. \square

¹⁵ Im Folgenden kann sich der Begriff „Beobachtung“ sowohl auf die einzelnen Beobachtungen als auch auf die „Gesamtbeobachtung“ beziehen.

Beispiel (Schraubenlängen – Fortsetzung von Beispiel 6.2). In Beispiel 6.2 können wir als statistisches Modell – je nachdem, ob wir die Varianz als unbekannt oder bekannt ansehen – das Standardmodell

$$\mathcal{X} = \mathbb{R}^n, \quad f_{\mu, \sigma^2}(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2 / (2\sigma^2)} \quad (x \in \mathbb{R}^n),$$

$$\vartheta \triangleq (\mu, \sigma^2), \quad \Theta = \mathbb{R} \times \mathbb{R}_+ \quad (6.3)$$

oder das Standardmodell

$$\mathcal{X} = \mathbb{R}^n, \quad f_{\mu}(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i - \mu)^2 / (2\sigma_0^2)} \quad (x \in \mathbb{R}^n), \quad \vartheta \triangleq \mu, \quad \Theta = \mathbb{R} \quad (6.4)$$

wählen. Hier sind im ersten Fall sowohl $\mu \in \mathbb{R}$ als auch $\sigma^2 > 0$ unbekannt, wohingegen im zweiten Fall $\mu \in \mathbb{R}$ unbekannt und $\sigma_0^2 > 0$ bekannt ist. \square

Beachte:

Das statistische Modell hängt nicht nur von der Wahl der W.maße ab, sondern auch von den Annahmen darüber, welche Parameter bekannt bzw. unbekannt sind.

Man kann jetzt verschiedene Zielsetzungen verfolgen. Traditionell werden drei Arten von statistischen Problemstellungen unterschieden, die wir anhand von Beispiel 6.1 erläutern wollen:

1. *Punktschätzprobleme:* Versuche, ausgehend von den Beobachtungen einen Wert für den unbekannten Parameter p (oder auch für den zugehörigen erwarteten Gewinn pro Runde $g(p) = 3p - 1$) anzugeben.¹⁶

Hier erscheint es nahe liegend, als Schätzung für den unbekannten Parameter p (d. h. für die W. von „Kopf“) die relative Häufigkeit von „Kopf“ zu verwenden:

$$\gamma : \mathcal{X} \longrightarrow \mathbb{R} \quad \gamma(x) := \frac{1}{n} \sum_{i=1}^n x_i.$$

Eine solche Abbildung, die jeder möglichen Beobachtung eine Schätzung zuordnet, wird auch als *Schätzer* bezeichnet. Es bleibt zu klären, ob der angegebene Schätzer „vernünftig“ ist.

Im vorliegenden Beispiel erhalten wir (für die oben beschriebene Beobachtung x^*) als Schätzung für p den Wert $\gamma(x^*) = \frac{35}{100}$. Da der zugehörige erwartete Gewinn pro Runde $3 \cdot \frac{35}{100} - 1 = \frac{5}{100}$ positiv ist, könnte man daher über eine Teilnahme am Spiel nachdenken. Man könnte aber auch einwenden, dass die Beweislage „zu unsicher“ ist (da sie von den zufallsabhängigen Beobachtungen abhängt und somit zufallsabhängigen Schwankungen unterliegt).

¹⁶ Beachte hierzu, dass man unter \mathbb{P}_p in jeder Runde m.W. $p \cdot 2 \text{ €}$ gewinnt und m.W. $(1-p) \cdot 1 \text{ €}$ verliert, was zu einem erwarteten Gewinn von $g(p) = 2p + (-1)(1-p) = 3p - 1 \text{ €}$ führt.

2. *Bereichsschätzprobleme:* Versuche, ausgehend von den Beobachtungen eine Menge (z.B. ein Intervall) anzugeben, welche den unbekannten Parameter p (oder auch den zugehörigen erwarteten Gewinn pro Runde $g(p) = 3p - 1$) „mit großer Sicherheit“ enthält.

Ein möglicher Ansatz besteht darin, als Menge für den unbekannten Parameter p ein symmetrisches Intervall um die Schätzung für p zu wählen:

$$C : \mathcal{X} \longrightarrow \mathfrak{P}(\mathbb{R}) \quad C(x) = \left[\frac{1}{n} \sum_{i=1}^n x_i - h(x), \frac{1}{n} \sum_{i=1}^n x_i + h(x) \right]$$

(Dabei ist $h(x)$ noch zu bestimmen.) Eine solche Abbildung, die jeder möglichen Beobachtung eine Menge zuordnet, wird auch als *Konfidenzbereich* (bzw. im Fall, dass stets ein Intervall gewählt wird, auch als *Konfidenzintervall*) bezeichnet. Eine solche „Bereichsschätzung“ hat im Vergleich zu einer „Punktschätzung“ den Vorteil, dass sie auch einen Eindruck von der Genauigkeit der Schätzung vermittelt. Auch hier stellt sich die Frage, ob der angegebene Konfidenzbereich „vernünftig“ ist.

3. *Testprobleme:* Versuche ausgehend von den Beobachtungen zu entscheiden, ob der unbekannte Parameter p eine bestimmte Eigenschaft besitzt oder nicht. Hier ist es z.B. von Interesse zu wissen, ob $H : p \leq \frac{1}{3}$ oder $K : p > \frac{1}{3}$ gilt, weil dann der erwartete Gewinn pro Runde $g(p) = 3p - 1$ nicht-positiv bzw. positiv ist. H und K werden auch als *Hypothese* und *Alternative* (*Gegenhypothese*) bezeichnet.

Codiert man die Entscheidung für H durch 0 und die Entscheidung für K durch 1, so erscheint es nahe liegend, eine Entscheidungsregel vom folgenden Typ zu verwenden:

$$\varphi : \mathcal{X} \longrightarrow \{0, 1\} \quad \varphi(x_1, \dots, x_n) := \begin{cases} 0 & \text{falls } \sum_{i=1}^n x_i \leq t, \\ 1 & \text{falls } \sum_{i=1}^n x_i > t. \end{cases}$$

(Dabei ist der Schwellenwert t noch zu bestimmen.) Eine solche Abbildung, die jeder möglichen Beobachtung eine Entscheidung für H bzw. K zuordnet, wird auch als *Test* bezeichnet. Indem man den Schwellenwert t hinreichend groß wählt, kann man sicherstellen, dass man sich nur dann für K (und damit für die Teilnahme am Spiel) entscheidet, wenn die Beweislage „recht eindeutig“ ist. Auch hier bleibt zu klären, ob der angegebene Test „vernünftig“ ist.

Es geht also in allen Fällen darum, ausgehend von der Beobachtung $x \in \mathcal{X}$ eine Aussage über den zugrunde liegenden (aber unbekannten) Parameter p zu treffen.

Es sei angemerkt, dass diese Problemstellungen in ähnlicher Form auch in vielen anderen „Verkleidungen“ auftreten, *zum Beispiel*:

- *Geburtenstatistik*. In der Stadt Rostock wurden im Jahr 2015 997 Jungen und 1052 Mädchen geboren.¹⁷ Lässt dies den Schluss zu, dass die W. für ein Mädchen größer ist als die W. für einen Jungen?
- *Wahlumfragen*. Bei einer Wahlumfrage haben sich von 1000 Befragten 98 Befragte für die Partei X ausgesprochen. Mit welchem Stimmenanteil kann die Partei X in der Gesamtbevölkerung rechnen?
- *Qualitätskontrolle*. Aus einer Lieferung von 10000 Bauteilen werden 500 Bauteile entnommen und daraufhin überprüft, ob sie den technischen Vorgaben entsprechen. Darunter befinden sich 66 Bauteile, die nicht den Vorgaben entsprechen. Lässt dies den Schluss zu, dass mehr als 10% der Bauteile nicht den Vorgaben entsprechen?

6.3 Punktschätztheorie

Hier sind ein statistisches Modell $(\mathcal{X}, \mathcal{B}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ sowie eine zu schätzende Funktion $g : \Theta \rightarrow \mathbb{R}^k$ gegeben, und man möchte eine „möglichst gute“ Prognose (*Schätzung*) für den Wert $g(\vartheta)$ angeben.

Definition 6.6. Ist $(\mathcal{X}, \mathcal{B}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $g : \Theta \rightarrow \mathbb{R}^k$ eine zu schätzende Funktion, so ist ein *Schätzer* (oder auch *Punktschätzer*) für g eine (messbare) Abbildung $\gamma : \mathcal{X} \rightarrow \mathbb{R}^k$.

Interpretation von Schätzern:

Ist $x \in \mathcal{X}$ unsere Beobachtung, so ist $\gamma(x)$ unsere Schätzung für $g(\vartheta)$.

Beispiel: In Beispiel 6.1 ist es z. B. von Interesse, den Parameter p selbst zu schätzen; es ist dann $\gamma : \Theta \rightarrow \mathbb{R}$, $g(p) := p$ die zu schätzende Funktion und $\gamma : \mathcal{X} \rightarrow \mathbb{R}$, $\gamma(x) := \frac{1}{n} \sum_{i=1}^n x_i$ ein möglicher Schätzer für p . Andere Schätzer für p sind z. B.

- $\gamma_1 : \mathcal{X} \rightarrow \mathbb{R}$, $\gamma_1(x) := \frac{1}{m} \sum_{i=1}^m x_i$ mit $1 \leq m \leq n$ fest,
- $\gamma_2 : \mathcal{X} \rightarrow \mathbb{R}$, $\gamma_2(x) := \frac{1}{n} \sum_{i=1}^n \alpha_i x_i$ mit $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ fest,
- $\gamma_3 : \mathcal{X} \rightarrow \mathbb{R}$, $\gamma_3(x) := \frac{1}{2}$.

Offensichtlich stellt sich die Frage, welche von diesen Schätzern vernünftig sind.

Bemerkung: Häufig ist g die Identität oder eine Projektion auf Θ ; in diesem Fall sei natürlich $\Theta \subseteq \mathbb{R}^k$ vorausgesetzt.

Bemerkung: Schätzer für g werden häufig mit \hat{g} bezeichnet.

Da es i. d. R. sehr viele verschiedene Schätzer gibt, stellt sich die Frage, welchen Schätzer wir verwenden sollen. Es gibt verschiedene Ansätze, Schätzer zu finden und zu beurteilen.

¹⁷ Quelle: Statistisches Jahrbuch Mecklenburg-Vorpommern 2017.

Maximum-Likelihood-Methode

Anders als in Beispiel 6.1, wo mit der relativen Häufigkeit ein nahe liegender Schätzer für die unbekannte Wahrscheinlichkeit p zur Verfügung steht, ist es im Allgemeinen unklar, wie ein vernünftiger Schätzer aussehen könnte. Hier sind Methoden nützlich, vernünftige Schätzer (bzw. genauer Kandidaten für vernünftige Schätzer) zu bestimmen. Eine solche Methode ist die *Maximum-Likelihood-Methode*.

Bei der Maximum-Likelihood-Methode entscheidet man sich (von der Idee her) für denjenigen Parameter $\vartheta \in \Theta$, unter dem die Beobachtung x die größte W. $\mathbb{P}_{\vartheta}(\{x\})$ besitzt. In stetigen Modellen haben Einpunktmengen stets W. 0; hier ersetzt man die W. $\mathbb{P}_{\vartheta}(\{x\})$ durch die Riemann-Dichte $f_{\vartheta}(x)$ an der Stelle x :

Definition 6.7 (Maximum-Likelihood-Schätzer). Sei $(\mathcal{X}, \mathcal{B}, (\mathbb{P}_{\vartheta})_{\vartheta \in \Theta})$ ein Standardmodell mit den Dichten $(f_{\vartheta})_{\vartheta \in \Theta}$, wobei $\Theta \subseteq \mathbb{R}^k$. Dann heißt die Funktion

$$L_x : \Theta \rightarrow \mathbb{R}, \quad L_x(\vartheta) = f_{\vartheta}(x)$$

Likelihood-Funktion (zur Beobachtung $x \in \mathcal{X}$), und jeder Schätzer $\hat{\vartheta}_{ML} : \mathcal{X} \rightarrow \Theta$ mit

$$L_x(\hat{\vartheta}_{ML}(x)) = \max_{\vartheta \in \Theta} L_x(\vartheta) \quad \text{für alle } x \in \mathcal{X}$$

heißt ein *ML-Schätzer* für ϑ . Ist $\hat{\vartheta}_{ML}$ eindeutig, so nennen wir $\hat{\vartheta}_{ML}$ den *ML-Schätzer* für ϑ .

Bemerkung: Häufig schreibt man die Bedingung auch in der Form

$$\hat{\vartheta}_{ML}(x) = \operatorname{argmax}_{\vartheta \in \Theta} L_x(\vartheta) \quad \text{für alle } x \in \mathcal{X}.$$

Bemerkung: Bei der Anwendung der ML-Methode können einige Probleme auftreten: argmax existiert nicht, argmax ist nicht eindeutig, argmax ist schwer zu bestimmen, ...

Bemerkung: Soll nicht ϑ , sondern $g(\vartheta)$ geschätzt werden, so verwenden wir den Schätzer $\hat{g}_{ML} := g(\hat{\vartheta}_{ML})$ (*ML-Schätzer für $g(\vartheta)$*).

Bemerkung: Häufig sind Θ ein Intervall und $L_x(\vartheta)$ stetig differenzierbar bzgl. ϑ ; man kann dann versuchen, $\hat{\vartheta}_{ML}$ mittels Differenzialrechnung zu bestimmen.

Bemerkung: In der Praxis bestimmt man in der Regel nicht $\operatorname{argmax} L_x(\vartheta)$, sondern $\operatorname{argmax} \ell_x(\vartheta)$, wobei $\ell_x(\vartheta) := \log L_x(\vartheta)$ die sog. *Log-Likelihood-Funktion* bezeichnet. Da \log streng monoton wachsend ist, führt dies natürlich zum selben Ergebnis. (Dabei ist \log die natürliche Logarithmusfunktion, die Umkehrfunktion zur natürlichen Exponentialfunktion \exp .)

Beispiele 6.8.

(a) (*Münzwurf*)

Wir kehren zurück zu Beispiel 6.1. Das statistische Modell ist durch

$$\mathcal{X} := \{0, 1\}^n, \quad f_p(x) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i}, \quad \vartheta \triangleq p, \quad \Theta = [0, 1]$$

gegeben. Damit erhalten wir als Likelihood-Funktion

$$L_x(p) = f_p(x) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \rightarrow \max!$$

und als Log-Likelihood-Funktion

$$\begin{aligned} \ell_x(p) &= \log L_x(p) = \sum_{i=1}^n \left(x_i \log p + (1-x_i) \log(1-p) \right) \\ &= \left(\sum_{i=1}^n x_i \right) \log p + \left(n - \sum_{i=1}^n x_i \right) \log(1-p) \rightarrow \max! \end{aligned}$$

Differentiation bzgl. p (!!!) liefert

$$\ell'_x(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} \stackrel{!}{=} 0 \Leftrightarrow \dots \Leftrightarrow p = \frac{1}{n} \sum_{i=1}^n x_i;$$

wir erhalten also

$$\hat{p}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n x_i.$$

(Streng genommen ist noch zu überprüfen, dass an der Nullstelle von $\ell'_x(p)$ wirklich ein Maximum von $\ell_x(p)$ vorliegt und dass das Resultat auch für die beiden Randfälle $x = (0, \dots, 0)$ und $x = (1, \dots, 1)$ gültig bleibt, aber dieser Schritt wird in der Praxis meist übersprungen.) Der ML-Schätzer für die Wahrscheinlichkeit von Kopf ist also durch die relative Häufigkeit von Kopf gegeben!

