Wahrscheinlichkeit und Statistik

Autor: V. Tassion
Übersetzung: L. Brummet[1]

D-INFK Frühling 2022 (Aktualisiert: 1. Juli 2022)

[1]Für Korrekturen gerade Meldung an: Luis.Brummet@math.ethz.ch

Einführung in die Wahrscheinlichkeitstheorie

Ein paar Einstiegsfragen

Was ist überhaupt Wahrscheinlichkeitstheorie?

➜ Eine mathematische Sprache zur Beschreibung von Systemen, die Zufälligkeiten beinhalten.

Warum ist Wahrscheinlichkeitstheorie nützlich?

➜ **Beschreiben zufälliger Experimente**. Viele Vorgänge der reellen Welt können per Zufallsexperiment beschrieben werden. Beispiele sind unter anderem der Münzwurf, Würfeln, Ankunftszeiten von Kunden, Wettervorhersage,...).

➜ **Quantifizierung von Unsicherheit**. Maschinelle Messdaten sind selten exakt. Man kann mittels Wahrscheinlichkeitstheorie und Statistik die Unsicherheit beschreiben und quantifizieren.

➜ **Unterstützung per Risikotheorie** Wahrscheinlichkeitstheorie hilft unter Informationsunsicherheit eine rationale Entscheidung zu treffen. Zum Beispiel hilft der SLF Lawinenbericht dem Skifahrer die Lawinengefahr richtig einzuschätzen.

➜ **Nutzen zufallsbasierter Algorithmen**. In der Informatik kann das Einbauen von Zufall in Algorithmen zu einer Verbesserung der Laufzeiteffizenz führen. Beispiele sind Google websearch, Ameisen bei der Nahrungssuche.

➜ **Vereinfachung komplexer Systeme**. Beispiele: Wassermoleküle in Wasser, Autos auf der Autobahn, Perkolationsprozesse.

Der Begriff des Grundraums

Für eine präzise physikalische Beschreibung eines Münzwurfs bräuchte man eine riesige Menge an Informationen: die genaue Position der Finger und der Münzen, die Anfangsgeschwindigkeit, die anfängliche Winkelgeschwindigkeit, Unvollkommenheiten der Münze, die Oberflächenbeschaffenheit des Tisches, Luftströmungen, die Gehirnaktivität des Spielers... Diese Parameter sind fast unmöglich genau zu messen, und eine winzige Änderung eines dieser Parameter kann das Ergebnis vollständig beeinflussen. In der Praxis verwenden wir eher eine wahrscheinlichkeitstheoretische Beschreibung, die hier in einer drastischen Vereinfachung des Systems besteht: Wir vergessen die physikalische Beschreibung des Münzwurfs völlig und konzentrieren uns nur auf den **Grundraum** also die **möglichen Ereignisse** des Experiments: Kopf oder Zahl.

Der Grundraum für den Münzwurf ist durch 2 mögliche Ereignisse (Kopf und Zahl) gegeben. Jedes Ergebnis hat eine Wahrscheinlichkeit von $p_{\text{Kopf}} = p_{Zahl} = 1/2$, realisiert zu werden. Mit anderen Worten,

$$\textbf{Münzwurf} = \{\{\text{Kopf}, \text{Zahl}\}, p_{\text{Kopf}} = 1/2, p_{\text{Zahl}} = 1/2\}$$

Überraschender Funfact: Münzwürfe sind gezinkt! Wenn man eine Münze wirft, ist die Wahrscheinlichkeit größer, dass sie auf die gleiche Seite fällt wie ihre ursprüngliche Seite! Dem interessierten Leser ist das Youtube-Video von Persi Diaconis ans Herz zu legen: How random is a coin toss? - Numberphile

Wahrscheinlichkeitsgesetzte: Struktur im Zufall

Wenn man ein einzelnes Zufallsexperiment durchführt (z.B. einmaliger Münzwurf) ist das Ergebnis nicht vorhersehbar. Führt man dagegen viele Zufallsexperimente hintereinander durch, so lassen sich einige allgemeine Gesetze beobachtet. Wenn man zum Beispiel 10000 unabhängige Münzen wirft, sollte man im Allgemeinen ungefähr 5000 Kopf und 5000 Zahl beobachten: Dies ist ein Beispiel für ein grundlegendes Wahrscheinlichkeitsgesetz, dem so genannte Gesetz der großen Zahlen. Ein Ziel der Wahrscheinlichkeitstheorie ist das Beschreiben von Wahrscheinlichkeitsgesetzen. Diese geben allgemeine Hinweise darauf wie sich aus wiederholtem Anwenden vieler Zufallsexperimenten eine allgemeingültige Struktur ergeben kann.

Kapitel 1

Mathematische Grundkenntnisse

Ziele

- Grundsätzliches Verständnis des Wahrscheinlichkeitraums $(\Omega, \mathcal{F}, \mathbb{P})$,
  - ➜ Verallgemeinerung diskreter Wahrscheinlichkeitsräume (eingeführt in[LSW21]),
  - ➜ Verständnis des Begriffes der σ-Algebra.
- Die Konzepte der Unabhängigkeit und der bedingten Wahrscheinlichkeit.

1 Wahrscheinlichkeitsräume

Grundraum

Wir wollen ein Zufallsexperiment modellieren. Das erste benötigte mathematische Objekt ist die Menge aller möglichen Ergebnisse des Experiments, bezeichnet mit Ω.

Terminologie: Die Menge Ω nennen wir **Grundraum**. Ein Element $\omega \in \Omega$ nennen wir **Elementarereignis** (oder **Ausgang des Experiments**).

Beispiel. *Einmaliges Würfeln*
Ein einfaches Beispiel eines Grundraums ist das einmalige Würfeln. Die Menge aller elementar Ereignisse ist gerade

$$\Omega = \{1,2,3,4,5,6\}.$$

Ereignisse

Reminder: Die Menge $\mathcal{P}(\Omega) = \{A \mid A \subseteq \Omega\}$ enthält alle möglichen Teilmengen A der Menge Ω.

In der vorherigen Vorlesung [LSW21], war die Menge alle möglicher Ereignisse stets $\mathcal{P}(\Omega)$. Wir hingegen werden mit einem etwas allgemeineren System von Mengen, der sogenannten Sigma-Algebra (geschrieben σ-Algebra) arbeiten. Dabei schreiben wir stets $\mathcal{F} \subset \mathcal{P}(\Omega)$ für die gewählte σ-Algebra.

Definition 1.1. *Ein Mengensystem $\mathcal{F} \subset \mathcal{P}(\Omega)$ heisst* ***Sigma-Algebra*** *(σ-Algebra), falls es die folgenden Eigenschaften erfüllt*

E1. $\Omega \in \mathcal{F}$

E2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ — *Ist A ein Ereignis, so ist "nicht A" ebenfalls ein Ereignis.*

E3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$. — *Ist $A_1, A_2, \ldots$ ein Ereignis, dann ist "A_1 oder A_2 oder ..." ebenfalls ein Ereignis*

Dabei nennen wir die Elemente der σ-Algebra Ereignisse.

Beispiele für σ-Algebren beim einmaligen Würfeln ($\Omega = \{1,2,3,4,5,6\}$):

- $\mathcal{F} = \{\varnothing, \{1,2,3,4,5,6\}\}$.
- $\mathcal{F} = \mathcal{P}(\Omega)$. (Wobei $|\mathcal{F}| = 64$).

- $\mathcal{F} = \{\varnothing, \{1,2\}, \{3,4,5,6\}, \{1,2,3,4,5,6\}\}$.

Gegenbeispiele für Sigma-Algebren beim einmaligen Würfeln ($\Omega = \{1,2,3,4,5,6\}$)**:**

- $\mathcal{F} = \{\{1,2,3,4,5,6\}\}$ ist keine σ-Algebra, da **E2** verletzt ist.
- $\mathcal{F} = \{\varnothing, \{1,2,3\}, \{4,5,6\}, \{1\}, \{2,3,4,5,6\}, \Omega\}$ ist keine σ-Algebra , da **E3** verletzt ist.

Wahrscheinlichkeitsmass

Definition 1.2. *Sei Ω ein Grundraum und sei $\mathcal{F}$ eine σ-Algebra. Eine Abbildung*

$$\begin{aligned} \mathbb{P}: \ \mathcal{F} &\to [0,1] \\ A &\mapsto \mathbb{P}[A] \end{aligned}$$

*heisst **Wahrscheinlichkeitsmass** auf $(\Omega, \mathcal{F})$, falls folgende Eigenschaften gelten*

E1. $\mathbb{P}[\Omega] = 1$.

E2. *(σ-**Additivität**)* $\mathbb{P}[A] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$ *if* $A = \bigcup_{i=1}^{\infty} A_i$ *(disjunkte Vereinigung).*

Zusammenfassend: Ein Wahrscheinlichkeitsmass ordnet jedem Ereignis der σ-Algebra eine Zahl zwischen 0 und 1 (Wahrscheinlichkeit) zu.

Beispiel fürs Würfeln $\Omega = \{1,2,3,4,5,6\}$ **und** $\mathcal{F} = \mathcal{P}(\{1,2,3,4,5,6\})$**:**

- Die folgende Abbildung $\mathbb{P} : \mathcal{F} \mapsto [0,1]$, welche definiert ist durch

  $$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \frac{|A|}{6}$$

  ist ein Wahrscheinlichkeitsmass $(\Omega, \mathcal{F})$.

- Seien $p_1, \ldots, p_6$ positive Zahlen, sodass $p_1 + \cdots + p_6 = 1$, die folgende Abbildung $\mathbb{P} : \mathcal{F} \mapsto [0,1]$

  $$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \sum_{i \in A} p_i$$

  ist ein Wahrscheinlichkeitsmass auf $(\Omega, \mathcal{F})$. Dabei entspricht der Fall $p_i = \frac{1}{6}$ (für jedes i) dem Modellieren eines fairen Würfels. Der Fall $p_1 = \cdots = p_5 = \frac{1}{7}$, und $p_6 = \frac{2}{7}$ würde einem gezinkten Würfel mit Bias zur 6 entsprechen.

Der Wahrscheinlichkeitsraum

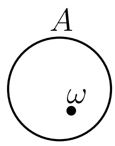
Definition 1.3. *Sei Ω ein Grundraum, $\mathcal{F}$ eine σ-Algebra, und $\mathbb{P}$ ein Wahrscheinlichkeitsmass. Wir nennen das Tripel $(\Omega, \mathcal{F}, \mathbb{P})$* ***Wahrscheinlichkeitsraum****.*

Zusammenfassend benötigt ein Wahrscheinlichkeitsraum stets folgende drei Objekte

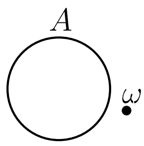
- Ein Grundraum Ω, “Alle möglichen Ausgänge des Zufallexperiments”
- Eine σ-Algebra $\mathcal{F} \subset \mathcal{P}(\Omega)$, “Die Menge aller Ereignisse”
- Ein Wahrscheinlichkeitsmass $\mathbb{P}$. “Ordnet jedem Ereignis eine Wahr. in $[0, 1]$ zu”

Terminologie

Sei A ein Ereignis und sei ω ein beliebiges Elementarereignis.
Wir sagen A **tritt ein** (für ω) falls $\omega \in A$.



