

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:* October 3, 2019

²*File:* stochinf-001.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	56
4	Erwartungswert und Varianz	68
4.1	Der Erwartungswert von diskreten Zufallsgrößen	68
4.2	Der Erwartungswert von allgemeinen Zufallsgrößen	74
4.3	Varianz und Kovarianz	77
5	Grenzwertsätze für Summen von u.i.v. Zufallsgrößen	83
5.1	Das Gesetz der großen Zahlen	83
5.2	Der zentrale Grenzwertsatz	87
6	Statistik	95
6.1	Einführende Beispiele	95
6.2	Grundbegriffe der Statistik	99
6.3	Punktschätztheorie	101
6.4	Bereichsschätztheorie	108
6.5	Testtheorie	114
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

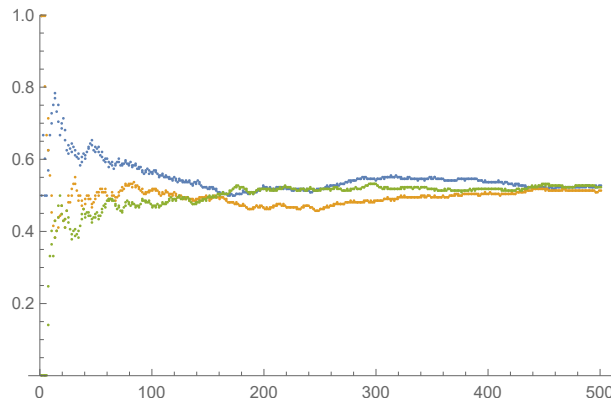
Eigenschaften (idealer) zufallsabhängiger Vorgänge.

- 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.

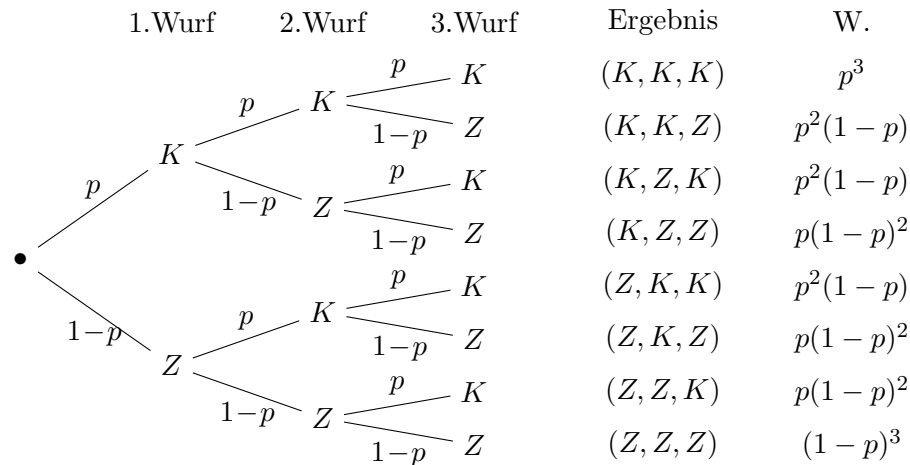
Von diesen np_{ω_1} Versuchen liefern etwa $np_{\omega_1} p_{\omega_2}$ Versuche ω_2 im 2. Wurf.

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:

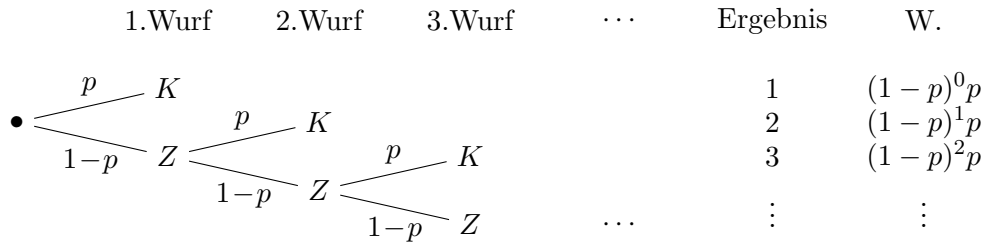
- $\Omega = \mathbb{N} = \{1, 2, 3, \dots\}$, wobei $\omega \triangleq$ Anzahl der Versuche
- $f(\omega) = (1-p)^{\omega-1}p$, $\omega \in \Omega$

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:

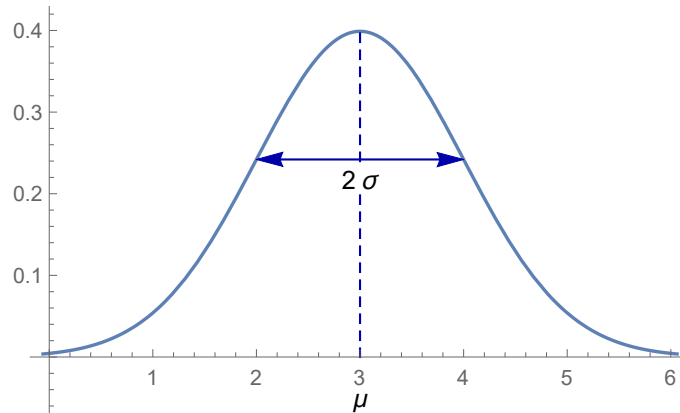
$$\sum_{\omega \in \Omega} f(\omega) = \sum_{n=1}^{\infty} p(1-p)^{n-1} = p \sum_{n=1}^{\infty} (1-p)^{n-1} \underset{\text{geom. Reihe}}{=} p \cdot \frac{1}{1-(1-p)} = 1.$$