(b) (*normalverteilte Beobachtungen; bekannte Varianz*)

Gegeben seien u.i.v. Beobachtungen $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma_0^2)$, wobei μ unbekannt und σ_0^2 bekannt ist. Das statistische Modell ist durch

$$\mathcal{X} = \mathbb{R}^n, \quad f_\mu(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i - \mu)^2 / (2\sigma_0^2)}, \quad \vartheta \triangleq \mu, \quad \Theta = \mathbb{R}$$

gegeben. Damit erhalten wir als Likelihood-Funktion

$$L_x(\mu) = f_\mu(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(x_i - \mu)^2 / (2\sigma_0^2)} \rightarrow \max!$$

und als Log-Likelihood-Funktion

$$\begin{aligned}\ell_x(\mu) = \log L_x(\mu) &= \sum_{i=1}^n \left(-\log \sqrt{2\pi\sigma_0^2} - (x_i - \mu)^2 / (2\sigma_0^2) \right) \\ &= -\frac{1}{2}n \log(2\pi\sigma_0^2) - \frac{1}{2\sigma_0^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma_0^2} \sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma_0^2} \rightarrow \max!\end{aligned}$$

Differentiation bzgl. μ (!!!) liefert

$$\ell'_x(\mu) = \frac{1}{\sigma_0^2} \sum_{i=1}^n x_i - \frac{n\mu}{\sigma_0^2} \stackrel{!}{=} 0 \Leftrightarrow \dots \Leftrightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i;$$

wir erhalten also

$$\hat{\mu}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n x_i.$$

Der ML-Schätzer für den Erwartungswert ist also durch das *Stichprobenmittel* gegeben!

(c) (*normalverteilte Beobachtungen; unbekannte Varianz*)

Gegeben seien u.i.v. Beobachtungen $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, wobei μ und σ^2 beide unbekannt sind. Das statistische Modell ist durch

$$\mathcal{X} = \mathbb{R}^n, \quad f_{\mu, \sigma^2}(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2 / (2\sigma^2)}, \quad \vartheta \triangleq (\mu, \sigma^2), \quad \Theta = \mathbb{R} \times \mathbb{R}_+$$

gegeben. Damit erhalten wir als Likelihood-Funktion

$$L_x(\mu, \sigma^2) = f_{\mu, \sigma^2}(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i - \mu)^2 / (2\sigma^2)} \rightarrow \max!$$

und als Log-Likelihood-Funktion

$$\begin{aligned}\ell_x(\mu, \sigma^2) = \log L_x(\mu, \sigma^2) &= \sum_{i=1}^n \left(-\log \sqrt{2\pi\sigma^2} - (x_i - \mu)^2 / (2\sigma^2) \right) \\ &= -\frac{1}{2}n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2} \rightarrow \max!\end{aligned}$$

Im Folgenden schreiben wir v statt σ^2 , weil sich dann einfacher differenzieren lässt. Hier liegt ein zweidimensionales Optimierungsproblem vor, dass man im Allgemeinen dadurch lösen könnte, dass man die partiellen Ableitungen $\frac{\partial}{\partial \mu}$ und $\frac{\partial}{\partial v}$ berechnet und (gleichzeitig!) gleich Null setzt. Im vorliegenden Fall geht es aber auch einfacher:

Zuerst optimieren wir bei festem v bzgl. μ und erhalten wie in Teil (b)

$$\frac{\partial}{\partial \mu} \ell_x(\mu, v) = \frac{1}{v} \sum_{i=1}^n x_i - \frac{n\mu}{v} \stackrel{!}{=} 0 \Leftrightarrow \dots \Leftrightarrow \mu = \frac{1}{n} \sum_{i=1}^n x_i =: \bar{x}_n.$$

Anschließend setzen wir für μ den erhaltenen Wert \bar{x}_n ein und optimieren den verbleibenden Ausdruck

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

bzgl. v und erhalten

$$\frac{\partial}{\partial v} \ell_x(\bar{x}_n, v) = -\frac{n}{2v} + \frac{1}{2v^2} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \stackrel{!}{=} 0 \Leftrightarrow \dots \Leftrightarrow v = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Damit erhalten wir insgesamt

$$\hat{\mu}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad \hat{v}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2.$$

Beurteilung von Schätzern

Es gibt verschiedene Kriterien zur Beurteilung von Schätzern, die nicht unbedingt den gleichen Schätzer empfehlen.

Ein Schätzer heißt *erwartungstreu*, wenn er *im Mittel* zum richtigen Ergebnis führt:

Definition 6.9 (Erwartungstreuer Schätzer). Ein Schätzer $\gamma : \mathcal{X} \rightarrow \mathbb{R}^1$ heißt *erwartungstreu* (oder auch *unverzerrt*) für g , wenn für jedes $\vartheta \in \Theta$ $\mathbb{E}_{\vartheta}(\gamma(X)) = g(\vartheta)$ gilt.

Dabei bezeichnet \mathbb{E}_{ϑ} den Erwartungswert unter dem W.maß \mathbb{P}_{ϑ} .

Ein erwartungstreuer Schätzer heißt *gleichmäßig varianzminimierend*, wenn er unter allen erwartungstreuen Schätzern *die kleinste Streuung* besitzt:

Definition 6.10 (Gleichmäßig varianzminimierender erwartungstreuer Schätzer).

Ein erwartungstreuer Schätzer γ^* für g heißt *gleichmäßig varianzminimierend*, falls für jeden erwartungstreuen Schätzer γ für g

$$\text{Var}_{\vartheta}(\gamma^*(X)) \leq \text{Var}_{\vartheta}(\gamma(X)) \quad \text{für alle } \vartheta \in \Theta$$

gilt.

Ein Schätzer heißt konsistent, wenn der Schätzfehler bei einer großen Anzahl von Beobachtungen klein wird. Streng genommen haben wir es hier mit einer Folge $(\gamma_n)_{n \in \mathbb{N}}$ von Schätzern zu tun. Genauer wollen wir annehmen, dass eine Folge $(X_n)_{n \in \mathbb{N}}$ von u.i.v. Beobachtungen vorliegt und für jeden Stichprobenumfang $n \in \mathbb{N}$ ein Schätzer $\gamma_n = f_n(X_1, \dots, X_n)$ gegeben ist, der nur von den ersten n Beobachtungen abhängt.

Definition 6.11 (Konsistenter Schätzer). Die Folge $(\gamma_n)_{n \in \mathbb{N}}$ heißt *konsistent* für g , falls für jedes $\vartheta \in \Theta$

$$(\forall \varepsilon > 0) \quad \lim_{n \rightarrow \infty} \mathbb{P}_{\vartheta}(|\gamma_n - g(\vartheta)| \geq \varepsilon) = 0$$

gilt (d. h. falls für jedes $\vartheta \in \Theta$ die Folge $(\gamma_n)_{n \in \mathbb{N}}$ \mathbb{P}_{ϑ} -stochastisch gegen $g(\vartheta)$ konvergiert).

Natürlich ist – ähnlich wie bei der ML-Schätzung – im Allgemeinen nicht klar, ob es überhaupt Schätzer mit diesen Eigenschaften gibt.

Beispiele 6.12 (Fortsetzung von Beispiel 6.8).

(a) (*Münzwurf*)

Der ML-Schätzer \hat{p}_{ML} mit $\hat{p}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n x_i$ ist erwartungstreu und konsistent für p , denn für jedes $p \in [0, 1]$ gilt

$$\mathbb{E}_p\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_p(X_i) = \frac{1}{n} \sum_{i=1}^n p = p$$

sowie (nach dem Schwachen Gesetz der großen Zahlen)

$$(\forall \varepsilon > 0) \quad \mathbb{P}_p\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

Der ML-Schätzer ist zudem varianzminimierend (was aber schwieriger zu zeigen ist).

(b) (*normalverteilte Beobachtungen; bekannte Varianz*)

Der ML-Schätzer $\hat{\mu}_{\text{ML}}$ mit $\hat{\mu}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n x_i$ ist erwartungstreu und konsistent für μ , denn für jedes $\mu \in \mathbb{R}$ gilt

$$\mathbb{E}_{\mu}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mu}(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$

sowie (nach dem Schwachen Gesetz der großen Zahlen)

$$(\forall \varepsilon > 0) \quad \mathbb{P}_{\mu}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0$$

Der ML-Schätzer ist zudem varianzminimierend (was aber schwieriger zu zeigen ist).

(c) (*normalverteilte Beobachtungen; unbekannte Varianz*)

Der ML-Schätzer für μ mit $\hat{\mu}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n x_i$ ist erwartungstreu für μ , denn es gilt für jedes $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$

$$\mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mu, \sigma^2}(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$

Der ML-Schätzer für v mit $\hat{v}_{\text{ML}}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ ist hingegen *nicht* erwartungstreu für v , denn es gilt für jedes $c \in \mathbb{R}$

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 &= \frac{1}{n} \sum_{i=1}^n ((x_i - c) - (\bar{x}_n - c))^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - c)^2 + (\bar{x}_n - c)^2 - 2(\bar{x}_n - c) \underbrace{\frac{1}{n} \sum_{i=1}^n (x_i - c)}_{= (\bar{x}_n - c)} \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - c)^2 - (\bar{x}_n - c)^2 \end{aligned} \quad (6.5)$$

und damit für jedes $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$

$$\begin{aligned} \mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mu, \sigma^2}(X_i - \mu)^2 - \mathbb{E}_{\mu, \sigma^2}(\bar{X}_n - \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \sigma^2 - \frac{1}{n} \sigma^2 = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Ohne Beweis sei angemerkt, dass beide ML-Schätzer konsistent sind.

Allerdings kann man \hat{v}_{ML} für $n \geq 2$ leicht in einen erwartungstreuen Schätzer überführen, indem man noch mit $\frac{n}{n-1}$ multipliziert. Betrachtet man die obigen Rechnungen, so sieht man, dass man für den Nachweis der Erwartungstreue nur benötigt, dass X_1, \dots, X_n unter $\mathbb{P}_{\mu, \sigma^2}$ u.i.v. mit Erwartungswert μ und Varianz σ^2 sind; die Normalverteilung selbst spielt keine Rolle. In der Tat gilt der folgende Satz:

Satz 6.13 (Schätzung von Erwartungswert und Varianz). *Gegeben seien eine Familie von W.mäßen $(\mathbb{P}_{\vartheta})_{\vartheta \in \Theta}$ sowie eine Folge von Zufallsgrößen X_1, X_2, X_3, \dots , so dass X_1, X_2, X_3, \dots unter \mathbb{P}_{ϑ} unabhängig und identisch verteilt (u.i.v.) mit Erwartungswert $\mu(\vartheta)$ und Varianz $\sigma^2(\vartheta)$ sind. Dann sind das Stichprobenmittel $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$ ein erwartungstreuer und konsistenter Schätzer für $\mu(\vartheta)$ und für $n \geq 2$ die Stichprobenvarianz $S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ ein erwartungstreuer und konsistenter Schätzer für $\sigma^2(\vartheta)$.*

Bemerkung. Die Werte von \bar{X}_n und S_n^2 für eine gegebene Stichprobe $x \in \mathbb{R}^n$ werden im Folgenden mit \bar{x}_n und s_n^2 bezeichnet.

Warnung. Die Bezeichnung *Stichprobenvarianz* ist in der Literatur nicht einheitlich; einige Autoren bezeichnen damit auch $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

6.4 Bereichsschätztheorie

Hier sind ein statistisches Modell $(\mathcal{X}, \mathcal{B}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ sowie eine zu schätzende Funktion $g : \Theta \rightarrow \mathbb{R}^k$ gegeben, und es ist eine Teilmenge von \mathbb{R}^k gesucht, die den „wahren“ Wert $g(\vartheta)$ „mit großer Sicherheit“ enthält. Eine solche *Bereichsschätzung* vermittelt im Unterschied zu einer *Punktschätzung* auch einen Eindruck von der Genauigkeit der Schätzung.

Definition 6.14. Ist $(\mathcal{X}, \mathcal{B}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $g : \Theta \rightarrow \mathbb{R}^k$ eine zu schätzende Funktion, so ist ein *Konfidenzbereich* (oder *Bereichsschätzer*) für $g(\vartheta)$ eine Abbildung $C : \mathcal{X} \rightarrow \mathfrak{P}(\mathbb{R}^k)$, so dass $\{g(\vartheta) \in C(X)\} = \{x \in \mathcal{X} : g(\vartheta) \in C(x)\} \in \mathcal{B}$ für alle $\vartheta \in \Theta$. Ist dabei $k = 1$ und $C(x)$ für jedes $x \in \mathcal{X}$ ein Intervall, so sprechen wir auch von einem *Konfidenzintervall*.

Interpretation von Konfidenzbereichen:

Ist x unsere Stichprobe, so ist $C(x)$ unsere „Bereichsschätzung“ für $g(\vartheta)$, d. h. wir vermuten, dass $g(\vartheta)$ in $C(x)$ liegt.

Definition 6.15. Sei $\alpha \in (0, 1)$. Ein Konfidenzbereich heißt *Konfidenzbereich zum Niveau $1 - \alpha$* für $g(\vartheta)$, falls $\mathbb{P}_\vartheta(g(\vartheta) \in C(X)) \geq 1 - \alpha$ für alle $\vartheta \in \Theta$.

Jetzt ist auch der Sinn der „technischen Bedingung“ in Def. 6.14 klar: Sie stellt sicher, dass die Menge $\{g(\vartheta) \in C(X)\}$ ein *Ereignis* ist und somit eine W. besitzt.

Interpretation von Konfidenzbereichen zum Niveau $1 - \alpha$:

Ist C ein Konfidenzbereich zum Niveau $1 - \alpha$, so enthält $C(X)$ den Wert $g(\vartheta)$ m. W. $\geq 1 - \alpha$, egal welches \mathbb{P}_ϑ vorliegt.

Ist die Stichprobe $x \in \mathcal{X}$ erst einmal gegeben, so ist die Aussage „ $C(x)$ enthält $g(\vartheta)$ m. W. $\geq 1 - \alpha$ “ aber unsinnig, weil dann kein Zufall mehr im Spiel ist.

Oder anders: Ein Konfidenzbereich zum Niveau 90% ist so konstruiert, dass er im Mittel (bei wiederholter Anwendung auf unterschiedlichen Daten) in mindestens 9 von 10 Fällen eine Menge liefert, die den wahren Wert $g(\vartheta)$ enthält. Zu der Frage, ob eine konkrete Anwendung den wahren Wert $g(\vartheta)$ „erwischt“ hat oder nicht, lässt sich aber (leider!) nichts aussagen.

Natürlich sind wir an einer möglichst genauen Bereichsschätzung interessiert und versuchen dementsprechend, einen „möglichst kleinen“ Konfidenzbereich C zum Niveau $1 - \alpha$ für $g(\vartheta)$ zu finden.

Dazu geht man üblicherweise von einem Schätzer $\hat{g}(X)$ für $g(\vartheta)$ aus und versucht, so eine Menge um $\hat{g}(X)$ anzuordnen, dass ein Konfidenzbereich $C(X)$ zum Niveau $1 - \alpha$ für $g(\vartheta)$ entsteht. Dazu benötigt man üblicherweise genauere Informationen über die Verteilung von $\hat{g}(X)$, wobei natürlich zu beachten ist, dass diese in aller Regel vom zugrunde liegenden Parameter ϑ abhängen wird!

Hier hilft manchmal der folgende Ansatz: Angenommen, man hat schon einen Schätzer $\hat{g}(X)$ für $g(\vartheta)$ und man findet stetige Transformationen $T_{g(\vartheta)} : \mathbb{R}^k \rightarrow \mathbb{R}^k$, so dass die Verteilung Q von $T_{g(\vartheta)}(\hat{g}(X))$ unter \mathbb{P}_ϑ nicht von ϑ abhängt. Wählt man dann (irgend)eine Menge $I \in \mathbb{B}^k$ mit $Q(I) \geq 1 - \alpha$ und $C(x) := \{g(\vartheta) \in \mathbb{R}^k : T_{g(\vartheta)}(\hat{g}(x)) \in I\}$, so folgt für jedes $\vartheta \in \Theta$

$$\mathbb{P}_\vartheta(g(\vartheta) \in C(X)) = \mathbb{P}_\vartheta(T_{g(\vartheta)}(\hat{g}(X)) \in I) = Q(I) \geq 1 - \alpha,$$

d. h. man erhält einen Konfidenzbereich zum Niveau $1 - \alpha$ für $g(\vartheta)$.