Wir sagen A **tritt nicht ein** falls $\omega \notin A$.



Als Beispiel nehme man das Ereignis "Würfeln einer geraden Zahl", also $A = \{2, 4, 6\}$. Falls $\omega = 2$, dann **tritt** A **ein**, wohingegen für $\omega = 3$ A **nicht eintritt**.

Bemerkung 1.4. *Das Ereignis $A = \varnothing$ tritt niemals ein.* *“Da $\omega \in \varnothing$ unmöglich ist!’*
Das Ereignis $A = \Omega$ tritt stets ein. *“Es gilt stets $\omega \in \Omega$”*

2 Beispiele von Wahrscheinlichkeitsräumen

Beispiele mit endlichem Grundraum

Wir erörtern nun eine bestimmte Art von Wahrscheinlichkeitsräumen, die in vielen konkreten Beispielen vorkommen. Im nachfolgenden ist Ω eine beliebige **endliche** Menge. Dabei habe jedes Elementarereignis $\omega \in \Omega$ gerade die **gleiche Wahrscheinlichkeit** $p_\omega = \frac{1}{|\Omega|}$.

Definition 1.5. *Sei Ω eine endlicher Grundraum. Das* ***Laplace Modell*** *auf Ω ist ein Tripel $(\Omega, \mathcal{F}, \mathbb{P})$, sodass*

- *$\mathcal{F} = \mathcal{P}(\Omega)$,*
- *$\mathbb{P} : \mathcal{F} \to [0,1]$ ist definiert durch*

$$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \frac{|A|}{|\Omega|}.$$

Man kann leicht überprüfen, dass die obige Abbildung $\mathbb{P}$ ein Wahrscheinlichkeitsmass im Sinne der Definition 1.2 ist. Dabei läuft die Berechnung der Wahrscheinlichkeit $\mathbb{P}[A]$ darauf hinaus, die Anzahl der Elemente in A und in Ω zu zählen.

Beispiel. *Wir betrachten $n \geq 3$ Punkte auf einem Kreis, von denen wir 2 zufällig auswählen. Wie gross ist die Wahrscheinlichkeit, dass diese beiden ausgewählten Punkte benachbart sind?*

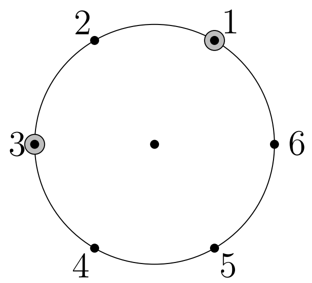


Abbildung 1.1: Kreis mit $n = 6$ Punkten, die Teilmenge $\{1,3\}$ ist ausgewählt.

Wir betrachten dazu das Laplace Modell auf

$$\Omega = \{E \subset \{1,2,\cdots,n\} \ : \ |E| = 2\}.$$

Das Ereignis "zwei Punkte aus E sind Nachbarn" ist gegeben durch

$$A = \{\{1,2\},\{2,3\},\cdots,\{n-1,n\},\{n,1\}\},$$

somit erhalten wir

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|} = \frac{n}{\binom{n}{2}} = \frac{2}{n-1}.$$

Ω als abzählbarer Grundraum

Wir werfen eine gezinkte Münze mehrmals, bei jedem Wurf fällt die Münze mit der Wahrscheinlichkeit p auf Kopf und mit Wahrscheinlichkeit $1-p$ auf Zahl (p ist ein fester Parameter in $[0,1]$). Wir stoppen das Zufallsexperiment, wenn wir das erste Mal Zahl sehen. Die Wahrscheinlichkeit, dass wir genau zum Zeitpunkt k anhalten, ist gegeben durch

$$p_k = p^{k-1}(1-p).$$

(Falls wir im k-ten Wurf stoppen, müssen wir exakt $k-1$-mal Kopf und 1-mal Zahl gesehen haben.)

Dabei können wir folgenden Wahrscheinlichkeitsraum wählen

- $\Omega = \mathbb{N}\backslash\{0\} = \{1, 2, 3, \ldots\}$,
- $\mathcal{F} = \mathcal{P}(\Omega)$,
- for $A \in \mathcal{F}$, $\mathbb{P}[A] = \sum_{k \in A} p_k$.

Ω als überabzählbarer Grundraum (Ein Blick in die Zukunft)

Ein verlockender Ansatz zur Definition eines Wahrscheinlichkeitsmasses besteht darin zunächst jedem Ausgang eines Zufallsexperiments ω die Wahrscheinlichkeit p_ω zuzuordnen. Dann definiere man für ein Ereignis $A \subset \Omega$ die Wahrscheinlichkeit von A durch

$$\mathbb{P}[A] = \sum_{\omega \in A} p_\omega. \tag{1.1}$$

Dieser Ansatz funktioniert einwandfrei, solange der Grundraum Ω endlich oder abzählbar ist (dies ist der Fall bei den beiden obigen Beispielen). Der Ansatz scheitert aber, wenn Ω nicht abzählbar ist.
Man beobachte den Fall eines Regentropfen auf ein 1-Kilometer langes Strassensegment ($\Omega = [0, 1]$). In diesem Fall ist die Wahrscheinlichkeit, einen winzigen festen Punkt zu treffen, immer 0 und die Gleichung (1.1) ergibt keinen Sinn.
Aus diesem Grund verwenden wir eine axiomatische Definition (in Definition 1.2) des Wahrscheinlichkeitsmasses. Dabei wählen wir als die natürliche Wahl des Wahrscheinlichkeitsraums

- $\Omega = [0, 1]$,
- $\mathcal{F}$ = Borel σ-algebra[1]
- for $A \in \mathcal{F}$, $\mathbb{P}(A)$ = Lebesguemass von$A$.

3 Eigenschaften/Interpretationen von Ereignissen

Eigenschaften von Ereignissen

Da Ereignisse als Teilmengen von Ω definiert sind, kann man sich Operationen der Mengenlehre (Vereinigung, Schnittmenge, Komplement, symmetrische Differenz, …) zu Nutze machen.
Aus der Definition der σ-Algebra wissen wir, dass wir das Komplement eines Ereignisses (nach E2) oder eine abzählbare Vereinigung von Ereignissen (nach E3) nehmen können. Der folgende Satz besagt, dass folgende Standardmengenoperationen ebenfalls zulässig sind. Man spricht dabei von der Abgeschlossenheit der σ-Algebra bzgl. Mengenoperationen.[2]

---

[1] die Borel σ-Algebra $\mathcal{F}$ erfüllt folgende Eigenschaften: Sie enthält alle Intervalle $A = [x_1, x_2] \times [y_1, y_2]$, mit $0 \le x_1 \le x_2 \le 1$, $0 \le y_1 \le y_2 \le 1$, und ist dabei die kleinste Vereinungen von Teilmengen von Ω, sodass **E1**, **E2** und **E3** in Definition 1.1 erfüllt ist.

[2] **Achtung:** Dies hat nichts mit Abgeschlossenheit im Sinne der Analysis/Mengentopologie zu tun

Satz 1.6 (Abgeschlossenheit der σ-Algebra bzgl. Operationen)**.** *Sei $\mathcal{F}$ eine σ-Algebra auf Ω. Es gilt*

E4. $\varnothing \in \mathcal{F}$,

E5. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$,

E6. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$,

E7. $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.

Beweis. Wir beweisen alle Aussagen der Reihe nach.

i. Per **E1** in Definition 1.1, gilt $\Omega \in \mathcal{F}$. Per **E3** in Definition 1.1, erhalten wir

$$\varnothing = \Omega^c \in \mathcal{F}.$$

ii. Seien $A_1, A_2, \ldots \in \mathcal{F}$. Mittels **E2**, gilt $A_1^c, A_2^c, \ldots \in \mathcal{F}$. Dann, per **E3**, haben wir $\bigcup_{i=1}^{\infty}(A_i)^c \in \mathcal{F}$. Erneutes Anwenden von **E2** liefert dann das Gewünschte mittels folgender Gleichheit

$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} (A_i)^c \right)^c \in \mathcal{F}.$$

iii. Sei $A, B \in \mathcal{F}$. Definiere $A_1 = A$, $A_2 = B$, und wähle für $i \geq 3$ $A_i = \varnothing$. Mittels **E3**, erhalten wir

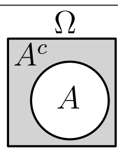
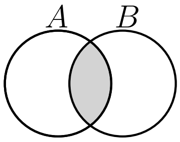
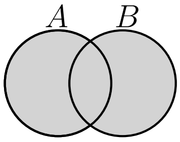
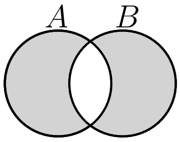
$$A \cup B = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

iv. Sei $A, B \in \mathcal{F}$. Mittels **E2** gilt, $A^c, B^c \in \mathcal{F}$. Dann nutze obrige Eigenschaft *iii.* , so dass $A^c \cup B^c \in \mathcal{F}$. Anwenden von **E2** liefert das Gewünschte

$$A \cap B = (A^c \cup B^c)^c \in \mathcal{F}.$$

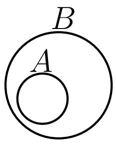
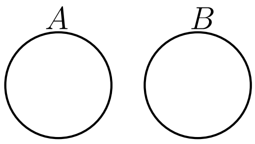
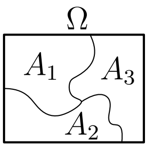
□

In der folgenden Tabelle betrachten wir zwei Ereignisse A und B und fassen die wahrscheinlichkeitstheoretischen Interpretation der wichtigsten Mengenoperation zusammen.

Ereignis	Grafische Darstellung	Probab. Interpretation
A^c		A tritt **nicht** ein
$A \cap B$		A **und** B treten ein
$A \cup B$		A **oder** B treten ein
$A \Delta B$		**Entweder** A oder B tritt ein

Beziehung/Interpretationen zwischen Ereignissen

Mengenrelationen (Inklusion, Unterscheidbarkeit, Partition) haben auch wahrscheinlichkeitstheoretische Interpretationen, die im Folgenden zusammengefasst werden.

Relation	Grafische Darstellung	Probab. Interpretation
$A \subset B$		B tritt ein, falls A eintritt
$A \cap B = \varnothing$		A und B können nicht gleichzeitig eintreten
$\Omega = A_1 \cup A_2 \cup A_3$ wobei A_1, A_2, A_3 paarweise disjunkt		für jedes Elementarereignis ω, höchstens eins der Ereignisse A_1, A_2, A_3 kann eintreten.