Beachte, dass wir sind in allen Beispielen *systematisch* vorgegangen, indem wir erst eine mathematische Beschreibung der Situation angeben 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

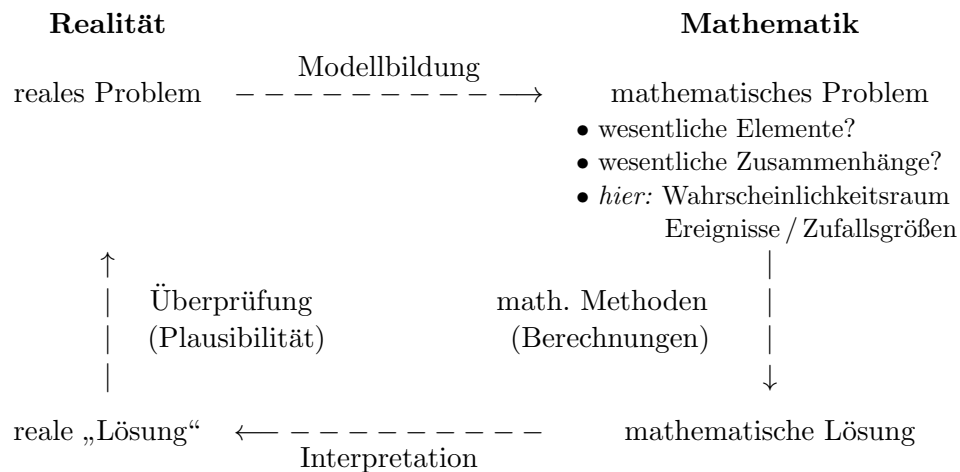
- persönliche Einschätzungen / Überzeugungen

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.

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)

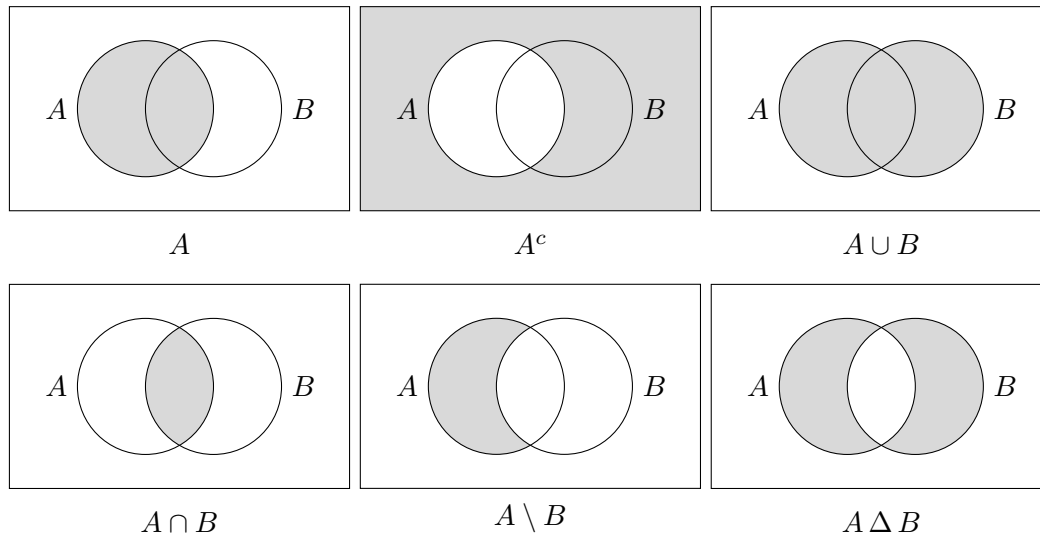
Je nachdem, ob die eckigen Klammern nach innen oder nach außen zeigen, gehören die „Intervallgrenzen“ also zum Intervall dazu oder nicht. Natürlich gibt es weitere Kombinationen, die nun allerdings selbsterklärend sein sollten. Anstelle nach außen geöffnete eckiger Klammern verwendet man (gleichwertig) 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*),

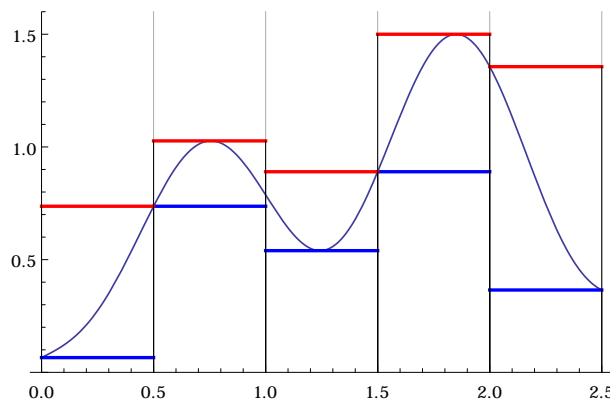
wobei insbesondere $f'(g(x)) \neq 0$ vorauszusetzen ist.

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.)

-
-
-