Für das Beispiel zu diesem Ansatz benötigen wir noch zwei Definitionen:

Definition 6.16 (Quantil). Ist \mathbb{P} eine W.verteilung auf \mathbb{R} und ist $t \in (0, 1)$, so heißt jede Zahl $x \in \mathbb{R}$ mit

$$\mathbb{P}([-\infty, x]) \geq t \quad \text{und} \quad \mathbb{P}([x, \infty]) \geq 1 - t$$

t -Quantil von \mathbb{P} .

Bemerkungen 6.17.

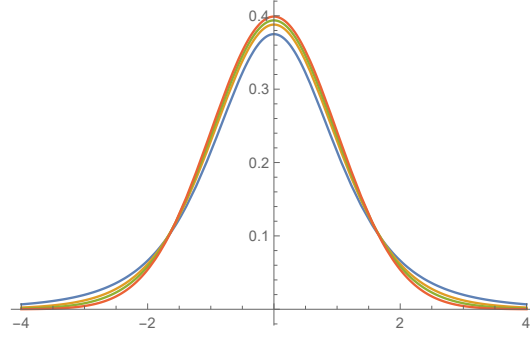
- Besitzt \mathbb{P} eine stetige und streng monoton wachsende Verteilungsfunktion F , so ist das t -Quantil von \mathbb{P} gerade diejenige Stelle $x \in \mathbb{R}$, an der F den Wert t annimmt, d. h. das t -Quantil von \mathbb{P} ist durch $F^{-1}(t)$ gegeben.
- Ist also $\mathbb{P} = \mathcal{N}(0, 1)$, so ist das t -Quantil durch $\Phi^{-1}(t)$ gegeben und kann (zumindest näherungsweise) mit Hilfe der Tabelle für die Verteilungsfunktion der Standardnormalverteilung bestimmt werden.
- Alternativ gibt es sowohl für die Standardnormalverteilung als auch für einige andere häufig auftretende Verteilungen (wie z. B. die nachfolgend eingeführten t -Verteilungen) Quantil-Tabellen, in denen man die Quantile nachschlagen kann. (Außerdem kann man natürlich geeignete Computer-Programme verwenden.)
- Das 0.5-Quantil von \mathbb{P} ist gerade der Median von \mathbb{P} .

Definition 6.18 (t -Verteilung). Für jedes $n \in \mathbb{N}$ ist die t -Verteilung mit n Freiheitsgraden (kurz t_n) die stetige W.verteilung auf (\mathbb{R}, \mathbb{B}) mit der Dichte

$$f_n(x) := \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n} \Gamma(\frac{1}{2}) \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad (x \in \mathbb{R}),$$

wobei $\Gamma(t) := \int_0^\infty x^{t-1} e^{-x} dx$ ($t > 0$) die *Gamma-Funktion* bezeichnet.

Im Folgenden sind die Dichten f_n für $n = 4$ (blau), $n = 9$ (gelb) und $n = 19$ (grün) dargestellt. Für große Werte von n lassen sich die Dichten f_n kaum von der Dichte φ der Standardnormalverteilung (rot) unterscheiden. Man sieht, dass die Dichten f_n ähnliche Symmetrie- und Monotonie-Eigenschaften wie φ besitzen.



Dichten einiger t -Verteilungen (Erläuterung im Text)

Beispiel 6.19.

(a) (*normalverteilte Beobachtungen; σ_0^2 bekannt*)

Gegeben sei die Situation aus Beispiel 6.8 (b), und gesucht sei ein Konfidenzintervall zum Niveau $1 - \alpha$ für μ , wobei $\alpha \in (0, 1)$ vorgegeben ist.

Es ist nahe liegend, das Konfidenzintervall C zum Niveau $1 - \alpha$ für μ ausgehend vom Schätzer \bar{X}_n für μ zu konstruieren. Dabei können wir die Besonderheit ausnutzen, dass bei den vorliegenden normalverteilten Beobachtungen für jedes $\mu \in \mathbb{R}$ unter \mathbb{P}_μ

$$T_\mu(X) := \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma_0} \sim \mathcal{N}_{0,1}$$

gilt:

$$\begin{aligned} X_1, \dots, X_n \text{ u.i.v. } \sim \mathcal{N}(\mu, \sigma_0^2) &\xRightarrow{\text{Lemma 3.25 (b)}} \sum_{i=1}^n X_i \sim \mathcal{N}(n\mu, n\sigma_0^2) \\ &\xRightarrow{\text{Lemma 2.22 (b)}} \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}(\mu, \sigma_0^2/n) \\ &\xRightarrow{\text{Lemma 2.22 (b)}} \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma_0} \sim \mathcal{N}(0, 1) \end{aligned}$$

Wir suchen jetzt ein möglichst kleines Intervall I mit $\mathcal{N}_{0,1}(I) \geq 1 - \alpha$: Dieses ist (wegen der Form der Dichte φ) durch $I = [-z_{1-\frac{\alpha}{2}}, +z_{1-\frac{\alpha}{2}}]$ gegeben, wobei $z_{1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$ -Quantil der $\mathcal{N}_{0,1}$ -Verteilung bezeichnet. Setzen wir nun

$$C(x) := \{\mu \in \mathbb{R} : T_\mu(x) \in I\} = [\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}],$$

so gilt für jedes $\mu \in \mathbb{R}$

$$\mathbb{P}_\mu(\mu \in C(X)) = \mathbb{P}_\mu(T_\mu(X) \in I) = \mathcal{N}_{0,1}(I) \geq 1 - \alpha,$$

d. h. wir haben ein Konfidenzintervall C zum Niveau $1 - \alpha$ für μ gefunden:

$$C(x) = \left[\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right]$$

(b) (*normalverteilte Beobachtungen; σ^2 unbekannt*)

Gegeben sei die Situation aus Beispiel 6.8 (c), und gesucht sei ein Konfidenzintervall zum Niveau $1 - \alpha$ für μ , wobei $\alpha \in (0, 1)$ vorgegeben ist.

Wir versuchen wieder, das Konfidenzintervall C zum Niveau $1 - \alpha$ für μ ausgehend vom Schätzer \bar{X}_n für μ zu konstruieren. Das Vorgehen in Teil (a) führt hier leider nicht zum Ziel, weil das dortige Intervall von der – nun unbekannten – Varianz σ^2 abhängt und damit kein zulässiges Konfidenzintervall darstellt. Wir ersetzen daher die Varianz σ^2 durch die Stichprobenvarianz S_n^2 , vgl. Satz 6.13. Hier können wir die Besonderheit ausnutzen, dass bei den vorliegenden normalverteilten Beobachtungen dass für jedes $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ unter $\mathbb{P}_{\mu, \sigma^2}$

$$T_\mu(X) := \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \sim t_{n-1}$$

gilt.¹⁸ Wir suchen jetzt ein möglichst kleines Intervall I mit $t_{n-1}(I) \geq 1 - \alpha$: Dieses ist (wegen der Form der Dichte f_{n-1}) durch $I = [-t_{n-1, 1-\frac{\alpha}{2}}, +t_{n-1, 1-\frac{\alpha}{2}}]$ gegeben, wobei $t_{n-1, 1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$ -Quantil der t_{n-1} -Verteilung bezeichnet. Setzen wir nun

$$C(x) := \{\mu \in \mathbb{R} : T_\mu(x) \in I\} = \left[\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right],$$

so gilt für jedes $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$

$$\mathbb{P}_{\mu, \sigma^2}(\mu \in C(X)) = \mathbb{P}_{\mu, \sigma^2}(T_\mu(X) \in I) = t_{n-1}(I) \geq 1 - \alpha,$$

d. h. wir haben ein Konfidenzintervall zum Niveau $1 - \alpha$ für μ gefunden:

$$C(x) = \left[\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right]$$

Bemerkung: Dass man bei n Beobachtungen die t -Verteilung mit $n - 1$ Freiheitsgraden (und nicht mit n Freiheitsgraden!!!) verwenden muss, kann man sich durch den Hinweis darauf merken, dass man durch die Schätzung des Mittelwertes bei der Berechnung der Stichprobenvarianz einen Freiheitsgrad verliert.

(c) (*Bernoulli-verteilte Beobachtungen*)

Gegeben sei die Situation aus Beispiel 6.1 bzw. Beispiel 6.8 (a) mit $\Theta = (0, 1)$, und gesucht sei ein Konfidenzintervall zum Niveau $1 - \alpha$ für p , wobei $\alpha \in (0, 1)$ vorgegeben ist.

¹⁸ Diese Aussage ist in der Vorlesung ohne Beweis angegeben worden.

Wir versuchen wieder, das Konfidenzintervall C zum Niveau $1 - \alpha$ für p ausgehend vom Schätzer \bar{X}_n für p zu konstruieren. Anders als in den vorhergehenden Beispielen lässt sich hier leider keine Transformation $T_p(X)$ finden, aus der sich \bar{X}_n zurückgewinnen lässt *und* deren Verteilung nicht vom unbekannten Parameter p abhängt. Ist allerdings n „groß“, so gilt nach dem schwachen Gesetz der großen Zahlen bzw. nach dem zentralen Grenzwertsatz für jedes $p \in (0, 1)$ unter \mathbb{P}_p

$$T_p(X) := \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1 - \bar{X}_n)/n}} \approx \frac{\bar{X}_n - p}{\sqrt{p(1 - p)/n}} \approx \mathcal{N}(0, 1)$$

(wobei $X \approx Q$ bedeuten soll, dass die Verteilung von X näherungsweise $= Q$ ist). Setzen wir wie in Beispiel (a) $I = [-z_{1-\frac{\alpha}{2}}, +z_{1-\frac{\alpha}{2}}]$, wobei $z_{1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$ -Quantil der $\mathcal{N}_{0,1}$ -Verteilung bezeichnet, und

$$C(x) := \{p \in \mathbb{R} : T_p(x) \in I\} = [\bar{x}_n - \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} z_{1-\frac{\alpha}{2}}],$$

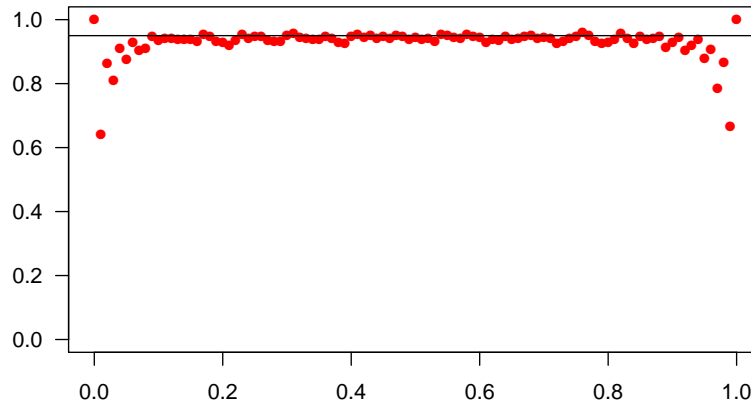
so gilt für jedes $p \in (0, 1)$

$$\mathbb{P}_p(p \in C(X)) = \mathbb{P}_p(T_p(X) \in I) \approx \mathcal{N}_{0,1}(I) \geq 1 - \alpha,$$

d. h. wir haben für „große“ Werte von n ein approximatives Konfidenzintervall C zum Niveau $1 - \alpha$ für p gefunden:

$$C(x) = \left[\bar{x}_n - \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} z_{1-\frac{\alpha}{2}} \right]$$

Beachte, dass die Herleitung dieses Konfidenzintervalls auf zwei Grenzwertsätzen beruht. Es ist daher nicht auszuschließen, dass das Niveau $1 - \alpha$ für „kleine“ Werte von n deutlich unterschritten wird. Eine genauere Analyse zeigt, dass das Niveau $1 - \alpha$ für alle Werte von p geringfügig und für $p \approx 0$ bzw. $p \approx 1$ sogar deutlich unterschritten werden kann:



W. dafür, dass das Konfidenzintervall $C(X)$ für $n = 100$ den wahren Parameter p enthält, in Abhängigkeit von $p \in \{0.00, 0.01, 0.02, \dots, 0.98, 0.99, 1.00\}$

Hinweise zur Konstruktion besserer Konfidenzintervalle finden sich z. B. bei HENZE.

(d) (*allgemeine Beobachtungen*)

Gegeben seien u.i.v. Zufallsgrößen X_1, \dots, X_n wie in Satz 6.13, und gesucht sei ein Konfidenzintervall zum Niveau $1 - \alpha$ für den unbekannten Erwartungswert μ , wobei $\alpha \in (0, 1)$ vorgegeben ist. Dann erhalten wir ähnlich wie in Teil (c) ausgehend von der Transformation $T_\mu(X) := \sqrt{n}(\bar{X}_n - \mu)/S_n$ für „große“ Werte von n das folgende approximative Konfidenzintervall C zum Niveau $1 - \alpha$ für μ :

$$C(x) = \left[\bar{x}_n - \frac{s_n}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right]$$

Auch bei diesem Konfidenzintervall kann das Niveau $1 - \alpha$ für „kleine“ Werte von n wieder unterschritten werden. Speziell für normalverteilte Beobachtungen lässt sich dies anhand von Tabelle 5 bestätigen: Da für „kleine“ Werte von n (bei festem $\alpha \in (0, 1)$) $z_{1-\frac{\alpha}{2}} < t_{n-1, 1-\frac{\alpha}{2}}$ gilt, sind die Konfidenzintervalle in Teil (d) kleiner als die Konfidenzintervalle in Teil (b), und da letztere das Niveau $1 - \alpha$ exakt einhalten, müssen erstere das Niveau $1 - \alpha$ unterschreiten. – Wenn die Annahme gerechtfertigt ist, dass normalverteilte Beobachtungen vorliegen, sollte man daher statt des approximativen Konfidenzintervalls in Teil (d) das exakte Konfidenzintervall in Teil (b) verwenden!

□

Beispiel 6.20. In einem Betrieb sollen Schrauben mit einer Länge von 5 cm hergestellt werden. Es sei bekannt, dass die tatsächlichen Schraubenlängen in Zentimetern (aufgrund von zufälligen Schwankungen im Produktionsprozess) $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind. Aufgrund einer Stichprobe

5.130 5.186 4.983 4.951 5.125 4.910 5.152 4.829 5.011 5.203 4.877 5.093

soll ein Konfidenzintervall zum Niveau 95 % für μ angegeben werden.

1. *Fall:* Die Standardabweichung der Schraubenlängen ist bekannt: 0.1 [cm]

Mit $n = 12$, $\bar{x}_n = 5.0375$, $\sigma_0 = 0.1$, $z_{0.975} \approx 1.96$ ergibt sich

$$C(x) = \left[\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right] \approx [4.98; 5.09]$$

(vgl. die Formel aus Beispiel 6.19 (a)).

2. *Fall:* Die Standardabweichung der Schraubenlängen ist unbekannt.

Mit $n = 12$, $\bar{x}_n = 5.0375$, $s_n = 0.1274$, $t_{11, 0.975} \approx 2.20$ ergibt sich

$$C(x) = \left[\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right] \approx [4.96; 5.12]$$

(vgl. die Formel aus Beispiel 6.19 (b)).

6.5 Testtheorie¹⁹

Bei einem *Testproblem* ist zusätzlich zum statistischen Modell eine Zerlegung $\Theta = H \cup K$ ($H \neq \emptyset$, $K \neq \emptyset$, $H \cap K = \emptyset$) des Parameterraums in zwei Teilmengen gegeben, und es ist zu entscheiden, in welcher dieser Teilmengen der Parameter ϑ liegt. Die Teilmengen H und K heißen auch *Hypothese* (*Nullhypothese*) und *Alternative* (*Gegenhypothese*).

Definition 6.21. Ist $(\mathcal{X}, \mathcal{B}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $\Theta = H \cup K$ eine Zerlegung von Θ , so ist eine *Testfunktion* / ein *Test* für H gegen K eine messbare Abbildung $\varphi : \mathcal{X} \rightarrow \{0, 1\}$.

Interpretation eines Tests:

Ist x unsere Beobachtung, so entscheiden wir uns für H bzw. K , falls $\varphi(x) = 0$ bzw. $\varphi(x) = 1$ gilt.