Warum wählen wir nicht stets $\mathcal{F} = \mathcal{P}(\Omega)$?

Während der letzten Lesung des Kurses [LSW21] wurde die Menge der Ereignisse immer als $\mathcal{P}(\Omega)$ fixiert. Es mag zunächst nutzlos erscheinen, allgemeinere Ereignismengen auszuwählen und den "komplizierten" Begriff der σ-Algebra zu betrachten. Folgende zwei Beispiele dienen daher als Motivation:

- **Stufenweise Experimente.** Das Arbeiten mit dem abstrakten Konzept der σ-Algebra ermöglicht eine natürliche Zerlegung der Wahrscheinlichkeitsräume. Dies ist besonders nützlich, wenn wir das Ergebnis eines Zufallsexperiments stufenweise aufdecken.
  Wir betrachten den Wurf zwei unabhängiger Würfel. Ein mögliches Ereignis ist ein Paar $\omega = (\omega_1, \omega_2)$, wobei ω_1 und ω_2 die jeweiligen Werte des ersten und zweiten Würfels sind. Wir wählen den Grundraum

  $$\Omega = \{1,2,3,4,5,6\}^2.$$

  Wir können nun **zwei verschiedene** σ-Algebren wählen

  $$\mathcal{F}_1 = \{A \times \{1,2,3,4,5,6\}, A \subset \{1,2,3,4,5,6\}\}$$

  und

  $$\mathcal{F}_2 = \mathcal{P}(\Omega).$$

  Die erste σ-Algebra $\mathcal{F}_1$ entspricht allen Ereignissen, welche **nur** Informationen des **ersten Würfelauges** spezifiziert. Die zweite σ-Algebra $\mathcal{F}_2$ hingegen erlaubt spezifizierte Informationen des ersten und zweiten Würfelauges.
  Zum Beispiel ist das Ereignis $A = \{2,4,6\} \times \{1,2,3,4,5,6\}$ ("der erste Würfel zeigt eine gerade Zahl") sowohl in $\mathcal{F}_1$ als auch in $\mathcal{F}_2$ enthalten, während das Ereignis $B = \{2,4,6\}^2$ ("Beide Würfel zeigen gerade Zahlen") zu $\mathcal{F}_2$, aber nicht zu $\mathcal{F}_1$ gehört, da es Informationen des zweiten Würfels spezifiziert.

  Nehmen wir also an das Ergebnis des Experiments wird stufenweise aufgedeckt. $\mathcal{F}_1$ eignet sich zum Beschreiben von Experimenten in denen nur **vollständiger Information zum ersten Würfel** vorliegt. $\mathcal{F}_2$ hingegeben beschreibt Experimente bezüglich **vollständiger Information beider Würfel**.

- **Ein theoretischer Aspekt.**

  Im Falle eines überabzählbaren Grundraums (z. B. $\Omega = [0,1]$ oder $\Omega = \{0,1\}^{\mathbb{N}}$), werden häufig einige Bedingungen für Ereignisse aufgestellt. Dies ist darauf zurückzuführen, dass wir in der Lage sein wollen, ein Wahrscheinlichkeitsmass für die Menge der Ereignisse zu definieren. Dies ist für $\mathcal{F} = \mathcal{P}(\Omega)$ nicht immer möglich.
  Bei der Definition des Wahrscheinlichkeitsmasses einer Gleichverteilung auf Ω kann man zum Beispiel Mengen $A \subset [0,1]$ konstruieren, die "exotisch genug" sind (vergleiche mit [Wil01], sodass $\mathbb{P}[A]$ nicht wohldefiniert ist. Daher müssen wir uns auf Mengensysteme der Form $\mathcal{F} \subsetneq \mathcal{P}(\Omega)$ beschränken.
  Diese Konstruktion schliesst zu exotische Mengen aus. Für diesen Kurs ist nicht entscheidend dieses theoretische Hindernis im Detail zu verstehen. Wir wollen es jedoch erwähnen, da dies von grundlegender Bedeutung für die Masstheorie, welche die theoretische Grundlage für die Wahrscheinlichkeitstheorie darstellt, ist.

4 Eigenschaften von Wahrscheinlichkeitsmassen

Direkte Konsequenzen aus der Definition

Satz 1.7. *Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmass auf* $(\Omega, \mathcal{F})$.

E3. *Es gilt*
$$\mathbb{P}[\varnothing] = 0.$$

E4. *(**Additivität**) Sei $k \geq 1$, seien $A_1, \ldots, A_k$ k-viele paarweise disjunkte Ereignisse, dann gilt*
$$\mathbb{P}[A_1 \cup \cdots \cup A_k] = \mathbb{P}[A_1] + \cdots + \mathbb{P}[A_k].$$

E5. *Sei A ein Ereignis, dann gilt*
$$\mathbb{P}[A^c] = 1 - \mathbb{P}[A].$$

E6. *Falls A und B zwei (nicht notwendigerweise disjunkte) Ereignisse, dann gilt*
$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

Beweis. Wir beweisen die Eigenschaften der Reihe nach.

E3. Definiere $x = \mathbb{P}[\varnothing]$. Da $\mathbb{P}$ ein Wahrscheinlichkeitsmass ist wissen wir bereits, dass $x \in [0,1]$, da x ein Ereignis ist. Definiere $A_1 = A_2 = \cdots = \varnothing$, dann erhalten wir
$$\varnothing = \bigcup_{i=1}^{\infty} A_i.$$
Die Ereignisse A_i sind disjunkt und Anwenden der σ-Additivität liefert
$$\sum_{i=1}^{\infty} P[A_i] = \mathbb{P}[\varnothing].$$
Da $\mathbb{P}[A_i] = x$ für jedes i und $P[\varnothing] \leq 1$, erhalten wir
$$\sum_{i=1}^{\infty} x \leq 1,$$
und somit muss $x = 0$ gelten.

E4. Definiere $A_{k+1} = A_{k+2} = \cdots = \varnothing$. Dann gilt
$$A_1 \cup \cdots \cup A_k = A_1 \cup \cdots \cup A_k \cup \varnothing \cup \varnothing \cup \cdots = \bigcup_{i=1}^{\infty} A_i.$$

Da die Ereignisse A_i paarweise disjunkt sind, erhalten wir unter Anwendung der σ-Additivität das Gewünschte

$$\begin{aligned}\mathbb{P}[A_1 \cup \cdots \cup A_k] &= \mathbb{P}[\bigcup_{i=1}^{\infty} A_i] \\ &\overset{\substack{\text{Sigma} \\ \text{Additivität}}}{=} \sum_{i=1}^{\infty} \mathbb{P}[A_i] \\ &= \mathbb{P}[A_1] + \cdots + \mathbb{P}[A_k] + \sum_{i>k} \underbrace{\mathbb{P}[A_i]}_{=0}.\end{aligned}$$

E5. Per Definition des Komplements gilt $\Omega = A \cup A^c$, somit erhalten wir

$$1 = \mathbb{P}[\Omega] = \mathbb{P}[A \cup A^c].$$

Da A, A^c disjunkt sind, liefert Anwendung der Addivität das Gewünschte

$$1 = \mathbb{P}[A] + \mathbb{P}[A^c].$$

E6. $A \cup B$ ist die disjunkte Vereinigung aus A und $B \backslash A$. Somit liefert Anwendung der Addivität

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B \backslash A]. \tag{1.2}$$

Zudem gilt $B = (B \cap A) \cup (B \cap A^c) = (B \cap A) \cup (B \backslash A)$. Erneutes Anwenden von Addivität liefert

$$\mathbb{P}[B] = \mathbb{P}[B \cap A] + \mathbb{P}[B \backslash A],$$

somit erhalten wir $\mathbb{P}[B \backslash A] = \mathbb{P}[B] - \mathbb{P}[A \cap B]$. Einsetzen in Eq. (1.2) liefert das Gewünschte.

□

Nützliche Ungleichungen

Satz 1.8 (Monotonie)**.** *Seien $A, B \in \mathcal{F}$, dann gilt*

$$A \subset B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B].$$

Beweis. Sei $A \subset B$, dann gilt $B = A \cup (B \backslash A)$ (disjunkte Vereinigung). Anwendung der Additivität liefert das Gewünschte,

$$\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[B \backslash A] \geq \mathbb{P}[A].$$

□

Satz 1.9 (Union Bound)**.** *Sei $A_1, A_2, \ldots$ eine Folge von (nicht notwendigerweise disjunkten) Ereignissen, dann gilt die folgende Ungleichung*

$$\mathbb{P}[\bigcup_{i=1}^{\infty} A_i] \leq \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

Bemerkung 1.10. *Die obrige Ungleichung gilt ebenfalls für eine Folge von endlich vielen nicht-leeren Ereignissen*

Beweis. Für $i \geq 1$, definiere

$$\widetilde{A}_i = A_i \cap A_{i-1}^c \cap \cdots \cap A_1^c.$$

Man überprüfe folgende Gleichheit

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \widetilde{A}_i.$$

(" $\subseteq$ " : Für die erste Inklusion, nehme ein ω in der linken Menge. Definiere das Kleinste i sodass $\omega \in A_i$. Für i haben wir $\omega \in \widetilde{A}_i$, somit muss ω ebenfalls in der rechten Menge liegen. " $\supseteq$ " : Die zweite Inklusion ist klar, da $\widetilde{A}_i \subset A_i$ für jedes i gilt.)
Anwendung der σ-Additivität auf disjunkte $\widetilde{A}_i$, liefert das Gewünschte

$$\begin{aligned}\mathbb{P}[\bigcup_{i=1}^{\infty} A_i] &= \mathbb{P}[\bigcup_{i=1}^{\infty} \widetilde{A}_i] \\ &= \sum_{i=1}^{\infty} \mathbb{P}[\widetilde{A}_i] \\ &\leq \sum_{i=1}^{\infty} \mathbb{P}[A_i].\end{aligned}$$

□

Anwendung der Ungleichungen

In Anwendungen kommt es oft vor, dass die Wahrscheinlichkeit $\mathbb{P}[A]$ schwer exakt zu berechnen ist: In solchen Fällen ist es oft nützlich, das Ereignis A in anderen einfacheren Ereignissen auszudrücken und dann die Monotonie nach (Proposition 1.8) und/oder die Sub-Additivität anzuwenden, um einige Abschätzungen für $\mathbb{P}[A]$ in Form von einfacheren Wahrscheinlichkeiten zu erhalten.

Problem. Wir würfeln einen fairen Würfeln n mal. Wir wählen als Wahrscheinlichkeitsraum

- $\Omega = \{1, 2, 3, 4, 5, 6\}^n$, Ein Elementarevent ist $\omega = (\underbrace{\omega_1}_{\text{erste Würfelzahl}}, \cdots, \underbrace{\omega_n}_{\text{n-te Würfelzahl}})$,

- $\mathcal{F} = \mathcal{P}(\Omega)$,
- for $A \in \mathcal{F}$, $\mathbb{P}[A] = \frac{|A|}{|\Omega|}$.

Wir wollen beweisen, dass die Wahrscheinlichkeit, mehr als $\ell := 7\log n$ aufeinanderfolgende Einsen zu sehen, klein ist, wenn n gross ist. Betrachten wir das Ereignis A, dass es ℓ aufeinanderfolgende Einsen gibt. Für jedes $0 \le k \le n - \ell$ definiere man das Ereignis

$$A_k = \{\omega : \omega_{k+1} = \omega_{k+2} = \cdots = \omega_{k+\ell} = 1\}.$$

Wobei A_k besagt, dass wir ℓ aufeinander folgende Einsen zwischen $k+1$ und $k+\ell$ haben. Dann gilt

$$A = \bigcup_{k=0}^{n-\ell} A_k.$$

Beweismethode 1: Monotonie

Beweismethode 2: Union Bound Anwenden der Union Bound liefert

$$\mathbb{P}[A] \le \sum_{k=0}^{n-\ell} \mathbb{P}[A_k] \le n \cdot \left(\frac{1}{6}\right)^{\ell} \le n \cdot n^{-\log(7)/\log(6)},$$

somit sehen wir, dass das Ereignis mehr als $7\log n$ aufeinander folgende Einsen gegen 0 konvergiert falls wir n gegen ∞ laufen lassen.

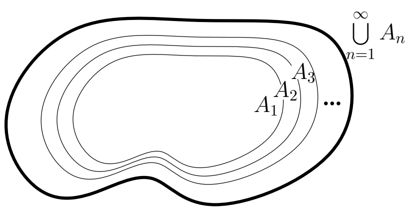
Stetigkeit von W-Massen

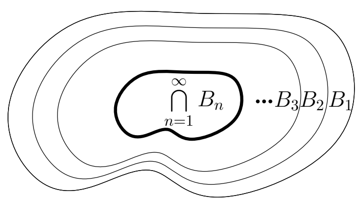
Satz 1.11. *Sei (A_n) eine monoton wachsende Folge von Ereignissen ($A_n \subset A_{n+1}$ für jedes n). Dann gilt*

$$\lim_{n\to\infty} P[A_n] = \mathbb{P}[\bigcup_{n=1}^{\infty} A_n]. \qquad \textit{monoton wachsender Grenzwert}$$

Sei (B_n) eine monoton fallende Folge von Ereignissen ($B_n \supset B_{n+1}$ for every n). Dann gilt

$$\lim_{n\to\infty} P[B_n] = \mathbb{P}[\bigcap_{n=1}^{\infty} B_n]. \qquad \textit{monoton fallender Grenzwert}$$





Bemerkung 1.12. *Durch Monotonie erhalten wir $\mathbb{P}[A_n] \le \mathbb{P}[A_{n+1}]$ und $\mathbb{P}[B_n] \ge \mathbb{P}[B_{n+1}]$ für jedes n. Daher sind die Grenzwerte in den obrigen Gleichungen wohldefiniert.*

Beweis. Sei $(A_n)_{n\geq 1}$ eine monoton steigende Folge von Ereignissen. Definiere $\widetilde{A}_1 = A_1$ für jedes $n \geq 2$ setzte

$$\widetilde{A}_n = A_n \backslash A_{n-1}.$$

Die Ereignisse $\widetilde{A}_n$ sind disjunkt und erfüllen

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \widetilde{A}_n \quad \text{und} \quad A_N = \bigcup_{n=1}^{N} \widetilde{A}_n.$$

Anwendung von Sigmaadditivität und Additivität liefert die erste Behauptung

$$\begin{aligned}
\mathbb{P}[\bigcup_{n=1}^{\infty} A_n] &= \mathbb{P}[\bigcup_{n=1}^{\infty} \widetilde{A}_n] \\
&= \sum_{n=1}^{\infty} \mathbb{P}[\widetilde{A}_n] \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{P}[\widetilde{A}_n] \\
&= \lim_{N\to\infty} \mathbb{P}[A_N].
\end{aligned}$$

Für die zweite Behauptung nehme eine monoton fallende Folge von Ereignissen (B_n). Die Folge (B_n^c) is monoton steigend und Anwendung der ersten Behauptung liefert das Gewünschte:

$$\begin{aligned}
\mathbb{P}[\bigcap_{n=1}^{\infty} B_n] &= 1 - \mathbb{P}[\bigcup_{n=1}^{\infty} B_n^c] \\
&= 1 - \lim_{n\to\infty} \mathbb{P}[B_n^c] \\
&= \lim_{n\to\infty} \mathbb{P}[B_n].
\end{aligned}$$

□

5 Bedingte Wahrscheinlichkeit

Betrachten wir ein Zufallsexperiment, welches durch einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ beschrieben wird. Häufig können wir nur durch Teilereignisse also unvollständige Information auf das tatsächliche Ergebnis schließen. Ein Beispiel hierfür ist das einmalige Würfeln wobei die tatsächliche Würfelzahl für uns unbekannt ist. Ein Freund bestätigt jedoch, dass eine gerade Zahl gewürfelt wurde. Diese Information beeinflusst unsere gesamte Wahrscheinlichkeitsberechnung. Seien A und B zwei Ereignisse und uns wird mitgeteilt, dass B eintritt. Dann kann die neue Wahrscheinlichkeit unter neuem Kenntnisstand nicht mehr $\mathbb{P}[A]$ sein. Warum ist dies so? Schließlich wissen wir, dass unter neuer Informationslage (Würfeln einer geraden Zahl) A dann und nur dann eintritt, wenn $A \cap B$ eintritt. Das bedeutet, dass die neue Wahrscheinlichkeit von A proportional zu $\mathbb{P}[A \cap B]$ ist.

Definition 1.13 (Bedingte Wahrscheinlichkeit). *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien A, B zwei Ereignisse mit $\mathbb{P}[B] > 0$. Wir definieren die* ***bedingte Wahrscheinlichkeit von A gegeben B*** *wie folgt*

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Bemerkung 1.14. $\mathbb{P}[B \mid B] = 1$. *B findet unter bedingt B stets statt.*

Beispiel: Wir betrachten den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$, der dem Wurf eines Würfels entspricht. Sei $A = \{1, 2, 3\}$ das Ereignis, dass das Würfelauge kleiner als oder gleich 3 ist. Sei $B = \{2, 4, 6\}$ das Ereignis, dass das Würfelauge gerade ist. Dann berechne

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{1/6}{1/2} = 1/3.$$

Satz 1.15. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei B ein Ereignis mit positiver Wahrscheinlichkeit. Dann ist $\mathbb{P}[\,.\mid B]$ ein W-Mass auf Ω.*

Satz 1.16 (Gesetz der totalen Wahrscheinlichkeit). *Sei $B_1, \cdots, B_n$ eine Partition[a] des Grundraums Ω, so dass $\mathbb{P}[B_i] > 0$ für jedes $1 \leq i \leq n$ gilt. Dann gilt*

$$\forall A \in \mathcal{F} \quad \mathbb{P}[A] = \sum_{i=1}^{n} \mathbb{P}[A \mid B_i]\ \mathbb{P}[B_i].$$

[a] i.e. $\Omega = B_1 \cup \cdots \cup B_n$ mit paarweise disjunkten Ereignissen.

Beweis. Unter Verwendung der Distributivität der Schnittmengenbildung gilt

$$A = A \cap \Omega = A \cap (B_1 \cup \cdots \cup B_n) = (A \cap B_1) \cup \cdots \cup (A \cap B_n).$$

Die Ereignisse $A \cap B_i$ sind paarweise disjunkt, zudem haben wir

$$\mathbb{P}[A] = \mathbb{P}[A \cap B_1] + \cdots + \mathbb{P}[A \cap B_n].$$

Zusätzlich gilt per Definition $\mathbb{P}[A \cap B_i] = \mathbb{P}[A \mid B_i]\ \mathbb{P}[B_i]$ für jedes i. Einsetzten in die obige Gleichung liefert das Gewünschte

$$\mathbb{P}[A] = \mathbb{P}[A \mid B_1]\ \mathbb{P}[B_1] + \cdots + \mathbb{P}[A \mid B_n]\ \mathbb{P}[B_n].$$

□

Satz 1.17 (Satz von Bayes). *Sei $B_1, \ldots, B_n \in \mathcal{F}$ eine Partition von Ω sodass, $\mathbb{P}[B_i] > 0$ für jedes i gilt. Für jedes Ereignis A mit $\mathbb{P}[A] > 0$ gilt*

$$\forall i = 1, \ldots, n \quad \mathbb{P}[B_i \mid A] = \frac{\mathbb{P}[A \mid B_i] \; \mathbb{P}[B_i]}{\sum_{j=1}^{n} \mathbb{P}[A \mid B_j] \; \mathbb{P}[B_j]}.$$

Klassische Anwendung: Es wird ein Test durchgeführt, um eine bestimmte seltene Krankheit zu diagnostizieren, die 1/10000 der Bevölkerung befällt. Dieser Test ist recht zuverlässig und gibt in 99 Prozent der Fälle ein richtiges Testergebnis. Wenn ein Patient ein "positives Testergebnis" hat (d.h. der Test behauptet, dass der Patient infiziert ist), wie groß ist die ist die Wahrscheinlichkeit, dass er tatsächlich krank ist?

Zuerst modellieren wir die Situation mittels Grundraum

$$\Omega = \{0, 1\} \times \{0, 1\}.$$

und $\mathcal{F} = \mathcal{P}(\Omega)$. Ein Ereignis beschreibt einen Patienten mittels Paar $\omega = (\omega_1, \omega_2)$, wobei

$$\omega_1 = \begin{cases} 0 & \text{falls der Patient gesund ist,} \\ 1 & \text{falls der Patient infiziert ist,} \end{cases}$$

$$\omega_2 = \begin{cases} 0 & \text{falls der Test negativ ausfällt,} \\ 1 & \text{falls der Test positiv ausfällt.} \end{cases}$$

Sei S ein Ereignis sodass, der Patient infiziert ist und sei T ein Ereignis sodass, der Patient positiv getestet ist. Alle Elemente aus S sind Paare $\omega = (\omega_1, \omega_2)$ mit $\omega_1 = 1$ sodass,

$$S = \{(1, 0), (1, 1)\}.$$

Ebenso haben wir

$$T = \{(0, 1), (1, 1)\}.$$

Aufgrund der Beschreibung des Zufallsexperiments haben wir zudem folgende Informationen über unser W-Mass

$$\mathbb{P}[S] = \frac{1}{10000}, \quad \mathbb{P}[T \mid S] = \frac{99}{100}, \quad \mathbb{P}[T \mid S^c] = \frac{1}{100}.$$

Uns interessiert nun die posteriori Wahrscheinlichkeit , dass ein getesteter Mensch infiziert ist bedingt unter dem Ereignis eines positiven Tests ($\mathbb{P}[S \mid T]$). Anwenden des Satz von Bayes mittels Partition $\Omega = S \cup S^c$ liefert

$$\begin{aligned} \mathbb{P}[S \mid T] &= \frac{\mathbb{P}[T \mid S] \; \mathbb{P}[S]}{\mathbb{P}[T \mid S] \; \mathbb{P}[S] + \mathbb{P}[T \mid S^c] \; \mathbb{P}[S^c]} \\ &= \frac{0.99 \times 0.0001}{0.99 \times 0.0001 + 0.01 \times 0.9999} \simeq 0.0098. \end{aligned}$$

Dieses Ergebnis ist recht überraschend: Die Wahrscheinlichkeit, bei einem positiven Test tatsächlich krank zu sein, ist sehr gering!