Alternativ kann man einen Test auch durch eine der beiden Mengen

$$A := \varphi^{-1}(\{0\}) = \{x \in \mathcal{X} : \varphi(x) = 0\} \quad (\text{Annahmehereich / acceptance region})$$

bzw.

$$R := \varphi^{-1}(\{1\}) = \{x \in \mathcal{X} : \varphi(x) = 1\} \quad (\text{Ablehnungsbereich / rejection region})$$

beschreiben, auf denen die Hypothese angenommen bzw. abgelehnt wird.

Problem: Auch hier gibt es i. d. R. wieder viele verschiedene Tests. Welchen Test sollen wir verwenden?

Innerhalb unseres Modells gibt es 2 Arten von Fehlern:

- Fehler 1. Art: wahrer Parameter $\vartheta \in H$, Entscheidung für K
- Fehler 2. Art: wahrer Parameter $\vartheta \in K$, Entscheidung für H

Die zugehörigen W.en $\mathbb{P}_\vartheta(\varphi = 1)$ ($\vartheta \in H$) und $\mathbb{P}_\vartheta(\varphi = 0)$ ($\vartheta \in K$) heißen Fehler-W.en 1. Art bzw. 2. Art. Beachte, dass diese Fehler-W.en für den Fall, dass die Mengen H und K aus mehreren Elementen bestehen, in aller Regel keine Konstanten, sondern Funktionen von $\vartheta \in H$ bzw. $\vartheta \in K$ sind! Für einen „guten“ Test sollten die Fehler-W.en natürlich klein sein.

Definition 6.22. Ist φ ein Test, so heißt die Abbildung

$$\beta : \Theta \rightarrow \mathbb{R} \quad \vartheta \mapsto \mathbb{E}_\vartheta(\varphi(X))$$

Gütefunktion von φ .

Für einen Test φ wie in Definition 6.21 gilt also $\mathbb{E}_\vartheta(\varphi(X)) = \mathbb{P}_\vartheta(X \in R)$ für alle $\vartheta \in \Theta$. Es ist also $\beta(\vartheta) = \text{Fehler-W. 1. Art}$, falls $\vartheta \in H$, und $\beta(\vartheta) = 1 - \text{Fehler-W. 2. Art}$, falls $\vartheta \in K$. Für einen „guten“ Test sollte die Gütefunktion also auf H möglichst klein und auf K möglichst groß sein.

¹⁹ Dieser Abschnitt ist in der Vorlesung aus Zeitgründen weggelassen worden.

In der Regel ist eine Verkleinerung der Fehler-W. 1. Art mit einer Vergrößerung der Fehler-W. 2. Art verbunden (und umgekehrt), so dass es nicht gelingt, beide Fehler-W. gleichzeitig klein zu halten. Wir geben daher für die Fehler-W. 1. Art eine obere Schranke vor und suchen in der Menge aller Tests, die diese obere Schranke einhalten, einen „vernünftigen“ Test:

Definition 6.23. Sei $\alpha \in [0, 1]$ gegeben. Ein Test φ heißt *Test zum Niveau α* , falls $\mathbb{E}_{\vartheta}(\varphi(X)) \leq \alpha$ für alle $\vartheta \in H$ gilt.

Man beachte, dass bei diesem Konzept die „Symmetrie“ zwischen H und K zerstört wird. Verwendet man einen Test zum Niveau α für H gegen K , so tritt ein Fehler 1. Art höchstens mit W. α auf – die Entscheidung zugunsten von K ist (auf dem Niveau α) *statistisch abgesichert*. Dagegen kann ein Fehler 2. Art auch mit größerer W. auftreten. Dies wird man bei der Wahl von H und K berücksichtigen: Man wird H und K i. d. R. so wählen, dass ein Fehler 1. Art die schwerwiegenden Konsequenzen besitzt. Häufig ist H der „Normalzustand“, und die Entscheidung für K soll nur dann erfolgen, wenn die Indizienlage eindeutig ist.

Beispiel 1: Glücksspiel \rightarrow nur teilnehmen, wenn das Spiel „eindeutig vorteilhaft“ ist

Beispiel 2: Pharmazie \rightarrow Medikament nur einführen, wenn es „deutlich besser“ ist

Angenommen, wir wollen $H : \vartheta = \vartheta_0$ gegen $K : \vartheta \neq \vartheta_0$ testen und wir haben bereits einen Konfidenzbereich C zum Niveau $1 - \alpha$ für ϑ konstruiert. Dann ist es nahe liegend, die Hypothese $H : \vartheta = \vartheta_0$ genau dann zu verwerfen, wenn ϑ_0 nicht in C liegt:

$$\varphi(x) := \begin{cases} 1 & \text{falls } \vartheta_0 \notin C(x) \\ 0 & \text{falls } \vartheta_0 \in C(x) \end{cases} \quad (6.6)$$

Dieser Test ist wegen

$$\mathbb{E}_{\vartheta_0}(\varphi(X)) = \mathbb{P}_{\vartheta_0}(\varphi(X) = 1) = 1 - \mathbb{P}_{\vartheta_0}(\vartheta_0 \in C(X)) \leq 1 - (1 - \alpha) = \alpha$$

in der Tat ein Test zum Niveau α .

Beispiele 6.24.

(a) (Zweiseitiger Gauß-Test)

Gegeben sei die Situation aus Beispiel 6.8 (b), und gesucht sei ein Test zum Niveau α für $H : \mu = \mu_0$ gegen $K : \mu \neq \mu_0$.

Da wir schon einen Konfidenzbereich zum Niveau $1 - \alpha$ für μ kennen, erhalten wir

$$\varphi(x) := \begin{cases} 1 & \mu_0 \notin [\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}] \\ 0 & \mu_0 \in [\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{x}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\frac{\alpha}{2}}] \end{cases}$$

bzw. nach Umformung:

Zweiseitiger Gauß-Test

$$\varphi(x) = \begin{cases} 1 & \sqrt{n} |\bar{x}_n - \mu_0| / \sigma_0 > z_{1-\frac{\alpha}{2}}, \\ 0 & \sqrt{n} |\bar{x}_n - \mu_0| / \sigma_0 \leq z_{1-\frac{\alpha}{2}}, \end{cases}$$

Dabei ist $z_{1-\frac{\alpha}{2}}$ wieder das $(1 - \frac{\alpha}{2})$ -Quantil der $\mathcal{N}_{0,1}$ -Verteilung.

(b) (Zweiseitiger t -Test)

Gegeben sei die Situation aus Beispiel 6.8 (c), und gesucht sei ein Test zum Niveau α für $H : \mu = \mu_0$ gegen $K : \mu \neq \mu_0$.

Da wir schon einen Konfidenzbereich zum Niveau $1 - \alpha$ für μ kennen, erhalten wir analog zum obigen Ansatz

$$\varphi(x) := \begin{cases} 1 & \mu_0 \notin [\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}] \\ 0 & \mu_0 \in [\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}] \end{cases}$$

bzw. nach Umformung:

Zweiseitiger t -Test

$$\varphi(x) = \begin{cases} 1 & \sqrt{n} |\bar{x}_n - \mu_0| / s_n > t_{n-1, 1-\frac{\alpha}{2}}, \\ 0 & \sqrt{n} |\bar{x}_n - \mu_0| / s_n \leq t_{n-1, 1-\frac{\alpha}{2}}, \end{cases}$$

Dabei ist $t_{n-1, 1-\frac{\alpha}{2}}$ wieder das $(1 - \frac{\alpha}{2})$ -Quantil der t_{n-1} -Verteilung.

(c) (Zweiseitiger Binomialtest mit Normalapproximation)

Gegeben sei die Situation aus Beispiel 6.8 (a); gesucht sei ein Test zum Niveau α für $H : p = p_0$ gegen $K : p \neq p_0$, wobei $p_0 \in (0, 1)$ fest.

Hier empfiehlt es sich, den Test nicht über das approximative Konfidenzintervall zum Niveau $1 - \alpha$ (welches eine Schätzung der Varianz erfordert), sondern „direkt“ zu bestimmen. Da $\frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}}$ unter \mathbb{P}_{p_0} nach dem ZWGS für große Werte von n näherungsweise $\mathcal{N}(0, 1)$ -verteilt ist, ist der folgende Ansatz nahe liegend:

$$\varphi(x) := \begin{cases} 1 & \left| \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \right| > t, \\ 0 & \left| \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \right| \leq t, \end{cases} \quad \text{wobei} \quad \mathbb{P}_{p_0} \left(\left| \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \right| > t \right) \stackrel{!}{\leq} \alpha.$$

Mittels Normalapproximation (ohne SK) ergibt sich

$$\Phi(+t) - \Phi(-t) \geq 1 - \alpha \iff 2\Phi(t) - 1 \geq 1 - \alpha \iff t = \Phi^{-1}(1 - \frac{\alpha}{2}),$$

so dass wir als approximativen Test zum Niveau α das folgende Ergebnis erhalten:

Zweiseitiger Binomialtest mit Normalapproximation

$$\varphi(x) = \begin{cases} 1 & \sqrt{n} \left| \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \right| > z_{1-\frac{\alpha}{2}}, \\ 0 & \sqrt{n} \left| \frac{\bar{x}_n - p_0}{\sqrt{p_0(1-p_0)}} \right| \leq z_{1-\frac{\alpha}{2}}, \end{cases}$$

Dabei ist $z_{1-\frac{\alpha}{2}}$ wieder das $(1 - \frac{\alpha}{2})$ -Quantil der $\mathcal{N}_{0,1}$ -Verteilung.

Hinweise zum exakten Binomialtest finden sich z. B. bei HENZE.

Beispiel 6.25. In einem Betrieb sollen Schrauben mit einer Länge von 5 cm hergestellt werden. Es sei bekannt, dass die tatsächlichen Schraubenlängen in Zentimetern (aufgrund von zufälligen Schwankungen im Produktionsprozess) $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind. Aufgrund einer Stichprobe

5.130 5.186 4.983 4.951 5.125 4.910 5.152 4.829 5.011 5.203 4.877 5.093

ist zum Irrtumsniveau 5% zu testen, ob der Mittelwert μ mit dem Zielwert $\mu_0 = 5$ [cm] übereinstimmt. Gesucht ist also ein Test z. N. $\alpha = 0.05$ für $H : \mu = \mu_0$ gegen $K : \mu \neq \mu_0$.

1. *Fall:* Die Standardabweichung der Schraubenlängen ist bekannt: 0.1 [cm]

Mit $n = 12$, $\bar{x}_n = 5.0375$, $\sigma_0 = 0.1$, $z_{0.975} = 1.96$ ergibt sich

$$\sqrt{n}|\bar{x}_n - 5|/\sigma_0 = 1.30 < 1.96,$$

so dass die Nullhypothese $H : \mu = \mu_0$ nicht verworfen werden kann.

2. *Fall:* Die Standardabweichung der Schraubenlängen ist unbekannt.

Mit $n = 12$, $\bar{x}_n = 5.0375$, $s_n = 0.1274$, $t_{11,0.975} = 2.20$ ergibt sich

$$\sqrt{n}|\bar{x}_n - 5|/s_n = 1.02 < 2.20,$$

so dass die Nullhypothese $H : \mu = \mu_0$ nicht verworfen werden kann.

Hat man die Konfidenzbereiche in Beispiel 6.20 schon berechnet, so sind diese Testergebnisse natürlich wegen (6.6) voraussehbar, da in beiden Fällen $\mu_0 \in C(x)$ gilt.

Anwendung von Tests: p -Werte

Häufig (z. B. in vielen Statistik-Programmen) wird bei der Durchführung von Tests zu gegebener Stichprobe $x \in \mathcal{X}$ der sog. p -Wert angegeben – dies ist das größte Niveau, zu dem bei Beobachtung von x an der Nullhypothese H festgehalten wird:

$$p\text{-Wert}(x) = \max\{\alpha \in [0, 1] : \varphi_\alpha(x) = 0\}$$

Dabei sei für jedes $\alpha \in [0, 1]$ φ_α ein Test z. N. α für H gegen K , und wir unterstellen, dass für jedes $x \in \mathcal{X}$

$$\{\alpha \in [0, 1] : \varphi_\alpha(x) = 0\} = [0; p\text{-Wert}(x)]$$

gilt, was in der Praxis in der Regel der Fall ist. Die Entscheidungsregel lautet dann

$$\begin{array}{ll} \text{Entscheidung für } H & \text{falls } p\text{-Wert}(x) \geq \alpha, \\ \text{Entscheidung für } K & \text{falls } p\text{-Wert}(x) < \alpha. \end{array}$$

Wichtig ist hier allerdings, dass das Niveau α *vor* Durchführung des Tests festgelegt werden muss: Wenn man das Niveau bei gegebenen Daten so an den p -Wert anpasst, dass man die Nullhypothese H verwerfen kann, so verwirft man die Nullhypothese H ja *immer* und begeht damit im Fall $\vartheta \in H$ m. W. 1 einen Fehler 1. Art. – Generell ist bei sauberer statistischer Arbeit das statistische Verfahren festzulegen, *bevor* die Daten betrachtet bzw. ausgewertet werden.

Ergänzende Bemerkungen

(a) Neben dem zweiseitigen Gauß-Test bzw. t -Test gibt es den einseitigen Gauß-Test bzw. t -Test. Diese Tests können zum Testen der Hypothesen

$$H : \mu \leq \mu_0 \text{ gegen } K : \mu > \mu_0 \quad \Bigg| \quad H : \mu \geq \mu_0 \text{ gegen } K : \mu < \mu_0$$

herangezogen werden, falls σ^2 bekannt bzw. unbekannt ist. Sie können aus einseitigen Konfidenzintervallen der Form

$$C(x) = [C_u(x), +\infty[\quad \Bigg| \quad C(x) =]-\infty, C_o(x)]$$

für μ gewonnen werden; die darin auftretenden Funktionen $C_u(x)$ und $C_o(x)$ werden auch als untere und obere *Konfidenzschranken* für μ bezeichnet.

Man erhält dann bei bekannter Varianz

$$C(x) = [\bar{x}_n - \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha}, +\infty[\quad \Bigg| \quad C(x) =]-\infty, \bar{x}_n + \frac{\sigma_0}{\sqrt{n}} z_{1-\alpha}]$$

bzw.

$$\varphi(x) = \begin{cases} 1 & \sqrt{n}(\bar{x}_n - \mu_0)/\sigma_0 > +z_{1-\alpha}, \\ 0 & \sqrt{n}(\bar{x}_n - \mu_0)/\sigma_0 \leq +z_{1-\alpha}, \end{cases} \quad \Bigg| \quad \varphi(x) = \begin{cases} 1 & \sqrt{n}(\bar{x}_n - \mu_0)/\sigma_0 < -z_{1-\alpha}, \\ 0 & \sqrt{n}(\bar{x}_n - \mu_0)/\sigma_0 \geq -z_{1-\alpha}, \end{cases}$$

und bei unbekannter Varianz

$$C(x) = [\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\alpha}, +\infty[\quad \Bigg| \quad C(x) =]-\infty, \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\alpha}]$$

bzw.

$$\varphi(x) = \begin{cases} 1 & \sqrt{n}(\bar{x}_n - \mu_0)/s_n > +t_{n-1, 1-\alpha}, \\ 0 & \sqrt{n}(\bar{x}_n - \mu_0)/s_n \leq +t_{n-1, 1-\alpha}, \end{cases} \quad \Bigg| \quad \varphi(x) = \begin{cases} 1 & \sqrt{n}(\bar{x}_n - \mu_0)/s_n < -t_{n-1, 1-\alpha}, \\ 0 & \sqrt{n}(\bar{x}_n - \mu_0)/s_n \geq -t_{n-1, 1-\alpha}, \end{cases}$$

(b) Interessiert man sich (in der Situation, dass bei unabhängigen normalverteilten Beobachtungen sowohl μ als auch σ^2 unbekannt sind) nicht für den Erwartungswert μ , sondern für die Varianz σ^2 , so ist bei festem $\sigma_0^2 > 0$ ein Test zum Niveau α für $H : \sigma^2 \leq \sigma_0^2$ gegen $K : \sigma^2 > \sigma_0^2$ durch den sogenannten *Chi-Quadrat-Test für die Varianz*

$$\varphi(x) := \begin{cases} 1 & \sum_{i=1}^n (x_i - \bar{x}_n)^2 / \sigma_0^2 > \chi_{n-1, 1-\alpha}^2, \\ 0 & \sum_{i=1}^n (x_i - \bar{x}_n)^2 / \sigma_0^2 \leq \chi_{n-1, 1-\alpha}^2, \end{cases}$$

gegeben, wobei $\chi_{n-1, 1-\alpha}^2$ das $(1 - \alpha)$ -Quantil der sogenannten *Chi-Quadrat-Verteilung mit $n - 1$ Freiheitsgraden* (kurz χ_{n-1}^2) bezeichnet. Für Details sei auf die Literatur verwiesen.