Wie lässt sich dies erklären? Betrachtet man die Gesamtbevölkerung, so gibt es zwei Arten von Personen, die positiv Test getestet werden:

- Gesunde Personen mit einem (fälschlicherweise) positiven Test, welche etwa ein Prozent der Bevölkerung ausmachen.

- Erkrankte Personen mit einem (korrekten) positiven Test, die etwa 1/10000 der Bevölkerung ausmachen.

Bedingt auf einen positiven Test, hat eine Person eine viel größere Chance, zur ersten Gruppe zu gehören.

6 Unabhängigkeit

Unabhängigkeit von Ereignissen

Definition 1.18 (Unabhängigkeit von zwei Ereignissen)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum. Zwei Ereignisse* A *und* B *heissen* ***unabhängig*** *falls*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\ \mathbb{P}[B].$$

Bemerkung 1.19. *Falls* $\mathbb{P}[A] \in \{0, 1\}$, *dann ist* A *unabhängig von jedem Ereignis sodass,*

$$\forall B \in \mathcal{F} \quad \mathbb{P}[A \cap B] = \mathbb{P}[A]\ \mathbb{P}[B].$$

Falls ein Ereignis A *unabhängig von sich selbst ist, also* $\mathbb{P}[A \cap A] = \mathbb{P}[A]^2$*) gilt, dann muss* $\mathbb{P}[A] \in \{0, 1\}$ *gelten.*

A *ist unabhängig von* B *genau dann wenn* A *unabhängig von* B^c *ist.*

Der Begriff der Unabhängigkeit ist in der Wahrscheinlichkeitsrechnung von grundlegender Bedeutung: Er entspricht der intuitiven Vorstellung, dass sich zwei Ereignisse nicht gegenseitig beeinflussen. Die nachfolgende Proposition illustriert dies

Satz 1.20. *Seien* $A, B \in \mathcal{F}$ *zwei Ereignisse mit* $\mathbb{P}[A], \mathbb{P}[B] > 0$. *Dann sind folgende Aussagen äquivalent:*

(i) $\mathbb{P}[A \cap B] = \mathbb{P}[A]\ \mathbb{P}[B]$, *A und B sind unabhängig*

(ii) $\mathbb{P}[A \mid B] = \mathbb{P}[A]$, *Eintreten von B hat keinen Einfluss auf A*

(iii) $\mathbb{P}[B \mid A] = \mathbb{P}[B]$. *Eintreten von A hat keinen Einfluss auf B*

Beweis. Sei $\mathbb{P}[B] > 0$ dann gilt

$$(i) \Leftrightarrow \left(\frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \mathbb{P}[A]\right) \Leftrightarrow (\mathbb{P}[A \mid B] = \mathbb{P}[A]) \Leftrightarrow \ (ii).$$

Da (iii) gerade (ii) mit vertauschten Rollen von A und B ist, kann die Äquivalenz (i)$\Leftrightarrow$ (iii) nach obiger Logik genauso bewiesen werden. $\square$

Typische Beispiele für unabhängige Ereignisse treten beim mehrfachen Wiederholen von Zufallsexperiment auf. Das nachfolgende Beispiel von zweier unabhängiger Würfel illustriert dies

Beispiel: Wurf zweier unabhängiger Würfel

Wir werfen zwei voneinander unabhängige Würfel. Dies wird durch das Laplace-Modell des Grundraums modelliert

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2.$$

Betrachte folgende Ereignisse

$$\begin{aligned}
A &= \{\omega : \omega_1 \in 2\mathbb{Z}\}, && \text{“Erstes Würfelauge ist gerade”}\\
B &= \{\omega : 1 + \omega_2 \in 2\mathbb{Z}\}, && \text{“Zweites Würfelauge ist ungerade”}\\
C &= \{\omega : \omega_1 + \omega_2 \le 3\}, && \text{“Die Summe beider Augen ist höchstens 3”}\\
D &= \{\omega : \omega_1 \le 2, \omega_2 \le 2\}. && \text{“Beide Augen sind kleiner als 2”}
\end{aligned}$$

Überprüfe, dass

- A und B sind unabhängig,
- A und C sind nicht unabhängig,
- A und D sind unabhängig.

Definition 1.21. *Sei I eine beliebe Indexmenge. Eine Familie von Ereignissen $(A_i)_{i\in I}$ heisst **unabhängig** falls*

$$\forall J \subset I \text{ endlich} \qquad \mathbb{P}[\bigcap_{j\in J} A_j] = \prod_{j\in J} \mathbb{P}[A_j].$$

Bemerkung: Drei Ereignisse A, B und C sind unabhängig falls alle 4 folgenden Gleichungen erfüllt sind (nicht nur die Letzte!):

$$\begin{aligned}
\mathbb{P}[A \cap B] &= \mathbb{P}[A]\ \mathbb{P}[B],\\
\mathbb{P}[A \cap C] &= \mathbb{P}[A]\ \mathbb{P}[C],\\
\mathbb{P}[B \cap C] &= \mathbb{P}[B]\ \mathbb{P}[C],\\
\mathbb{P}[A \cap B \cap C] &= \mathbb{P}[A]\ \mathbb{P}[B]\ \mathbb{P}[C].
\end{aligned}$$

Beispiel: Wir verwenden die gleiche Notation wie im obigen Beispiel Definition 1.21. Die Ereignisse A, B, und D sind unabhängig (Überprüfe dies!).

Kapitel 2

Zufallsvariablen und Verteilungsfunktionen

Ziele

- Verstehen der Definition der Zufallsvariable und ihrer Verteilungsfunktion.
- Anwendung von Abstrakten Wahrscheinlichkeitsräumen $(\Omega, \mathcal{F}, \mathbb{P})$.
- Nachvollziehen der eingeführten Notation bezüglich dem Beschreiben von Ereignissen durch Zufallsvariablen.
- Konkrete Konstruktion von Zufallsvariablen durch Folgen von u.i.v. Bernoulli Zufallsvariablen.

1 Abstrakte Definition

Häufig ist das Modellieren eines wahrscheinlichkeitstheoretischen Modells ziemlich zu abstrakt und kompliziert und man ist nur an bestimmten Größen des Modells interessiert. Aus diesem Grund führt man den Begriff der Zufallsvariable ein.

Definition 2.1. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine* ***Zufallsvariable*** *(Z.V.) ist eine Abbildung $X : \Omega \to \mathbb{R}$ sodass, für alle $a \in \mathbb{R}$ gilt,*

$$\{\omega \in \Omega \,:\, X(\omega) \leq a\} \in \mathcal{F}.$$

➜ Die Bedingung $\{\omega \in \Omega \,:\, X(\omega) \leq a\} \in \mathcal{F}$ ist notwendig für die Wohldefiniertheit von $\mathbb{P}[\{\omega \in \Omega \,:\, X(\omega) \leq a\}]$.

Example 1: *Wetten auf ein Würfelergebnis*

Wir werfen einen fairen Würfel. Der Stichprobenraum ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ und wir betrachten das Laplace-Modell $(\Omega, \mathcal{F}, \mathbb{P})$ nach Definition 1.5. Nehmen wir an, dass wir so auf das Ergebnis wetten sodass, unser Gewinn beschrieben wird durch

$$\begin{aligned} -1 &\text{ Würfelauge zeigt 1, 2 oder 3,} \\ 0 &\text{ Würfelauge zeigt 4,} \\ 2 &\text{ Würfelauge zeigt 5 oder 6,} \end{aligned}$$

wobei ein negativer Gewinn einen Verlust darstellt. Wir wollen nun unseren Gewinn durch eine Zufallsvariable X modellieren. Wir definieren daher

$$\forall \omega \in \Omega \quad X(\omega) = \begin{cases} -1 & \text{if } \omega = 1, 2, 3, \\ 0 & \text{if } \omega = 4, \\ 2 & \text{if } \omega = 5, 6. \end{cases} \tag{2.1}$$

Da $\mathcal{F} = \mathcal{P}(\Omega)$, gilt $\{\omega \,:\, X(\omega) \leq a\} \in \mathcal{F}$ für jedes a. Nach Definition ist X somit eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$.

Example 2: *Indikatorfunktion auf einem Ereigniss*

Sei $A \in \mathcal{F}$. Wir definieren die **Indikatorfunktion** $\mathbb{1}_A$ auf A, durch

$$\forall \omega \in \Omega \quad \mathbb{1}_A(\omega) = \begin{cases} 0 & \text{if } \omega \notin A, \\ 1 & \text{if } \omega \in A. \end{cases}$$

Somit ist $\mathbb{1}_A$ per Definition eine Zufallsvariable, denn es gilt

$$\{\omega \,:\, \mathbb{1}_A(\omega) \leq a\} = \begin{cases} \varnothing & \text{if } a < 0, \\ A^c & \text{if } 0 \leq a < 1, \\ \Omega & \text{if } a \geq 1, \end{cases}$$

wobei $\varnothing$, A^c und Ω Elemente in $\mathcal{F}$ sind.

Bemerkung: Die Rolle der σ-Algebra Nehmen wir die gleiche Notation wie in Beispiel 1. Wir betrachten zusätzlich folgende zwei Sigma-Algebren:

$$\mathcal{F}_1 = \{\varnothing, \{1,2,3\}, \{4,5,6\}, \{1,2,3,4,5,6\}\},$$
$$\mathcal{F}_2 = \{\varnothing, \{1,2,3\}, \{1,2,3,4\}, \{4,5,6\}, \{5,6\}, \{1,2,3,5,6\}, \{4\}, \{1,2,3,4,5,6\}\}.$$

Ist X eine Zufallsvariable auf $(\Omega, \mathcal{F}_i, \mathcal{P}\}$? Um diese Frage zu beantworten, muss man die Menge $\{\omega : X(\omega) \leq a\}$ genauer untersuchen. Hier sehen wir, dass

$$\{\omega : X(\omega) \leq a\} = \begin{cases} \varnothing & \text{für } a < -1, \\ \{1,2,3\} & \text{für } -1 \leq a < 0, \\ \{1,2,3,4\} & \text{für } 0 \leq a < 2, \\ \{1,2,3,4,5,6\} & \text{für } a \geq 2. \end{cases}$$

Insbesondere sehen wir, dass X eine Zufallsvariable auf $(\Omega, \mathcal{F}_2, \mathcal{P})$ aber nicht auf $(\Omega, \mathcal{F}_1, \mathcal{P})$ ist.

Notation: Für Ereignisse im Bezug auf Z.V. werden wir auf darauf **verzichten** sie **mittels Beziehung zu ω darzustellen**. Stattdessen schreiben wir für $a \leq b$

$$\{X \leq a\} = \{\omega \in \Omega : X(\omega) \leq a\},$$
$$\{a < X \leq b\} = \{\omega \in \Omega : a < X(\omega) < b\},$$
$$\{X \in \mathbb{Z}\} = \{\omega \in \Omega : X(\omega) \in \mathbb{Z}\}.$$

Betrachten wir die Wahrscheinlichkeit nach obigen Beispiel. Dann lassen wir gerade die Klammern weg und schreiben einfach

$$\mathbb{P}[X \leq a] = \mathbb{P}[\{X \leq a\}] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \leq a\}].$$

2 Verteilungsfunktion

Definition 2.2. *Sei X eine Zufallsvariable auf einem W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$. Die **Verteilungsfunktion von** X ist eine Funktion $F_X : \mathbb{R} \to [0,1]$, definiert durch*

$$\forall a \in \mathbb{R} \quad F_X(a) = \mathbb{P}[X \leq a].$$

Example 1: *Wetten auf einen Würfel*

Sei X eine Zufallsvariable definiert durch Eq. (2.1). Für $a \in \mathbb{R}$, schreiben wir

$$F_X(a) = \begin{cases} 0 & \text{if } a < -1, \\ 1/2 & \text{if } -1 \le a < 0, \\ 2/3 & \text{if } 0 \le a < 2, \\ 1 & \text{if } a \ge 2. \end{cases}$$

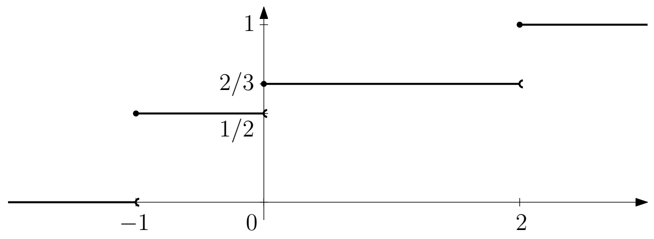


Abbildung 2.1: Graph der Verteilungsfunktion F_X.

Example 2: *Indikatorfunktion eines Ereignisses*

Sei A ein Ereignis. Sei $X = \mathbb{1}_A$ eine Indikatorfunktion auf einem Ereignis A. Dann gilt

$$F_X(a) = \begin{cases} 0 & \text{falls } a < 0, \\ 1 - \mathbb{P}[A] & \text{falls } 0 \le a < 1, \\ 1 & \text{falls } a \ge 1. \end{cases}$$

Satz 2.3 (Einfache Identität)**.** *Seien $a < b$ zwei reelle Zahlen. Dann gilt*

$$\mathbb{P}[a < X \le b] = F(b) - F(a).$$

Beweis. Betrachte die disjunkte Vereinigung $\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}$. Somit gilt

$$\mathbb{P}[X \le b] = \mathbb{P}[X \le a] + \mathbb{P}[a < X \le b],$$

und aus obiger Gleichung folgt die Behauptung. $\square$

Theorem 2.4 (Eigenschaften der Verteilungsfunktion)**.** *Sei X eine Z.V. auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Die Verteilungsfunktion $F = F_X : \mathbb{R} \to [0,1]$ von X erfüllt folgende Eigenschaften.*

(i) *F ist monoton wachsend.*

(ii) *F ist rechtsstetig*[a]*.*

(iii) $\lim_{a\to-\infty} F(a) = 0$ *und* $\lim_{a\to\infty} F(a) = 1$.

[a] Formal: $F(a) = \lim_{h\downarrow 0} F(a+h)$ für jedes $a \in \mathbb{R}$.

Beweis. Wir beweisen (i), dann (iii) und schließlich (ii).

(i) Für $a \le b$, haben wir $\{X \le a\} \subset \{X \le b\}$. Unter Anwendung der Monotonie des W-Masses gilt $\mathbb{P}[X \le a] \le \mathbb{P}[X \le b]$. Dies liefert das Gewünschte

$$F(a) \le F(b).$$

(iii) Sei $a_n \uparrow \infty$. Für jedes $\omega \in \Omega$, existiert ein n groß genug, sodass $X(\omega) \le a_n$. Also gilt,

$$\Omega = \bigcup_{n\ge1} \{X \le a_n\}.$$

Zudem, haben wir $\{X \le a_n\} \subset \{X \le a_{n+1}\}$ und durch die Stetigkeit des W-Masses gilt

$$\begin{aligned} 1 = \mathbb{P}[\Omega] &= \mathbb{P}[\bigcup_{n\ge1}\{X \le a_n\}] \\ &= \lim_{n\to\infty} \mathbb{P}[X \le a_n] \\ &= \lim_{n\to\infty} F(a_n). \end{aligned}$$

Analog folgt $\lim_{a\to-\infty} F(a) = 0$. Dafür nehme man $a_n \downarrow -\infty$, dann gilt

$$\varnothing = \bigcap_{n\ge1}\{X \le a_n\}$$

und $\{X \le a_n\} \supset \{X \le a_{n+1}\}$. Erneutes Anwenden der Stetigkeit des W-Masses liefert das Gewünschte

$$0 = \mathbb{P}[\varnothing] = \lim_{n\to\infty} \mathbb{P}[X \le a_n] = \lim_{n\to\infty} F(a_n).$$

(ii) Sei $a \in \mathbb{R}$, sei $h_n \downarrow 0$. Es gilt

$$\{X \le a\} = \bigcap_{n\ge1}\{X \le a + h_n\},$$

wobei $\{X \le a + h_n\} \supset \{X \le a + h_{n+1}\}$. Anwendung der Stetigkeit des W-Masses liefert das Gewünschte

$$F(a) = \mathbb{P}[X \le a] = \mathbb{P}[\bigcap_{n\ge1}\{X \le a + h_n\}] = \lim_{n\to\infty} \mathbb{P}[X \le a + h_n] = \lim_{n\to\infty} F[a + h_n].$$

□

3 Unabhängigkeit von Zufallsvariablen

Definition 2.5. *Seien $X_1, \dots, X_n$ Zufallsvariablen auf einem W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$. Dann heissen $X_1, \dots, X_n$* ***unabhängig*** *falls*

$$\forall x_1, x_2, \dots, x_n \in \mathbb{R} \qquad \mathbb{P}[X_1 \leq x_1, \dots, X_n \leq x_n] = \mathbb{P}[X_1 \leq x_1] \dots \mathbb{P}[X_n \leq x_n]. \tag{2.2}$$

Bemerkung 2.6. *Man kann zeigen, dass $X_1, \dots, X_n$ genau dann unabhängig sind, wenn folgende Bedingung gilt*

$$\forall\, I_1 \subset \mathbb{R}, \dots, I_n \subset \mathbb{R} \text{ Intervalle } \{X_1 \in I_1\}, \dots, \{X_n \in I_n\} \text{ sind unabhängig }.$$

Example 1: 2-*faches unabhängiges Würfeln*

Wir betrachten das Laplace modell $(\Omega, \mathcal{F}, P)$ auf $\Omega = \{1,2,3,4,5,6\}^2$. Ein Element $\omega \in \Omega$ ist gerade ein Paar (ω_1, ω_2). Die erste Koordinate entspricht dem Wert des ersten Würfelauges. Die zweite Koordinate entspricht dem Wert des zweiten Würfelauges. Wir definieren folgende Zufallsvariablen $X, Y, Z : \Omega \to \mathbb{R}$ durch

$$X(\omega) = \omega_1 \, Y(\omega) = \omega_2 \, Z(\omega) = \omega_1 + \omega_2.$$

Dabei ist Z gerade der Summe beider Würfelaugen. Per Konstruktion sind X und Y gerade unabhängig. Seien $I, J \subset \{1, \dots, 6\}$, dann erhalten wir

$$\begin{aligned} P[X \in I, Y \in J] =& P[I \times J] = \frac{|I \times J|}{|\Omega|} = \frac{|I|}{6} \cdot \frac{|J|}{6} \\ =& \frac{|I \times \{1, \dots, 6\}|}{6} \cdot \frac{|J \times \{1, \dots, 6\}|}{6} = P[X \in I] P[Y \in J]. \end{aligned}$$

Somit existiert für jedes $x, y \in \mathbb{R}$, Mengen $I, J \subset \{1, \dots, 6\}$, sodass $[X \leq x] = [X \in I]$ und $[Y \leq y] = [Y \in J]$ gilt. Dies liefert das Gewünschte

$$P[X \leq x, Y \leq y] = P[X \leq x] P[Y \leq y].$$

Beobachte allerdings, dass X und Z nicht unabhängig sind

$$\frac{1}{6^2} = P[X \leq 1, Z \leq 2] \neq P[X \leq 1] P[Z \leq 2] = \frac{1}{6^3}.$$

Gruppierung von Zufallsvariablen

Wenn wir eine Menge unabhängiger Zufallsvariablen haben und disjunkte Gruppen solcher Zufallsvariablen bilden, dann sind diese Gruppen auch wiederum unabhängig voneinander. Diese Idee wird durch den folgenden Satz formalisiert.

Satz 2.7 (Gruppieren von Zufallsvariablen). *Seien $X_1, \dots, X_n$ n unabhängige Zufallsvariablen. Seien $1 \leq i_1 < i_2 < \dots < i_k \leq n$ Indexes und $\phi_1, \dots, \phi_k$ Abbildungen. Dann sind*

$$Y_1 = \phi_1(X_1, \dots, X_{i_1}), Y_2 = \phi_2(X_{i_1+1}, \dots, X_{i_2}) \dots, Y_k = \phi_k(X_{i_{k-1}+1}, \dots, X_{i_k})$$

unabhängig.