(c) Häufig stellt sich das Problem, dass zwei unterschiedliche Verfahren verglichen werden sollen. Angenommen, Verfahren I liefert die Messungen x_1, \dots, x_m und Verfahren II liefert die Messungen y_1, \dots, y_n , wobei die Messungen als Realisierungen von *unabhängigen* normalverteilten Zufallsgrößen $X_1, \dots, X_m \sim \mathcal{N}(\mu_1, \sigma^2)$ bzw. $Y_1, \dots, Y_n \sim \mathcal{N}(\mu_2, \sigma^2)$ mit derselben Varianz angesehen werden können. Dann ist ein Test zum Niveau α für $H : \mu_1 \leq \mu_2$ gegen $K : \mu_1 > \mu_2$ bei bekannter Varianz σ_0^2 durch den *Gauß-Test zum Vergleich zweier Stichproben*

$$\varphi(x, y) := \begin{cases} 1 & \sqrt{\frac{mn}{m+n}} \frac{\bar{x}_m - \bar{y}_n}{\sigma_0} > z_{1-\alpha}, \\ 0 & \sqrt{\frac{mn}{m+n}} \frac{\bar{x}_m - \bar{y}_n}{\sigma_0} \leq z_{1-\alpha}, \end{cases}$$

und bei unbekannter Varianz σ^2 durch den *t-Test zum Vergleich zweier Stichproben*

$$\varphi(x, y) := \begin{cases} 1 & \sqrt{\frac{mn}{m+n}} \frac{\bar{x}_m - \bar{y}_n}{s_{m,n}} > t_{m+n-2, 1-\alpha}, \\ 0 & \sqrt{\frac{mn}{m+n}} \frac{\bar{x}_m - \bar{y}_n}{s_{m,n}} \leq t_{m+n-2, 1-\alpha}, \end{cases}$$

gegeben, wobei $s_{m,n}^2 := \frac{1}{m+n-2} (\sum_{i=1}^m (x_i - \bar{x}_m)^2 + \sum_{i=1}^n (y_i - \bar{y}_n)^2)$. Für Details sei wieder auf die Literatur verwiesen.

Zusammenfassung:

- Mit Hilfe von statistischen Verfahren lassen sich ausgehend von Daten Rückschlüsse auf die zugrunde liegenden Gesetzmäßigkeiten ziehen.
- Schätzer liefern Werte für unbekannte Parameter.
- Konfidenzbereiche liefern Mengen, die die unbekannten Parameter mit hoher Sicherheit enthalten, und vermitteln dadurch einen Eindruck von der Genauigkeit einer Schätzung.
- Mit Tests lassen sich Hypothesen über unbekannte Parameter überprüfen.
- Es gibt verschiedene Ansätze zur Konstruktion bzw. zur Beurteilung „guter“ statistischer Verfahren, die nicht immer zum selben Ergebnis führen müssen.

A Grundwissen Mathematik

In diesem Kapitel sind einige grundlegende Begriffe und Sätze aus der Mathematik zusammengestellt, die für die Stochastik von besonderer Bedeutung sind und die wir als bekannt voraussetzen. Die Aufstellung erhebt keinen Anspruch auf Vollständigkeit.

A.1 Mengen

Wir beginnen mit einer Einführung in die (naive) Mengenlehre.

Eine *Menge* entsteht durch Zusammenfassung von verschiedenen Objekten zu einem Ganzen.

Im einfachsten Fall kann man eine Menge dadurch beschreiben, dass man die Objekte, die die Menge bilden, auflistet und geschweifte Klammern darum setzt. Dabei sind die Reihenfolge der Objekte sowie die Vielfachheiten, mit denen die Objekte auftreten, unerheblich.

Beispiele:

- $\{1, 2, 3\}$: Menge, die aus den Zahlen 1,2,3 besteht
- $\{2, 1, 3\}$: dieselbe Menge
- $\{3, 1, 2, 1, 2, 1\}$: nochmals dieselbe Menge
- $\Sigma = \{a, b, c, \dots, x, y, z\}$: Menge aller (Klein-)Buchstaben
- S : Menge der Studierenden, die die Veranstaltung „Stochastik“ besuchen
- $\mathbb{N} = \{1, 2, 3, \dots\}$: Menge der natürlichen Zahlen (ohne die Null)
- $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$: Menge der natürlichen Zahlen einschließlich der Null
- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$: Menge der ganzen / rationalen / reellen / komplexen Zahlen
- $\mathcal{C}(\mathbb{R}, \mathbb{R})$: Menge der stetigen Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$
- $\emptyset := \{ \}$: leere Menge
- $\{\{1\}, \{2\}, \{3\}\}$: Menge aller einelementigen Teilmengen der Menge $\{1, 2, 3\}$

Die Objekte, die zu einer Menge gehören, bezeichnet man auch als *Elemente* der Menge. Ist ein Objekt a ein [kein] Element einer Menge A , so schreibt man $a \in A$ [$a \notin A$]. Beispielsweise gilt $\frac{1}{2} \in \mathbb{Q}$ und $\sqrt{2} \notin \mathbb{Q}$.

Eine Menge M heißt *endliche Menge*, wenn sie aus nur endlich vielen Elementen besteht, und sonst *unendliche Menge*.

Beispiele:

$\{1, 2, 3\}$ ist eine endliche Menge, $\mathbb{N} = \{1, 2, 3, \dots\}$ und \mathbb{R} sind unendliche Mengen.

In der Mathematik ist es wichtig, zwischen verschiedenen Arten von unendlichen Mengen zu unterscheiden: Eine unendliche Menge M heißt *abzählbar-unendlich*, wenn sie sich der Form $M = \{a_1, a_2, a_3, \dots\}$ angeben lässt, und sonst *überabzählbar*.

Beispiele:

\mathbb{N} ist abzählbar-unendlich, \mathbb{R} ist überabzählbar.

Eine Menge heißt *abzählbar*, wenn sie entweder endlich oder abzählbar-unendlich ist. Die abzählbaren Mengen sind in der Stochastik von besonderem Interesse, weil sie die Grundmengen der diskreten W.räume bilden.

Die Anzahl der Elemente einer Menge M wird auch als *Mächtigkeit* (oder *Kardinalität*) von M bezeichnet, kurz $|M|$, wobei wir für eine unendliche Menge $|M| := \infty$ setzen. (Streng genommen könnte / müsste man auch hier noch zwischen verschiedenen Arten von Unendlich unterscheiden, aber dies ist für unsere Zwecke unwichtig.)

Beispiele:

$$|\{1, 2, 3\}| = 3, |\{3, 1, 2, 1, 2, 1\}| = 3, |\mathbb{N}| = \infty, |\mathbb{R}| = \infty.$$

Ist Ω eine beliebige Menge, so heißt eine Menge A *Teilmenge* von Ω (kurz: $A \subseteq \Omega$), wenn jedes Element von A auch ein Element von Ω ist. Ist Ω eine Menge, so ist das System aller Teilmengen $A \subseteq \Omega$ (einschließlich der leeren Menge \emptyset sowie der Menge Ω selbst) wieder eine Menge, die sog. *Potenzmenge* von Ω (kurz: $\mathfrak{P}(\Omega)$).

Beispiele:

$$\mathfrak{P}(\{1, 2, 3\}) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

Allgemein gilt: Ist M eine endliche Menge mit n Elementen ($n \in \mathbb{N}_0$), so besitzt $\mathfrak{P}(\Omega)$ 2^n Elemente.

Meistens erhält man Teilmengen $A \subseteq \Omega$ dadurch, dass man nur die Elemente von Ω betrachtet, die eine bestimmte Eigenschaft E besitzen; man schreibt dann

$$A = \{\omega \in \Omega \mid \omega \text{ hat die Eigenschaft } E\}$$

(lies: A ist die Menge aller Elemente von Omega, die die Eigenschaft E besitzen).

Beispiel:

$$A = \{n \in \mathbb{Z} \mid n \text{ ist durch } 2 \text{ teilbar}\} : \text{Menge der geraden Zahlen}$$

Als Teilmengen von \mathbb{R} treten häufig *Intervalle* auf: Für beliebiges $a, b \in \mathbb{R}$ mit $a \leq b$ setzen wir beispielsweise:

- $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ (abgeschlossenes Intervall)
- $]a, b[:= \{x \in \mathbb{R} : a < x < b\}$ (offenes Intervall)
- $]a, b] := \{x \in \mathbb{R} : a < x \leq b\}$ (linksseitig offenes, rechtsseitig abgeschlossenes Intervall)
- $]a, \infty[:= \{x \in \mathbb{R} : x > a\}$ (offenes uneigentliches Intervall)
- $]-\infty, b] := \{x \in \mathbb{R} : x \leq b\}$ (abgeschlossenes uneigentliches Intervall)

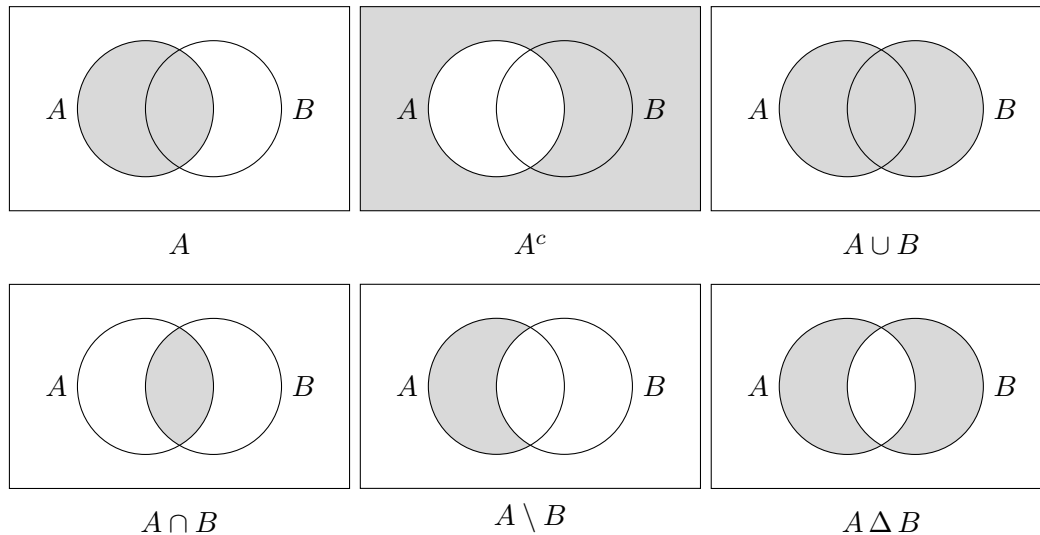
Je nachdem, ob die eckigen Klammern nach innen oder nach außen zeigen, gehören die „Intervallgrenzen“ also zum Intervall dazu oder nicht. Natürlich gibt es weitere Kombinationen, die nun allerdings selbsterklärend sein sollten. Anstelle nach außen geöffnete eckiger Klammern verwendet man (gleichwertig) nach innen geöffnete runde Klammern; es ist also z. B. $(a, b) =]a, b[$ und $(0, \infty) =]0, \infty[$ die Menge der positiven reellen Zahlen. Schließlich sei noch angemerkt, dass im Spezialfall $a = b$ $[a, a] = \{a\}$, $(a, a) = \emptyset$ sowie $(a, a] = \emptyset$ gilt.

Sind Ω eine feste Bezugsmenge und $A, B \subseteq \Omega$, so kann man neue Mengen konstruieren:

- $A^c = \{\omega \in \Omega \mid \omega \notin A\}$ (Komplement von A)
- $A \cup B = \{\omega \in \Omega \mid \omega \in A \vee \omega \in B\}$ (Vereinigung von A und B)
- $A \cap B = \{\omega \in \Omega \mid \omega \in A \wedge \omega \in B\}$ (Durchschnitt von A und B)
- $A \setminus B = \{\omega \in \Omega \mid \omega \in A \wedge \omega \notin B\}$
(mengentheoretische Differenz von A und B)
- $A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B) = \{\omega \in \Omega \mid \omega \in A \text{ xor } \omega \in B\}$
(symmetrische Differenz von A und B)

Dabei bezeichnet \vee das nicht-ausschließende Oder, \wedge das Und und xor das ausschließende Oder. Es sei angemerkt, dass wir bei der mengentheoretischen Differenz $A \setminus B$ nicht $B \subseteq A$ voraussetzen, wie schon die Formel für die symmetrische Differenz vermuten lässt.

Diese Mengenoperationen lassen sich durch Diagramme veranschaulichen:



Für diese Mengenoperationen gelten die folgenden Rechenregeln (wobei $A, B, C \subseteq \Omega$):

- $(A^c)^c = A$ (Involution)
- $(A \cup B)^c = A^c \cap B^c$ (de Morgan'sche Regel)
- $(A \cap B)^c = A^c \cup B^c$ (de Morgan'sche Regel)
- $A \cup B = B \cup A$ (Kommutativgesetz)
- $A \cap B = B \cap A$ (Kommutativgesetz)
- $(A \cup B) \cup C = A \cup (B \cup C)$ (Assoziativgesetz)
- $(A \cap B) \cap C = A \cap (B \cap C)$ (Assoziativgesetz)
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (Distributivgesetz)
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributivgesetz)
- $(A \cup B) \cap A = A$ (Absorptionsgesetz)
- $(A \cap B) \cup A = A$ (Absorptionsgesetz)

Die Mengenoperationen Vereinigung und Durchschnitt sind sogar für beliebige Familien von Teilmengen von Ω sinnvoll: Ist $(A_i)_{i \in I}$ eine beliebige Familie von Teilmengen von Ω (d. h. ist I eine beliebige *Indexmenge* und für jedes $i \in I$ eine Menge $A_i \subseteq \Omega$ gegeben), so setzt man

$$\bigcup_{i \in I} A_i := \{\omega \in \Omega \mid \exists i \in I : \omega \in A_i\} \quad \text{und} \quad \bigcap_{i \in I} A_i := \{\omega \in \Omega \mid \forall i \in I : \omega \in A_i\}.$$

Ist speziell $I = \mathbb{N}$, so schreibt man auch $\bigcup_{i=1}^{\infty} A_i$ bzw. $\bigcap_{i=1}^{\infty} A_i$.

Es gelten dann u. a. die folgenden Rechenregeln (wobei $B \subseteq \Omega$):

$$(\bigcup_{i \in I} A_i)^c = \bigcap_{i \in I} A_i^c \quad (\text{de Morgan'sche Regel})$$

$$(\bigcap_{i \in I} A_i)^c = \bigcup_{i \in I} A_i^c \quad (\text{de Morgan'sche Regel})$$

$$B \cup (\bigcap_{i \in I} A_i) = \bigcap_{i \in I} (B \cup A_i) \quad (\text{Distributivgesetz})$$

$$B \cap (\bigcup_{i \in I} A_i) = \bigcup_{i \in I} (B \cap A_i) \quad (\text{Distributivgesetz})$$

In der Stochastik werden allerdings nur abzählbare Vereinigungen und Durchschnitte (d. h. Vereinigungen und Durchschnitte von abzählbar vielen Mengen) auftreten.

A.2 Tupel

Ein n -*Tupel* ist eine Liste von n Objekten, wobei Wiederholungen zugelassen sind. Tupel der Länge 2 und 3 heißen auch *Paare* bzw. *Tripel*. Im Gegensatz zu Mengen kommt es hier darauf an, in welcher Reihenfolge und wie oft die Objekte auftreten.