Beweis. Ausgelassen. □

Folgen von u.i.v. Zufallsvariablen

Definition 2.8. *Eine Folge von Zufallsvariablen $X_1, X_2, \dots$ heisst*

1. ***unabhängig*** *falls $X_1, \dots, X_n$ unabhängig sind, für alle $n \in \mathbb{N}$.*
2. ***unabhängig und identisch verteilt (uiv)*** *falls sie unabhängig ist und die Zufallsvariablen dieselbe Verteilungsfunktion haben d.h.*

$$\forall i, j \quad F_{X_i} = F_{X_j}.$$

4 Transformation von Zufallsvariablen

Wenn wir einige Zufallsvariablen $X_1, X_2, \dots$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ haben, können mittels Komposition von Funktionen neue Zufallsvariablen auf demselben Wahrscheinlichkeitsraum erzeugen. So kann man z.B. $Z_1 = \exp(X_1)$, $Z_2 = X_1 + X_2$ als neue Zufallsvariablen betrachten. Dabei sollte man nicht vergessen, dass Zufallsvariablen Abbildungen von Ω nach $\mathbb{R}$ sind. Zum Beispiel entsprechen die Zufallsvariablen Z_1 und Z_2 den Abbildungen, die für jedes $\omega \in \Omega$ definiert sind durch

$$Z_1(\omega) = \exp(X_1(\omega)), \qquad Z_2(\omega) = X_1(\omega) + X_2(\omega).$$

Formal führen wir die folgende Notation ein, die es uns erlaubt, mit Zufallsvariablen so zu behandeln, als wären sie nur reelle Zahlen. Falls X eine Zufallsvariable ist, und $\phi : \mathbb{R} \to \mathbb{R}$, so schreiben wir

$$\phi(X) := \phi \circ X.$$

Somit ist $\phi(X)$ eine neue Abbildung von $\Omega \to \mathbb{R}$, welche in dem nachfolgenden Diagram dargestellt ist.

$$\begin{array}{ccccc} \Omega & \xrightarrow{X} & \mathbb{R} & \xrightarrow{\phi} & \mathbb{R} \\ \omega & \longmapsto & X(\omega) & \longmapsto & \phi(X(\omega)). \end{array}$$

Ganz allgemein, können wir ebenfalls Funktionen mehrerer Variablen betrachten. Seien $X_1, \ldots, X_n$ n Zufallsvariablen und $\phi : \mathbb{R}^n \to \mathbb{R}$ eine Abbildung, so schreiben wir

$$\phi(X_1, \ldots, X_n) := \phi \circ (X_1, \ldots, X_n).$$

5 Konstruktion von Zufallsvariablen

In Abschnitt 1 haben wir Zufallsvariablen definiert. In Abschnitt 2 haben wir gesehen, dass wir jeder Zufallsvariablen X eine Verteilungsfunktion $F = F_X : \mathbb{R} \to [0,1]$ zuordnen können, die ihre wahrscheinlichkeitstheoretischen Eigenschaften umfasst und folgende Bedingungen erfüllt

(i) F monoton steigend

(ii) F rechtsstetig

(iii) $\lim_{a \to -\infty} F(a) = 0$ und $\lim_{a \to \infty} F(a) = 1$.

Umgekehrt kann man nun fragen, ob es für jede Abbildung $F : \mathbb{R} \to [0,1]$, welche Eigenschaft (i)–(iii) erfüllt, eine Zufallsvariable X mit Eigenschaft $F_X = F$ existiert?

Das Ziel dieses Abschnitts ist es, allgemeine Zufallsvariablen zu konstruieren und die obige Frage positiv zu beantworten. Eine vollständige Konstruktion würde einige Hilfsmittel aus der Maßtheorie erfordern, die den Rahmen dieses Kurses sprengen würden: Unser Ansatz stützt sich auf ein abstraktes Theorem von Kolmogorov, das die Existenz von uiv Folgen von Zufallsvariablen garantiert. Der Beweis dieses Theorems wird ausgelassen, jedoch wollen wir die Konstruktion rigoros beweisen. Unsere Motivation ist eine doppelt motiviert

- Auf theoretischer Ebene ist die Existenz von Zufallsvariablen grundlegend: "Es ist notwendig, dass die mathematischen Objekte über die wir sprechen, überhaupt existieren!"
- Auf praktischer Ebene bietet die hier vorgestellte explizite Konstruktion ein allgemeines Rezept zur Konstruktion von Zufallsvariablen. Damit kann eine beliebige Zufallsvariable auch numerisch simuliert werden, sofern ihre Verteilungsfunktion gegeben ist.

Die Existenz folgt den folgenden 4 Schritten.

Schritt 1: Existenzsatz von Kolmogorov und uiv Folgen von Ber. Z.V.

Unsere Konstruktion beginnt mit Bernoulli-Zufallsvariablen, welche wir nun definieren.

Definition 2.9. *Sei* $p \in [0,1]$. *Eine Zufallsvariable* X *heißt* ***Bernoulli Zufallsvariable mit Parameter*** p *falls*

$$\mathbb{P}[X = 0] = 1 - p \quad \textit{und} \quad \mathbb{P}[X = 1] = p.$$

Dabei schreiben wir stets $X \sim \mathrm{Ber}(p)$.

Example: *n-facher Münzwurf*

Wir wollen ein Modell für n aufeinanderfolgende unabhängige Münzwürfe definieren. Betrachten wir den Grundraum $\Omega = \{0,1\}^n$ mittels Laplace-Modell $(\mathcal{F}, \mathbb{P})$. Definiere die Zufallsvariablen

$$X_i : \begin{array}{ccc} \Omega & \to & \{0,1\} \\ (\omega_1, \dots, \omega_n) & \mapsto & \omega_i \end{array}.$$

(X_i repräsentiert den i-ten Münzwurf, $X_i = 1$ besagt, dass der i-te Münzwurf gerade Kopf zeigt, wohingegen $X_i = 0$ gerade das Ergebnis von Zahl darstellt.) Per Konstruktion sind $X_1, \dots, X_n$ gerade unabhängige Bernoulli Zufallsvariabeln mit Parameter $p = 1/2$.

Um zu beweisen, dass $X_1 \sim \mathrm{Ber}(1/2)$ gilt, berechnen wir gerade die Ereignisse $\{X_1 = 0\} = \{0\} \times \{0,1\}^{n-1}$ und $\{X_1 = 1\} = \{1\} \times \{0,1\}^{n-1}$. Anwendung der Definition von $\mathbb{P}$ im Laplace Modell liefert:

$$\mathbb{P}[X_1 = 0] = \frac{|\{0\} \times \{0,1\}^{n-1}|}{|\Omega|} = \frac{1}{2} \quad \text{and} \quad \mathbb{P}[X_1 = 1] = \frac{|\{1\} \times \{0,1\}^{n-1}|}{|\Omega|} = \frac{1}{2}.$$

Äquivalent kann man zeigen, dass jede X_i, $1 \le i \le n$ eine Bernoulli-Zufallsvariable mit dem Parameter $1/2$ ist.

Für die Unabhängigkeit ist es hinreichend Gleichung (2.2) für $x_1, \dots, x_n \in \{0,1\}$ nachzuweisen. Für solche Zahlen gilt mit $|\{0, x_i\}| = 1 + x_i$ gerade

$$\begin{aligned}
\mathbb{P}[X_1 \le x_1, \dots, X_n \le x_n] &= \mathbb{P}[\{0, x_1\} \times \dots \times \{0, x_n\}] \\
&= \frac{|\{0, x_1\} \times \dots \times \{0, x_n\}|}{|\Omega|} \\
&= \frac{1 + x_1}{2} \cdots \frac{1 + x_n}{2} = \mathbb{P}[X_1 \le x_1] \cdots \mathbb{P}[X_n \le x_n].
\end{aligned}$$

Für jedes $n \geq 1$ konstruiert das obige Beispiel einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und n unabhängige Bernoulli-Zufallsvariablen $X_1, \ldots, X_n$ mit Parameter $1/2$. Ebenso ist es natürlich, eine unendliche Folge von unabhängigen Bernoulli-Zufallsvariablen $X_1, X_2, \ldots$ zu betrachten. Die Konstruktion eines geeigneten Wahrscheinlichkeitsraumes ist dabei deutlich komplizierter und ist Inhalt des folgenden Theorems.

Theorem 2.10 (Existenzsatz von Kolmogorov)**.** *Es existiert ein W-Raum* $(\Omega, \mathcal{F}, \mathbb{P})$ *und eine nicht endliche uiv Folge von Bernoulli Zufallsvariablen* $X_1, X_2, \ldots$ *auf* $(\Omega, \mathcal{F}, \mathbb{P})$ *mit Parameter* $1/2$.

Beweis. Ausgelassen. □

Schritt 2: Konstruktion von gleichverteilten Zufallsvariablen auf $[0,1]$

Wir verwenden Bernoulli-Zufallsvariablen, um eine gleichmäßige Zufallsvariable auf $[0,1]$ zu konstruieren. Intuitiv kann man sich einen Wassertropfen vorstellen, der in das Intervall $[0,1]$ fällt. Wir nehmen an, dass der Tropfen gleichmässig auf das Intervall fällt. Zum Beispiel ist die Wahrscheinlichkeit, in $[0,1, 0.2]$ zu fallen, dieselbe wie in $[0.8, 0.9]$ zu fallen. Eine gleichförmige Zufallsvariable in $[0,1]$ repräsentiert die Position, an der ein solcher Tropfen fällt.

Definition 2.11. *Eine Zufallsvariable U heißt* ***gleichverteilt auf*** $[0,1]$ *falls ihre Verteilungsfunktion gegeben ist durch*

$$F_U(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1. \end{cases}$$

Wir schreiben gerade $U \sim \mathcal{U}([0,1])$.

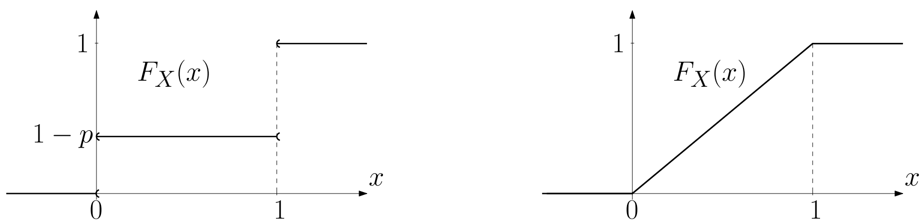


Abbildung 2.2: Links: Verteilungsfunktion einer Bernoulli-Zufallsvariable mit Parameter p . Rechts: Verteilungsfunktion einer gleichverteilten Zufallsvariable auf $[0,1]$.

Seien $X_1, X_2, \ldots$ eine Folge von unabhängigen Bernoulli-Zufallsvariablen mit Parameter $1/2$. Für jedes festes ω haben wir $X_1(\omega), X_2(\omega) \cdots \in \{0,1\}$. Daraus folgt, dass die unendliche Reihe

$$Y(\omega) = \sum_{n=1}^{\infty} 2^{-n} X_n(\omega) \tag{2.3}$$

absolut konvergiert, wobei $Y(\omega) \in [0,1]$ ist.

Satz 2.12. *Die Abbildung* $Y : \Omega \to [0,1]$ *definiert in* (2.3) *ist eine gleichverteilte Zufallsvariable auf* $[0,1]$.