Beispiel: Es gilt $\{1, 2, 3\} = \{2, 1, 3\} = \{1, 2, 3, 1\}$, weil es bei Mengen (per Konvention!) keinen Unterschied macht, in welcher Reihenfolge bzw. wie oft man die Elemente angibt. Hingegen gilt $(1, 2, 3) \neq (2, 1, 3)$ und $(1, 2, 3) \neq (1, 2, 3, 1)$.

Zwei Tupel (a_1, \dots, a_n) und (b_1, \dots, b_m) sind also genau dann gleich, wenn $n = m$ und $a_i = b_i$ für alle $i = 1, \dots, n$ gilt.

Sind A und B nicht-leere Mengen, so kann man das *kartesische Produkt* von A und B (kurz: $A \times B$) einführen:

$$A \times B = \{(a, b) \mid a \in A, b \in B\}$$

Dies ist also die Menge aller Paare, deren erste [zweite] Komponente in A [B] liegt.

Beispiel:

Ist $A = \{1, 2, 3\}$ und $B = \{3, 4\}$, so gilt $A \times B = \{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4)\}$.

Allgemein gilt $|A \times B| = |A| \cdot |B|$; dies erklärt übrigens auch die Bezeichnungen *Produkt* bzw. \times .

Entsprechend führt man kartesische Produkte $A_1 \times \cdots \times A_n$ von mehr als zwei ($n > 2$) nicht-leeren Mengen ein; man erhält dann die Menge aller n -Tupel der Form (a_1, \dots, a_n) mit $a_i \in A_i$ für alle $i = 1, \dots, n$. Es gilt dann $|A_1 \times \cdots \times A_n| = |A_1| \cdot \cdots \cdot |A_n|$.

Im Fall $A_1 = \cdots = A_n = A$ schreibt man statt $A \times \cdots \times A$ auch A^n (*kartesische Potenz*). Es gilt dann $|A^n| = |A|^n$.

Man kann sogar noch allgemeinere kartesische Produkte einführen: Ist $(A_i)_{i \in I}$ eine Familie von nicht-leeren Mengen (d. h. ist I eine beliebige *Indexmenge* und ist für jedes $i \in I$ eine nicht-leere Menge A_i gegeben), so setzt man

$$\prod_{i \in I} A_i = \{(a_i)_{i \in I} \mid a_i \in A_i \text{ für alle } i \in I\}.$$

Für uns ist vor allem der Fall $I = \mathbb{N}$ und $A_n = A$ für alle $n \in \mathbb{N}$ von Interesse; hier ist $A^\infty := \prod_{n \in \mathbb{N}} A$ die Menge aller unendlichen Folgen mit Komponenten in A .

A.3 Abbildungen

Seien A, B nicht-leere Mengen. Eine *Abbildung* (oder auch *Funktion*) $f : A \rightarrow B$ ordnet jedem Element $a \in A$ genau ein Element $f(a) = b \in B$ zu. Die vollständige Beschreibung einer Abbildung umfasst also

- die Angabe der *Ausgangsmenge* A ,
- die Angabe der *Zielmenge* B ,
- eine Beschreibung der Zuordnungsvorschrift, etwa durch eine mathematische Formel oder eine verbale Formulierung.

Die Menge A wird auch als *Definitionsbereich* von f bezeichnet.

Beispiel:

Sei $n \in \mathbb{N}$ fest. Die Abbildung $f : \{0, 1\}^n \rightarrow \mathbb{N}_0$ mit $f(x_1, \dots, x_n) := \sum_{i=1}^n x_i$ ordnet jedem n -Tupel über der Menge $\{0, 1\}$ die Anzahl der darin vorkommenden Einsen zu.

Abbildungen zwischen (kleinen) endlichen Mengen lassen sich durch Pfeildiagramme veranschaulichen. Zwischen zwei Elementen $a \in A$ und $b \in B$ wird genau dann ein Pfeil eingezeichnet, wenn $f(a) = b$ gilt. Von jedem Element $a \in A$ muss also genau ein Pfeil ausgehen.

Sind $f : A \rightarrow B$ und $g : B \rightarrow C$ Abbildungen, wobei die Zielmenge der ersten Abbildung mit der Ausgangsmenge der zweiten Abbildung übereinstimmt, so kann man die *Hinter-einanderschaltung* (*Komposition*) von f und g einführen:

$$g \circ f : A \rightarrow C \quad \text{mit} \quad (g \circ f)(a) := g(f(a)) \quad (a \in A)$$

In der Sprache der Pfeildiagramme erhält man $g \circ f$ dadurch, dass man die Pfeile zu f und g „hintereinanderlegt“. In der Stochastik ist es (im Gegensatz zu den meisten anderen Teilgebieten der Mathematik!) zulässig und sogar üblich, statt $g \circ f$ auch $g(f)$ zu schreiben.

Sei $f : A \rightarrow B$. Gilt $f(a) = b$, so heißt b *Bild von a unter f* und a *Urbild von b unter f* . Allgemeiner heißt für $A_0 \subseteq A$

$$f(A_0) := \{b \in B \mid \exists a \in A_0 : f(a) = b\} \subseteq B$$

Bild(menge) von A_0 unter f und für $B_0 \subseteq B$

$$f^{-1}(B_0) := \{a \in A \mid f(a) \in B_0\} \subseteq A$$

Urbild(menge) von B_0 unter f .

Eine Abbildung $f : A \rightarrow B$ heißt ...

- *injektiv*, wenn $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$ gilt
(wenn also jedes Element $b \in B$ *höchstens einmal* als Funktionswert auftritt),
- *surjektiv*, wenn $f(A) = B$ gilt
(wenn also jedes Element $b \in B$ *mindestens einmal* als Funktionswert auftritt),
- *bijektiv*, wenn f sowohl injektiv als auch surjektiv ist
(wenn also jedes Element $b \in B$ *genau einmal* als Funktionswert auftritt).

Ist $f : A \rightarrow B$ eine bijektive Abbildung, so kann man die *Umkehrabbildung* (oder auch *Umkehrfunktion*) $f^{-1} : B \rightarrow A$ einführen, die jedem $b \in B$ das eindeutige Element $a \in A$ mit $f(a) = b$ zuordnet. In der Sprache der Pfeildiagramme erhält man die Umkehrabbildung dadurch, dass man alle Pfeile umdreht.

Beachte dabei, dass wir die Bezeichnung f^{-1} für unterschiedliche Zwecke verwenden, zum einen zur Bezeichnung von Urbildmengen (die immer definiert sind, auch wenn f nicht bijektiv ist) und zum anderen zur Bezeichnung der Umkehrabbildung (die nur dann definiert ist, wenn f bijektiv ist).

Sind A und B endliche Mengen und ist $f : A \rightarrow B$ bijektiv, so folgt $|A| = |B|$. Dieses Grundprinzip der Kombinatorik ist manchmal bei der Bestimmung von Anzahlen nützlich, vgl. z. B. den Beweis zu Fall IV in Satz 1.9.

A.4 Folgen und Reihen

Eine Folge reeller Zahlen ist eine Familie $(a_n)_{n \in \mathbb{N}}$ mit $a_n \in \mathbb{R}$ für alle $n \in \mathbb{N}$. Offenbar kann man eine solche Folge auch als Funktion $a : \mathbb{N} \rightarrow \mathbb{R}$ mit $a(n) := a_n$ für alle $n \in \mathbb{N}$ ansehen und umgekehrt.

Definition A.1. Eine Folge $(a_n)_{n \in \mathbb{N}}$ heißt *konvergent*, wenn es eine Zahl $a \in \mathbb{R}$ gibt, so dass für jedes $\varepsilon > 0$ ein $N \in \mathbb{N}$ mit $|a_n - a| < \varepsilon$ für alle $n \geq N$ existiert.

Die Zahl a ist dann eindeutig bestimmt und wird als *Grenzwert* der Folge bezeichnet; man schreibt $\lim_{n \rightarrow \infty} a_n = a$.

Es ist zweckmäßig, auch die Werte $\pm\infty$ als Grenzwerte „im weiteren Sinne“ zuzulassen: Eine Folge $(a_n)_{n \in \mathbb{N}}$ heißt *bestimmt divergent* gegen $+\infty$ $[-\infty]$, wenn zu jedem $K > 0$ ein $N \in \mathbb{N}$ existiert, so dass für alle $n \geq N$ $a_n > +K$ [$a_n < -K$] gilt. Man schreibt dann $\lim_{n \rightarrow \infty} a_n = +\infty$ $[-\infty]$.

Beispiele A.2.

- Die Folge $(a_n)_{n \in \mathbb{N}}$ mit $a_n = \frac{1}{n}$ ist konvergent mit Grenzwert 0.
- Die Folge $(a_n)_{n \in \mathbb{N}}$ mit

$$a_n = \begin{cases} \frac{1}{n} & n \text{ gerade,} \\ 0 & n \text{ ungerade,} \end{cases}$$

ist ebenfalls konvergent mit Grenzwert 0. Dieses Beispiel zeigt, dass Konvergenz nicht „monoton“ vonstatten gehen muss.

- Die Folge $(a_n)_{n \in \mathbb{N}}$ mit $a_n = b^n$ ($b \in \mathbb{R}$ fest) ist konvergent mit Grenzwert 0, falls $|b| < 1$, konvergent mit Grenzwert 1, falls $b = 1$, und bestimmt divergent gegen ∞ , falls $b > 1$. (Für $b \leq -1$ ist die Folge weder konvergent noch bestimmt divergent.)

Ein Ausdruck der Form $\sum_{n=1}^{\infty} a_n$ mit $a_n \in \mathbb{R}$ für alle $n \in \mathbb{N}$ heißt *unendliche Reihe*.

Definition A.3. Eine Reihe $\sum_{k=1}^{\infty} a_k$ heißt *konvergent*, falls die Folge der zugehörigen Partialsummen $(s_n)_{n \in \mathbb{N}}$ mit $s_n := \sum_{k=1}^n a_k$ für alle $n \in \mathbb{N}$ konvergent ist.

In diesem Fall heißt der Grenzwert $s := \lim_{n \rightarrow \infty} s_n$ auch *Wert* der Reihe $\sum_{n=1}^{\infty} a_n$; man schreibt $\sum_{n=1}^{\infty} a_n = s$.

Auch hier kann man bestimmt divergente Reihen mit dem Wert $+\infty$ bzw. $-\infty$ einführen.

Beispiele A.4.

- Die Reihe $\sum_{n=1}^{\infty} \frac{1}{n}$ ist bestimmt divergent gegen ∞ . (*harmonische Reihe*)
- Die Reihe $\sum_{n=1}^{\infty} \frac{1}{n^s}$ ($s > 0$) ist konvergent, falls $s > 1$, und bestimmt divergent gegen $+\infty$, falls $s \leq 1$. Auch wenn man zeigen kann, dass die Reihe für $s > 1$ stets konvergent ist, ist es nur in Ausnahmefällen möglich, den Wert der Reihe zu berechnen. Für $s = 2$ ergibt sich z. B. der Wert $\frac{1}{6}\pi^2$.
- Die Reihe $\sum_{n=1}^{\infty} x^n$ ($x \in \mathbb{R}$) ist konvergent mit Wert $\frac{1}{1-x}$, falls $|x| < 1$. (*geometrische Reihe*) Für $x \geq +1$ ist sie bestimmt divergent gegen $+\infty$, für $x \leq -1$ ist sie weder konvergent noch bestimmt divergent.

Man kann sich fragen, ob sich der Wert einer Reihe ändern kann, wenn man die Elemente „umordnet“ (wobei jedes Element nach dem Umordnen genau so oft vorkommen muss wie vor dem Umordnen). Das folgende Beispiel zeigt, dass dies möglich ist:

Beispiel A.5 (Umordnen von Summen). Die Reihe $\sum_{n=1}^{\infty} a_n$ mit $a_n := \frac{(-1)^{n-1}}{n}$ ist konvergent mit dem Wert $\log 2$. (Dabei ist \log die natürliche Logarithmusfunktion; siehe unten.) Es gilt $a_n \rightarrow 0$, die Glieder a_{2n-1} sind positiv mit $\sum_{n=1}^{\infty} a_{2n-1} = +\infty$, und die Glieder a_{2n} sind negativ mit $\sum_{n=1}^{\infty} a_{2n} = -\infty$. Damit kann man zu einer *beliebigen* reellen Zahl $b \in \mathbb{R}$ eine Umordnung $\sum_{n=1}^{\infty} b_n$ von $\sum_{n=1}^{\infty} a_n$ konstruieren, die den Wert b besitzt:

- Beginne mit der „leeren“ Summe 0.
- Addiere positive Elemente $a_1, a_3, \dots, a_{2n_1-1}$, bis die Summe den Wert b überschreitet.
- Addiere negative Elemente $a_2, a_4, \dots, a_{2m_1}$, bis die Summe den Wert b unterschreitet.
- Addiere positive Elemente $a_{2n_1+1}, a_{2n_1+3}, \dots, a_{2n_2-1}$, bis die Summe den Wert b überschreitet.
- Addiere negative Elemente $a_{2m_1+2}, a_{2m_1+4}, \dots, a_{2m_2}$, bis die Summe den Wert b unterschreitet.
- usw.

Die einzelnen Summanden a_i bezeichnen wir in der Reihenfolge ihres Auftretens mit b_1, b_2, b_3, \dots . Man kann sich überlegen, dass die einzelnen Schritte terminieren (d.h. die „Abbruchbedingung“ jeweils nach endlich vielen Additionen erreicht wird) und dass die entstehende Reihe $\sum_{n=1}^{\infty} b_n$ in der Tat eine Umordnung von $\sum_{n=1}^{\infty} a_n$ ist und den Wert b besitzt. \square

Das Umordnen von Reihen ist also im Allgemeinen nicht ganz unproblematisch. Daher muss man sich bei der Bildung von Summen der Form

$$\sum_{\omega \in \Omega} a_{\omega}$$

(wobei Ω abzählbar und $a_{\omega} \in \mathbb{R}$), wie sie in der diskreten Stochastik bei der Berechnung von Wahrscheinlichkeiten und Erwartungswerten auftreten, ein paar Gedanken machen. Da a priori im Allgemeinen keine Summationsreihenfolge vorgeschrieben ist, wählt man eine beliebige Abzählung

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

(je nachdem, ob Ω endlich oder abzählbar-unendlich ist), wobei jedes Element von Ω genau einmal vorkommt, und setzt

$$\sum_{\omega \in \Omega} a_{\omega} := \sum_{i=1}^n a_{\omega_i} \quad \text{bzw.} \quad \sum_{\omega \in \Omega} a_{\omega} := \lim_{n \rightarrow \infty} \sum_{i=1}^n a_{\omega_i}.$$

Damit diese Festlegung sinnvoll ist, muss man sich aber noch überlegen, dass der entstehende Wert nicht davon abhängt, welche Summationsreihenfolge zugrunde gelegt wird.

Es sind daher hinreichende Bedingungen dafür von Interesse, dass jede Summationsreihenfolge zum selben Ergebnis führt:

Bemerkung A.6 (Umordnen von Summen). Es sei $(a_i)_{i \in I}$ eine Familie reeller Zahlen, wobei $I \subseteq \mathbb{N}$. Unter jeder der folgenden Voraussetzungen existiert die Summe $\sum_{i \in I} a_i$ (wobei allerdings der Wert ∞ auftreten kann) und ist unabhängig von der Summationsreihenfolge (d. h. man kann beliebig umordnen, ohne den Wert der Summe zu verändern):

- (a) I ist eine endliche Menge.
- (b) Es gilt $a_i \geq 0$ für alle $i \in I$.
- (c) Es gilt $\sum_{i \in I} |a_i| < \infty$. (*absolute Konvergenz*).

Insbesondere gilt unter jeder der o. g. Voraussetzungen:

- 1. Ist $(b_i)_{i \in I}$ eine Umordnung der Summanden a_i , so gilt $\sum_{i \in I} a_i = \sum_{i \in I} b_i$.
- 2. Ist $(c_{jk})_{j \in J, k \in K_j}$ eine Anordnung der Summanden a_i in einem „Rechteckschema“, so gilt $\sum_{i \in I} a_i = \sum_{j \in J} \sum_{k \in K_j} c_{jk}$.

Beim Rechnen mit Wahrscheinlichkeiten ist stets Bedingung (b) erfüllt, da Wahrscheinlichkeiten nicht-negativ sind. Bei der Einführung von Erwartungswerten stellt man durch geeignete Zusatzvoraussetzungen sicher, dass Bedingung (c) erfüllt ist und somit keine Probleme auftreten.

A.5 Funktionen

Eine Abbildung mit Werten in \mathbb{R} wird auch als *Funktion* bezeichnet. (Die Bezeichnungswiese ist nicht einheitlich; oft werden die Begriffe *Abbildung* und *Funktion* auch synonym verwendet.)

Definition A.7. Sei $f : M \rightarrow \mathbb{R}$ eine Funktion, wobei $M \subseteq \mathbb{R}$, und sei $a \in \mathbb{R}$. Wir schreiben $\lim_{x \rightarrow a} f(x) = b$, wenn es (mindestens) eine Folge $(x_n)_{n \in \mathbb{N}}$ in M mit $\lim_{n \rightarrow \infty} x_n = a$ gibt und wenn für *jede* solche Folge $\lim_{n \rightarrow \infty} f(x_n) = b$ gilt.

Definition A.8. Sei $f : M \rightarrow \mathbb{R}$ eine Funktion, wobei $M \subseteq \mathbb{R}$. Dann heißt f *stetig* an der Stelle $a \in M$, wenn $\lim_{x \rightarrow a} f(x) = f(a)$ gilt. Ferner heißt f stetig auf M , wenn f an jeder Stelle $a \in M$ stetig ist.

Definition A.9. Sei $f : M \rightarrow \mathbb{R}$ eine Funktion, wobei $M \subseteq \mathbb{R}$. Dann heißt f *differenzierbar* an der Stelle $a \in M$, wenn der Grenzwert des Differenzenquotienten

$$f'(a) := \lim_{\substack{x \rightarrow a \\ x \neq a}} \frac{f(x) - f(a)}{x - a}$$

existiert; $f'(a)$ heißt dann auch *Ableitung* von f an der Stelle a . Ferner heißt f differenzierbar auf M , wenn f an jeder Stelle $a \in M$ differenzierbar ist; die Funktion $f' : M \rightarrow \mathbb{R}$ heißt dann auch *Ableitung* (oder *Ableitungsfunktion*) von f .

Anschaulich beschreibt die Ableitung $f'(a)$ die Steigung der Tangente an den Graphen von f an der Stelle a .

Ableitungsregeln.

Es seien f, g Funktionen und $\alpha, \beta \in \mathbb{R}$ Konstanten. Dann gilt (unter geeigneten Voraussetzungen):

- $(\alpha f + \beta g)'(x) = \alpha f'(x) + \beta g'(x)$ (*Linearität*)
- $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$ (*Produktregel*)
- $(f/g)'(x) = (f'(x)g(x) - f(x)g'(x))/(g(x))^2$ (*Quotientenregel*)

Liegen die Werte von g im Definitionsbereich von f , so gilt außerdem:

- $(f \circ g)'(x) = f'(g(x))g'(x)$ (*Kettenregel*)

Ist g die Umkehrfunktion von f , so gilt schließlich

- $g'(x) = 1/f'(g(x))$ (*Ableitung der Umkehrfunktion*),

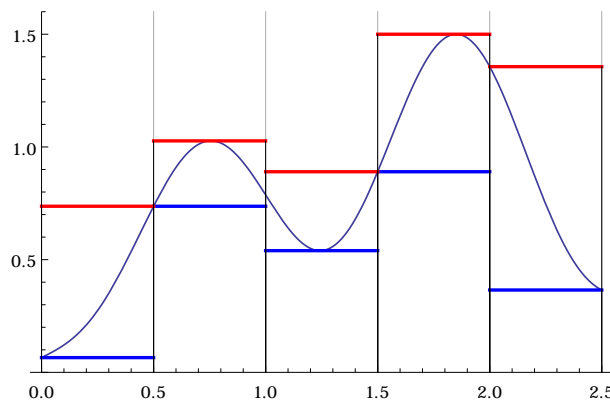
wobei insbesondere $f'(g(x)) \neq 0$ vorauszusetzen ist.

Definition A.10. Sei $I = [a, b]$ ein beschränktes Intervall und $f : [a, b] \rightarrow \mathbb{R}$ eine beschränkte Funktion. Dann heißt f *Riemann-integrierbar* auf $[a, b]$, wenn es zu jedem $\varepsilon > 0$ eine Unterteilung $a = t_0 < t_1 < \dots < t_n = b$ des Intervalls $[a, b]$ sowie Werte $m_1, \dots, m_n, M_1, \dots, M_n \in \mathbb{R}$ gibt, so dass gilt:

(a) $(\forall i = 1, \dots, n) (\forall x \in [t_{i-1}, t_i]) m_i \leq f(x) \leq M_i$.

(b) $\sum_{i=1}^n (M_i - m_i)(t_i - t_{i-1}) < \varepsilon$.

Das *Riemann-Integral* von f auf $[a, b]$, kurz $\int_a^b f(x) dx$, ist dann die eindeutig bestimmte reelle Zahl, die größer gleich allen *Untersummen* der Form $\sum_{i=1}^n m_i(t_i - t_{i-1})$ sowie kleiner gleich allen *Obersummen* der Form $\sum_{i=1}^n M_i(t_i - t_{i-1})$ vom obigen Typ ist.



Veranschaulichung der Konstruktion des Riemann-Integrals
über Untersummen (in Blau) und Obersummen (in Rot)

Ist f nicht-negativ, so bedeutet dies, dass man den Flächeninhalt unter dem Funktionsgraphen von f beliebig genau durch die Flächeninhalte unter dem Funktionsgraphen von *Treppenfunktionen* $\varphi \leq f$ bzw. $\psi \geq f$ approximieren kann. Insbesondere gilt also, dass sich das Riemann-Integral als *Flächeninhalt* unter dem Funktionsgraphen von f interpretieren lässt. Dies ist wesentlich, um das Konzept einer Riemann-Dichte verstehen zu können.

In der Praxis berechnet man das Riemann-Integral i. d. R. nicht über Ober- und Untersummen, sondern mit dem Hauptsatz der Differenzial- und Integralrechnung:

Satz A.11 (Hauptsatz der Differenzial- und Integralrechnung). *Es seien $a, b \in \mathbb{R}$ mit $a < b$ und $f : [a, b] \rightarrow \mathbb{R}$ eine stetige und damit (als stetige Funktion auf einem kompakten Intervall) beschränkte Funktion. Dann ist die Funktion*

$$F : [a, b] \rightarrow \mathbb{R} \quad \text{mit} \quad F(x) = \int_a^x f(t) dt \quad (x \in [a, b])$$

eine Stammfunktion von f (d.h. F ist differenzierbar mit $F' = f$), und ist G eine beliebige solche Stammfunktion von f , so gilt

$$\int_a^b f(x) dx = G(x) \Big|_{x=a}^{x=b} = G(b) - G(a).$$

Ist G eine Stammfunktion von f , so teilt man dies häufig in der Form $\int f(x) dx = G(x)$ mit. (Oft wird noch der Zusatz $+ \text{const}$ hinzugefügt, um zum Ausdruck zu bringen, dass eine Stammfunktion nur bis auf eine additive Konstante eindeutig ist.)

Sind die Funktion f unbeschränkt oder der Definitionsbereich von f ein unbeschränktes Intervall, so kann man unter Umständen das *uneigentliche Riemann-Integral* betrachten, welches auf das gerade eingeführte (eigentliche) Riemann-Integral zurückgeführt wird. Für die Details verweisen wir auf die einschlägige Literatur. Es gilt dann z. B.

$$\int_0^\infty \frac{1}{x^2 + \sqrt{x}} dx = \lim_{a \rightarrow 0} \int_a^1 \frac{1}{x^2 + \sqrt{x}} dx + \lim_{b \rightarrow \infty} \int_1^b \frac{1}{x^2 + \sqrt{x}} dx.$$

Integrationsregeln.

Es seien f, g Funktionen und $\alpha, \beta \in \mathbb{R}$ Konstanten. Dann gilt (unter geeigneten Voraussetzungen):

- $\int_a^b (\alpha f + \beta g)'(x) dx = \alpha \int_a^b f'(x) dx + \beta \int_a^b g'(x) dx$ (*Linearität*)
- $\int_a^b f'(x)g(x) dx = f(x)g(x) \Big|_{x=a}^{x=b} - \int_a^b f(x)g'(x) dx$ (*partielle Integration*)
- $\int_a^b f'(g(x))g'(x) dx = \int_{g(a)}^{g(b)} f(y) dy$ (*Substitutionsregel*)

Taylor-Approximation.

Satz A.12 (Taylor-Approximation). Ist $f : (a, b) \rightarrow \mathbb{R}$ n -mal differenzierbar (wobei $n \in \mathbb{N}$ fest) und ist $x \in (a, b)$, so gilt für alle $h \in \mathbb{R}$ mit $x + h \in (a, b)$

$$f(x + h) = \sum_{k=0}^n f^{(k)}(x) \frac{h^k}{k!} + R_{n+1}(x, h),$$

wobei für den Fehlerterm $R_{n+1}(x, h)$

$$\lim_{h \rightarrow 0} \frac{R_{n+1}(x, h)}{h^n} = 0$$

gilt.

Speziell für $n = 1$ und $n = 2$ ergibt sich also für $h \approx 0$

$$f(x + h) \approx f(x) + f'(x)h \quad \text{bzw.} \quad f(x + h) \approx f(x) + f'(x)h + \frac{1}{2}f''(x)h^2,$$

d. h. die Funktion f lässt sich an der Stelle x „gut“ durch eine lineare bzw. quadratische Funktion approximieren. (Beachte, dass dies für $n = 1$ auch direkt aus der Definition von Differenzierbarkeit folgt.) Allgemeiner lässt sich eine n -mal differenzierbare Funktion „gut“ durch eine Polynomfunktion n -ten Grades approximieren.

Bemerkung. Ist f sogar $(n + 1)$ -mal stetig differenzierbar, so besitzt der Fehlerterm $R_{n+1}(x, h)$ in Satz A.12 auch die Darstellungen

$$R_{n+1}(x, h) = \frac{h^{n+1}}{n!} \int_0^1 f^{(n+1)}(x + th)(1 - t)^n dt$$

(Integral-Darstellung) und

$$R_{n+1}(x, h) = f^{(n+1)}(\xi) \frac{h^{n+1}}{(n + 1)!}$$

(Lagrange-Darstellung) für geeignetes (!) $\xi = \xi(x, h)$ zwischen x und $x + h$. Mit Hilfe dieser Darstellungen lassen sich obere Schranken für den Approximationsfehler angeben.

Beispiel. Für die Funktionen $\exp(x)$ und $\log(1 + x)$ (vgl. dazu den folgenden Abschnitt) ergibt sich für $x = 0$, $n = 1$ und $h \approx 0$

$$\exp(h) \approx 1 + h \quad \text{bzw.} \quad \log(1 + h) \approx h,$$

wobei der Approximationsfehler für $|h| \leq c$ (wobei $c \in (0, 1)$ fest) beispielsweise durch

$$\frac{h^2}{2} \max_{|h| \leq c} |\exp''(h)| = \frac{h^2}{2} e^c \quad \text{bzw.} \quad \frac{h^2}{2} \max_{|h| \leq c} |\log''(1 + h)| = \frac{h^2}{2} \frac{1}{(1 - c)^2}$$

abgeschätzt werden kann.

A.6 Exponentialfunktion und Logarithmusfunktion

Die *natürliche Exponentialfunktion* \exp ist die (eindeutige) differenzierbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit $f(0) = 1$ und $f'(x) = f(x)$ für alle $x \in \mathbb{R}$. Sie wird zur Beschreibung *exponentiellen Wachstums* verwendet.

Die Exponentialfunktion besitzt die *Funktionalgleichung*

$$\exp(x + y) = \exp(x) \exp(y) \quad \text{für alle } x, y \in \mathbb{R}.$$

Sie kann mit Hilfe der *Exponentialreihe*

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (x \in \mathbb{R})$$

oder mit Hilfe des Grenzwerts

$$\exp(x) = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \quad (x \in \mathbb{R})$$

beschrieben werden. (Beide Darstellungen spielen in der Stochastik eine Rolle.) Allgemeiner gilt sogar

$$\lim_{n \rightarrow \infty} x_n = x \implies \lim_{n \rightarrow \infty} \left(1 + \frac{x_n}{n}\right)^n = \exp(x).$$

Setzt man

$$e := \exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!} = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2.718 \dots \quad (\text{Euler'sche Zahl}),$$

so gilt $\exp(x) = e^x$ für alle $x \in \mathbb{R}$. In dieser Notation ergibt sich die Funktionalgleichung für die Exponentialfunktion (formal) aus den Rechenregeln für Potenzen.

Die Exponentialfunktion lässt sich ins Komplexe fortsetzen, indem man für $z = x + \mathbf{i}y$ $\in \mathbb{C}$ (wobei $x, y \in \mathbb{R}$)

$$\exp(z) := \exp(x) (\cos(y) + \mathbf{i} \sin(y))$$

setzt. Die oben angegebene Funktionalgleichung gilt dann sogar für alle $z_1, z_2 \in \mathbb{C}$.

Die reelle Exponentialfunktion $\exp : \mathbb{R} \rightarrow \mathbb{R}$ ist stetig und streng monoton wachsend mit $\lim_{x \rightarrow -\infty} e^x = 0$ und $\lim_{x \rightarrow +\infty} e^x = \infty$ und bildet daher die Menge \mathbb{R} bijektiv auf das Intervall $(0, \infty)$ ab. Die zugehörige Umkehrfunktion $\log : (0, \infty) \rightarrow \mathbb{R}$ heißt *natürliche Logarithmusfunktion*.

Nach der Regel von der Ableitung der Umkehrfunktion gilt

$$\log'(x) = \frac{1}{\exp'(\log x)} = \frac{1}{\exp(\log x)} = \frac{1}{x} \quad (x > 0).$$

Die Logarithmusfunktion ist also eine Stammfunktion der Funktion $f : (0, \infty) \rightarrow (0, \infty)$, $f(x) := 1/x$.

In der Stochastik tritt die Logarithmusfunktion vor allem im Zusammenhang mit der Maximum-Likelihood-Methode auf. Hier ist es unerlässlich, die Rechenregeln für den Logarithmus sicher zu beherrschen: Für alle $x, y, x_1, \dots, x_n > 0$ und alle $r \in \mathbb{R}$ gilt

$$\log(xy) = \log x + \log y, \quad \log \prod_{j=1}^n x_j = \sum_{j=1}^n \log x_j, \quad \log x^r = r \log x.$$

Des Weiteren gilt für jede differenzierbare Funktion f mit Werten in $(0, \infty)$ (nach der Kettenregel)

$$(\log f)'(x) = \frac{f'(x)}{f(x)}.$$

B Tabellen

Tabelle 1: Wichtige diskrete Wahrscheinlichkeitsverteilungen auf \mathbb{R}

Verteilung	Parameter	Zähldichte	E.-wert	Varianz
Bernoulli-Verteilung $\mathcal{B}(1, p)$	$p \in [0, 1]$	$f(k) = p^k(1-p)^{1-k}$ wobei $k \in \{0, 1\}$	p	$p(1-p)$
Binomialverteilung $\mathcal{B}(n, p)$	$n \in \mathbb{N}, p \in [0, 1]$	$f(k) = \binom{n}{k} p^k (1-p)^{n-k}$ wobei $k \in \{0, \dots, n\}$	np	$np(1-p)$
Poisson-Verteilung $\mathcal{P}(\lambda)$	$\lambda > 0$	$f(k) = e^{-\lambda} \lambda^k / k!$ wobei $k \in \mathbb{N}_0$	λ	λ
geometrische Verteilung $\mathcal{G}(p)$	$p \in]0, 1]$	$f(k) = p(1-p)^{k-1}$ wobei $k \in \mathbb{N}$	$1/p$	$(1-p)/p^2$
hypergeometrische Verteilung $\mathcal{H}(r, s, n)$	$r, s \in \mathbb{N}$ $0 \leq n \leq r+s$	$f(k) = \binom{r}{k} \binom{s}{n-k} / \binom{r+s}{n}$ wobei $k \in \{0, \dots, n\}$	$\frac{nr}{r+s}$	\dots
(diskrete) Gleichverteilung $\mathcal{U}(\{1, \dots, n\})$	$n \in \mathbb{N}$	$f(k) = 1/n$ wobei $k \in \{1, \dots, n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$

Tabelle 2: Wichtige stetige Wahrscheinlichkeitsverteilungen auf \mathbb{R}

Verteilung	Parameter	Riemann-Dichte	E.-wert	Varianz
Standardnormalverteilung $\mathcal{N}(0, 1)$	$(\mu = 0, \sigma^2 = 1)$	$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	0	1
Normalverteilung $\mathcal{N}(\mu, \sigma^2)$	$\mu \in \mathbb{R}, \sigma^2 > 0$	$\varphi_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$	μ	σ^2
Exponentialverteilung $\mathcal{E}(\lambda)$	$\lambda > 0$	$f(x) = \lambda e^{-\lambda x} \cdot \mathbf{1}_{(0, \infty)}(x)$	$1/\lambda$	$1/\lambda^2$
(stetige) Gleichverteilung $\mathcal{U}(a, b)$	$a, b \in \mathbb{R}, a < b$	$f(x) = \frac{1}{b-a} \cdot \mathbf{1}_{(a, b)}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

Tabelle 3: Formelsammlung für diskrete und stetige W.verteilungen auf \mathbb{R}^d

Im Folgenden sei $X : (\Omega, \mathcal{B}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathbb{B}^d)$ eine Zufallsgröße mit einer diskreten Verteilung mit einer Zähldichte f (linke Seite) bzw. mit einer stetigen Verteilung mit einer Riemann-Dichte f (rechte Seite).

	diskrete Verteilung	stetige Verteilung
Dichte	$\exists \mathcal{X} \subseteq \mathbb{R}^d$ abzählbar : $\mathbb{P}(X \in \mathcal{X}) = 1$ $f : \mathcal{X} \rightarrow \mathbb{R}$ $f(x) \geq 0 \quad \forall x \in \mathcal{X}$ $\sum_{x \in \mathcal{X}} f(x) = 1$	$f : \mathbb{R}^d \rightarrow \mathbb{R}$ „integrierbar“ $f(x) \geq 0 \quad \forall x \in \mathbb{R}^d$ $\int_{\mathbb{R}^d} f(x) dx = 1$
Wahrscheinlichkeiten	$\mathbb{P}(X \in B) = \sum_{x \in B \cap \mathcal{X}} f(x) \quad (B \text{ beliebig})$	$\mathbb{P}(X \in B) = \int_B f(x) dx \quad (B \text{ Intervall})$
Verteilungsfunktion ($d = 1$)	$F_X : \mathbb{R} \rightarrow [0, 1]$ $F_X(x) = \mathbb{P}(X \leq x) = \sum_{y \in \mathcal{X} : y \leq x} f(y)$	$F_X : \mathbb{R} \rightarrow [0, 1]$ $F_X(x) = \mathbb{P}(X \leq x) = \int_{]-\infty, x]} f(y) dy$
Erwartungswert ($d = 1$)	$\mathbb{E}(X) = \sum_{x \in \mathcal{X}} x f(x)$ falls $\sum_{x \in \mathcal{X}} x f(x) < \infty$	$\mathbb{E}(X) = \int_{\mathbb{R}} x f(x) dx$ falls $\int_{\mathbb{R}} x f(x) dx < \infty$
Erwartungswert ($d = 1$)	$\mathbb{E}(X^2) = \sum_{x \in \mathcal{X}} x^2 f(x)$ falls $\sum_{x \in \mathcal{X}} x ^2 f(x) < \infty$	$\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 f(x) dx$ falls $\int_{\mathbb{R}} x ^2 f(x) dx < \infty$
Erwartungswert von Funktionen ($d \geq 1$)	$h : \mathbb{R}^d \rightarrow \mathbb{R}$ $\mathbb{E}(h(X)) = \sum_{x \in \mathcal{X}} h(x) f(x)$ falls $\sum_{x \in \mathcal{X}} h(x) f(x) < \infty$	$h : \mathbb{R}^d \rightarrow \mathbb{R}$, so dass $h \cdot f$ „integrierbar“ $\mathbb{E}(h(X)) = \int_{\mathbb{R}^d} h(x) f(x) dx$ falls $\int_{\mathbb{R}^d} h(x) f(x) dx < \infty$
Randverteilungen	(X, Y) habe die Dichte $f(x, y)$ auf $\mathcal{X} \times \mathcal{Y}$, wo \mathcal{X}, \mathcal{Y} abz., $\mathbb{P}(X \in \mathcal{X}) = 1, \mathbb{P}(Y \in \mathcal{Y}) = 1$. Dichte von X : $f_X(x) = \sum_{y \in \mathcal{Y}} f(x, y), x \in \mathcal{X}$ Dichte von Y : $f_Y(y) = \sum_{x \in \mathcal{X}} f(x, y), y \in \mathcal{Y}$	(X, Y) habe die Dichte $f(x, y)$ auf $\mathbb{R} \times \mathbb{R}$. Dichte von X : $f_X(x) = \int_{\mathbb{R}} f(x, y) dy$ Dichte von Y : $f_Y(y) = \int_{\mathbb{R}} f(x, y) dx$
Unabhängigkeit	X, Y sind unabhängig, wenn für alle Mengen $B \subseteq \mathcal{X}, C \subseteq \mathcal{Y}$ gilt: $\mathbb{P}(X \in B, Y \in C) = \mathbb{P}(X \in B)\mathbb{P}(Y \in C)$. Sind X und Y unabhängig mit den Dichten $f_X(x)$ und $f_Y(y)$, so hat (X, Y) die Dichte $f_X(x)f_Y(y)$. Hat umgekehrt (X, Y) eine Dichte der Form $f_X(x)f_Y(y)$, so sind X und Y unabhängig.	X, Y sind unabhängig, wenn für alle Intervalle $B, C \subseteq \mathbb{R}$ gilt: $\mathbb{P}(X \in B, Y \in C) = \mathbb{P}(X \in B)\mathbb{P}(Y \in C)$. Sind X und Y unabhängig mit den Dichten $f_X(x)$ und $f_Y(y)$, so hat (X, Y) die Dichte $f_X(x)f_Y(y)$. Hat umgekehrt (X, Y) eine Dichte der Form $f_X(x)f_Y(y)$, so sind X und Y unabhängig.

Tabelle 4: Verteilungsfunktion der Standardnormalverteilung

Die folgende Tabelle enthält die Werte $\Phi(x)$ der Verteilungsfunktion der Standardnormalverteilung für $x \geq 0$.

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Tabelle 5: Quantile der t -Verteilungen sowie der Standardnormalverteilung

Die folgende Tabelle enthält die Quantile $t_{n,u}$ der t -Verteilungen und die Quantile z_u der Standardnormalverteilung (letzte Zeile).

n	0.8	0.9	0.95	0.975	0.98	0.99	0.995
1	1.3764	3.0777	6.3138	12.7062	15.8945	31.8205	63.6567
2	1.0607	1.8856	2.9200	4.3027	4.8487	6.9646	9.9248
3	0.9785	1.6377	2.3534	3.1824	3.4819	4.5407	5.8409
4	0.9410	1.5332	2.1318	2.7764	2.9985	3.7469	4.6041
5	0.9195	1.4759	2.0150	2.5706	2.7565	3.3649	4.0321
6	0.9057	1.4398	1.9432	2.4469	2.6122	3.1427	3.7074
7	0.8960	1.4149	1.8946	2.3646	2.5168	2.9980	3.4995
8	0.8889	1.3968	1.8595	2.3060	2.4490	2.8965	3.3554
9	0.8834	1.3830	1.8331	2.2622	2.3984	2.8214	3.2498
10	0.8791	1.3722	1.8125	2.2281	2.3593	2.7638	3.1693
11	0.8755	1.3634	1.7959	2.2010	2.3281	2.7181	3.1058
12	0.8726	1.3562	1.7823	2.1788	2.3027	2.6810	3.0545
13	0.8702	1.3502	1.7709	2.1604	2.2816	2.6503	3.0123
14	0.8681	1.3450	1.7613	2.1448	2.2638	2.6245	2.9768
15	0.8662	1.3406	1.7531	2.1314	2.2485	2.6025	2.9467
16	0.8647	1.3368	1.7459	2.1199	2.2354	2.5835	2.9208
17	0.8633	1.3334	1.7396	2.1098	2.2238	2.5669	2.8982
18	0.8620	1.3304	1.7341	2.1009	2.2137	2.5524	2.8784
19	0.8610	1.3277	1.7291	2.0930	2.2047	2.5395	2.8609
20	0.8600	1.3253	1.7247	2.0860	2.1967	2.5280	2.8453
21	0.8591	1.3232	1.7207	2.0796	2.1894	2.5176	2.8314
22	0.8583	1.3212	1.7171	2.0739	2.1829	2.5083	2.8188
23	0.8575	1.3195	1.7139	2.0687	2.1770	2.4999	2.8073
24	0.8569	1.3178	1.7109	2.0639	2.1715	2.4922	2.7969
25	0.8562	1.3163	1.7081	2.0595	2.1666	2.4851	2.7874
30	0.8538	1.3104	1.6973	2.0423	2.1470	2.4573	2.7500
35	0.8520	1.3062	1.6896	2.0301	2.1332	2.4377	2.7238
40	0.8507	1.3031	1.6839	2.0211	2.1229	2.4233	2.7045
45	0.8497	1.3006	1.6794	2.0141	2.1150	2.4121	2.6896
50	0.8489	1.2987	1.6759	2.0086	2.1087	2.4033	2.6778
60	0.8477	1.2958	1.6706	2.0003	2.0994	2.3901	2.6603
70	0.8468	1.2938	1.6669	1.9944	2.0927	2.3808	2.6479
80	0.8461	1.2922	1.6641	1.9901	2.0878	2.3739	2.6387
90	0.8456	1.2910	1.6620	1.9867	2.0839	2.3685	2.6316
100	0.8452	1.2901	1.6602	1.9840	2.0809	2.3642	2.6259
∞	0.8416	1.2816	1.6449	1.9600	2.0537	2.3263	2.5758

C Symbolverzeichnis

Allgemeines

$ A $	Mächtigkeit von A
$\text{vol}(A)$	d -dimensionales Volumen von A (für $A \subseteq \mathbb{R}^d$)
$\mathbf{1}_A$	Indikatorfunktion von A
$H_n(A)$	absolute Häufigkeit von A
$h_n(A)$	relative Häufigkeit von A
$\mathbb{P}(A)$	Wahrscheinlichkeit von A
$\mathbb{P}(A B)$	bedingte Wahrscheinlichkeit von A gegeben B
$\mathbb{E}(X)$	Erwartungswert von X
$\text{Var}(X)$	Varianz von X
$\text{Cov}(X, Y)$	Kovarianz von X und Y
$\rho(X, Y)$	Korrelationskoeffizient von X und Y

Diskrete W.räume

Ω	abzählbare Menge
$\mathfrak{P}(\Omega)$	Potenzmenge (= σ -Algebra über Ω)
\mathbb{P}	diskretes W.maß
f	Zähldichte

Stetige W.räume

Ω	Teilmenge von \mathbb{R}^d
$\mathbb{B}^d _{\Omega}$	Borel'sche σ -Algebra über Ω
\mathbb{P}	stetiges W.maß
f	Riemann-Dichte

Allgemeine W.räume

Ω	beliebige Menge
\mathcal{A}	σ -Algebra über Ω
\mathbb{P}	W.maß
F	Verteilungsfunktion von \mathbb{P} (falls $\Omega \subseteq \mathbb{R}$)

Verteilungen

siehe Tabelle 1 und 2

φ	Dichte der Standardnormalverteilung
Φ	Verteilungsfunktion der Standardnormalverteilung

Zufallsgrößen

$X : \Omega \rightarrow \mathcal{X}$	Zufallsgröße (ohne Angabe der σ -Algebren)
$X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$	Zufallsgröße (mit Angabe der σ -Algebren)
\mathbb{P}_X	induzierte Verteilung
F_X	Verteilungsfunktion von X bzw. von \mathbb{P}_X (falls $\mathcal{X} \subseteq \mathbb{R}$)
f_X	Zähldichte von X bzw. von \mathbb{P}_X (falls \mathbb{P}_X diskrete Verteilung)
f_X	Riemann-Dichte von X bzw. von \mathbb{P}_X (falls \mathbb{P}_X stetige Verteilung)
$X(A) := \{X(\omega) : \omega \in A\}$	Bild der Menge $A \subseteq \Omega$
$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}$	Urbild der Menge $B \subseteq \mathcal{X}$
$\{X \in B\}$	alternative Schreibweise für $X^{-1}(B)$
$\{X = x\}$	alternative Schreibweise für $X^{-1}(\{x\})$
$\{a \leq X \leq b\}$	alternative Schreibweise für $X^{-1}([a, b])$
$X \sim P$	X hat Verteilung P
$X \sim f$	X hat Dichte f

Statistik

\mathcal{X}	Stichprobenraum
\mathcal{B}	σ -Algebra über \mathcal{X}
\mathbb{P}_ϑ	Wahrscheinlichkeitsmaß zum Parameter ϑ
\mathbb{E}_ϑ	Erwartungswertoperator zum Parameter ϑ
f_ϑ	Dichte zum Parameter ϑ (falls existent)
Θ	Parameterraum
X	i. d. R. identische Abbildung
X_1, \dots, X_n	i. d. R. Komponentenfunktionen von X
γ	allgemeiner Schätzer
g	zu schätzende Funktion
\hat{g}	Schätzer für g
L_x	Likelihood-Funktion
ℓ_x	Loglikelihood-Funktion
\hat{g}_{ML}	Maximum-Likelihood-Schätzer für g
φ	allgemeiner Test
β_φ	Gütefunktion von φ
H	Nullhypothese
K	Alternative
\bar{X}_n	Stichprobenmittel
S_n^2	Stichprobenvarianz

Errata

Damit Sie das Skript nicht immer wieder vollständig neu ausdrucken müssen, sind hier einige nachträgliche Änderungen aufgelistet. (Rein sprachliche Änderungen sind i. d. R. nicht aufgeführt.)

- Seite 22 unten: $\{X \in x\} \cap \{Y = y\} \rightarrow \{X = x\} \cap \{Y = y\}$
- Seite 47, Satz 2.25: Beweis geändert.
- Seite 52, Beispiel 3.10: Tippfehler „Sensitivität“ und „Spezifität“ korrigiert.
- Seite 61, Satz 3.21: $[h_i : (X_i, \mathcal{B}_i) \rightarrow (\mathcal{Y}_i, \mathcal{C}_i)] \rightarrow [h_i : (\mathcal{X}_i, \mathcal{B}_i) \rightarrow (\mathcal{Y}_i, \mathcal{C}_i)]$
- Seiten 44f, Beispiel 2.22 (b): Rollen von a und b vertauscht. (*reine Geschmackssache!*)
- Seite 63, Beweis von Satz 3.23: U_n ist $\mathcal{U}_{0,1}$ -verteilt. $\rightarrow U_n$ ist $\mathcal{U}_{(0,1)}$ -verteilt.
- Seiten 67 – 83, Kapitel 4: diverse Änderungen, Umstellungen und Ergänzungen; insbesondere ist ein neues Kapitel 4.1 eingefügt worden, in dem die wichtigsten Formeln zur Berechnung von Erwartungswerten zusammengestellt sind.
- Seite 80 unten: Dieser Satz ist in der Vorlesung übersprungen worden. \rightarrow Dieser Satz ist in der Vorlesung mit Ausnahme von Teil (b) übersprungen worden.
- Seite 81: Bemerkung zur Berechnung der Kovarianz eingefügt.
- Seite 91: Graphik überarbeitet.
- Seite 108: man erhält ein Konfidenzintervall \rightarrow man erhält einen Konfidenzbereich (Änderung nur auf den Folien: Nr. 237, Nr. 240, Nr. 241)
- Seite 112, Beispiel 6.20: zusätzliche Details eingefügt. (Änderung nur auf den Folien: Nr. 246)
-
-
-