Schritt 3: Konstruktion von Z.V. mit beliebiger Verteilungsfunktion F

Sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion, welche Eigenschaften (i)–(iii) erfüllt.

Falls F streng monoton steigend und stetig ist, dann ist F bijektiv und man kann eine Umkehrfunktion F^{-1} definieren. Für jedes $\alpha \in [0,1]$ ist $F^{-1}(\alpha)$ die eindeutige reelle Zahl x, für die $F(x) = \alpha$ gilt. In einem solchen Fall definiert dies die inverse Verteilungsfunktion. Allgemeiner kann man eine verallgemeinerte Umkehrfunktion die sogenannte verallgemeinerte Inverse für F definieren.

Definition 2.13 (Pseudoinverse)**.** *Die Pseudoinverse von* F *ist eine Abbildung* $F^{-1} : (0,1) \to \mathbb{R}$ *definiert durch*

$$\forall \alpha \in (0,1) \quad \boxed{F^{-1}(\alpha) = \inf\{x \in \mathbb{R} \; : \; F(x) \geq \alpha\}}.$$

Nach Definition des Infimums und unter Verwendung der rechten Stetigkeit von F ergibt sich für jedes $x \in \mathbb{R}$ und $\alpha \in (0,1)$

$$\left(F^{-1}(\alpha) \leq x\right) \iff (\alpha \leq F(x)).$$

Ausgehend von der Pseudoinversen bietet das folgende Theorem eine Möglichkeit, eine Zufallsvariable mit beliebiger Verteilungsfunktion zu konstruieren.

Theorem 2.14 (Inversionsmethode)**.** *Sei* $F : \mathbb{R} \to [0,1]$ *eine Abbildung, welche Eigenschaften (i)–(iii) erfüllt. Sei* U *eine Gleichverteilte Zufallsvariable. Dann besitzt die Zufallsvariable*

$$X = F^{-1}(U)$$

gerade die Verteilungsfunktion $F_X = F$.

Bemerkung 2.15. *Wir wollen nochmals kurz erläutern, warum die Definition von* X *nach 2.14 wohldefiniert ist. Sei* $U : \Omega \to [0,1]$ *und* $F^{-1} : (0,1) \to \mathbb{R}$ *analog zum obigen Theorem definiert. Dann gilt stets* $P[U \in (0,1) = 1]$. *Strenggenommen ist* X *bis jetzt nur auf einer Menge mit Wahrscheinlichkeit* 1 *aber nicht auf ganz* Ω *definiert. Wir beheben das Problem mittels folgender Definition*

$$X(\omega) = \begin{cases} F^{-1}(U(\omega)) & \textit{falls } U(\omega) \in (0,1) \\ 0 & \textit{sonst.} \end{cases}$$

(Dabei spielt 0 *selbst keine Rolle, man hätte jede beliebige reelle Zahl nehmen können.)*

Beweis. Für jedes $x \in \mathbb{R}$ gilt

$$\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x).$$

□

Schritt 4: Allgemeine Folgen von unabhängigen Zufallsvariablen

Theorem 2.16. *Seien $F_1, F_2 \dots$ eine Folge von Funktionen $\mathbb{R}$ auf $[0,1]$, die die Eigenschaften (i)–(iii) am Anfang des Abschnitts erfüllen. Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Folge von unabhängigen Zufallsvariablen $X_1, X_2, \dots$ auf diesem Wahrscheinlichkeitsraum, sodass*

- *für jedes i gilt: X_i hat Verteilungsfunktion F_i (d.h. $\forall x\ \mathbb{P}[X_i \le x] = F_i(x)$), und*
- *$X_1, X_2, \dots$ sind unabhängig.*

Beweis. Siehe Übungsserie. $\square$

Der obige Satz ist in der Theorie wichtig, weil er uns erlaubt, direkt mit Zufallsvariablen zu arbeiten, ohne den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ genauer zu definieren. Wenn zum Beispiel F und G zwei gegebene Verteilungsfunktionen sind, können wir stets zum Beispiel schreiben:

"Seien X, Y zwei unabhängige Zufallsvariablen mit Verteilungsfunktionen F und G.".

Kapitel 3

Diskrete und stetige Zufallsvariablen

Ziele

- Definition von diskreten und stetigen Zufallsvariablen.
- Klassische Beispiele für diskrete und kontinuierliche Zufallsvariablen: Motivation, Beziehung verschiedenen Zufallsvariablen.
- Wahrscheinlichkeitstheoretische Interpretation der Eigenschaften von F_X.
- Wahrscheinlichkeitsdichte f_X einer Zufallsvariable: Interpretation, Zusammenhang mit der Verteilungsfunktion F_X.

Rahmenbedingung In diesem Kapitel fixieren wir stets einen W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$.

1 Unstetigkeit/Stetigkeit der Verteilungsfunktion F

Wir haben gesehen, dass die Verteilungsfunktion $F = F_X$ einer Zufallsvariablen X immer rechtsstetig ist. Wie sieht es aber mit Linksstetigkeit aus?

Für eine Bernoulli-Zufallsvariable $X \sim \mathrm{Ber}(p)$ mit $p < 1$ haben wir $F_X(-h) = 0$ für jedes $h > 0$, aber $F_X(0) = 1 - p \neq 0$. Daher ist F_X nicht linksstetig in 0, d.h.

$$\lim_{h \downarrow 0} F_X(-h) = 0 \neq F_X(0).$$

Man kann dies in Abb. 2.2 sehen, beobachte dabei den Sprung von $F_X(x)$ an der Stelle $x = 0$.

Im Gegensatz dazu ist die Verteilungsfunktion der gleichverteilten Zufallsvariablen, die in Abb. 2.2 dargestellt ist, auf $\mathbb{R}$ stetig, insbesondere ist sie in jedem Punkt linksstetig: für eine gleichförmige Zufallsvariable U gilt

$$\forall a \in \mathbb{R} \quad \lim_{h \downarrow 0} F_U(a - h) = F_U(a).$$

Der folgende Satz gibt eine Interpretation des linksseitigen Grenzwertes

$$F(a-) := \lim_{h \downarrow 0} F(a - h)$$

im Punkt a für allgemeine Verteilungsfunktionen.

Satz 3.1 (Wahrscheinlichkeit eines Punktes). *Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit Verteilungsfunktion F. Für jedes a in $\mathbb{R}$ gilt*

$$\mathbb{P}[X = a] = F(a) - F(a-)$$

Wir verzichten auf den Beweis, der mit Hilfe der grundlegenden Identität des Satzes 2.3 und der Stetigkeitseigenschaft von Wahrscheinlichkeitsmaßen (Prop. 1.11) leicht zu führen ist. Wir wollen uns vielmehr der Interpretation dieses Satzes widmen.

Sei $a \in \mathbb{R}$ fixiert.

➜ Wenn F in einem Punkt $a \in \mathbb{R}$ nicht stetig ist, dann ist die "Sprunghöhe" $F(a) - F(a-)$ gleich der Wahrscheinlichkeit, dass $X = a$.

➜ Falls F stetig in einem Punkt $a \in \mathbb{R}$ ist, dann gilt $\mathbb{P}[X = a] = 0$.

2 Fast sichere Ereignisse

Ein wichtiger Begriff bei der Arbeit mit Zufallsvariablen ist der Begriff des fast sicheren Eintretens eines Ereignisses.

Definition 3.2. *Sei $A \in \mathcal{F}$ ein Ereignis. Wir sagen A tritt* ***fast sicher (f.s.)*** *ein, falls*

$$\mathbb{P}[A] = 1.$$

Bemerkung 3.3. *Wir erweitern gerade diese Notation auf allgemeinere Mengen $A \subset \Omega$ (nicht zwangsweise ein Ereignis): Wir sagen dann, dass A* ***fast sicher*** *eintritt, falls ein Ereignis $A' \in \mathcal{F}$ existiert, sodass $A' \subset A$ und $\mathbb{P}[A'] = 1$.*

Zusammengefasst: Etwas tritt f.s. ein, falls es mit Wahrscheinlichkeit 1 stattfindet. Ein klassisches Beispiel ist die folgende Aussage für die Grössenrelation einzelner oder zweier Zufallsvariablen X, Y. Wir schreiben dann gerade

$$X \leq Y \quad \text{f.s.}$$

falls $\mathbb{P}[X \leq Y] = 1$, und

$$X \leq a \quad \text{f.s.}$$

falls $\mathbb{P}[X \leq a] = 1$.

3 Diskrete Zufallsvariablen

Definition 3.4 (Diskrete Zufallsvariable)**.** *Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heisst* ***diskret*** *falls eine endliche oder abzählbare Menge $W \subset \mathbb{R}$ existiert, sodass*

$$\mathbb{P}[X \in W] = 1.$$

"Die Werte von X liegen in W fast sicher."

Bemerkung 3.5. *Wenn der Grundraum Ω endlich oder abzählbar ist, dann ist jede Zufallsvariable $X : \Omega \to \mathbb{R}$ diskret. In der Tat ist das Bild $X(\Omega) = \{x \in \mathbb{R} : \exists \omega \in \Omega \ X(\omega) = x\}$ endlich oder abzählbar und wir haben $\mathbb{P}[X \in W] = 1$, mit $W = X(\Omega)$.*

Definition 3.6. *Sei X eine diskrete Zufallsvariable mit Werten in einer endlichen oder abzählbaren Menge $W \subset \mathbb{R}$. Die Zahlenfolge $(p(x))_{x \in W}$ definiert durch*

$$\forall x \in W \quad p(x) := \mathbb{P}[X = x]$$

heisst ***Verteilung von*** *X.*

Satz 3.7. *Die Verteilung* $(p(x))_{x\in W}$ *einer diskreten Zufallsvariablen erfüllt*

$$\sum_{x\in W} p(x) = 1.$$

Beweis. Wir haben

$$\{X \in W\} = \bigcup_{x\in W} \{X = x\}.$$

Da die Vereinigung disjunkt ist und die Menge W höchstens abzählbar ist, haben wir

$$1 = \mathbb{P}[X \in W] = \mathbb{P}[\bigcup_{x\in W} \{X = x\}] = \sum_{x\in W} \mathbb{P}[X = x] = \sum_{x\in W} p(x).$$

□

Wir geben 3 Beispiele für diskrete Zufallsvariablen, die auf $(\Omega, \mathcal{F}, \mathbb{P})$, dem Laplace-Modell auf $\Omega = \{1, 2, 3, 4, 5, 6\}$ definiert sind.

Example 1: *Wert des Würfelauges*

Man betrachte die Zufallsvariable $X : \Omega \to \mathbb{R}$, definiert durch

$$\forall \omega \in \Omega \quad X(\omega) := \omega$$

(X steht für den Wert des Würfels). Dann nimmt X folgende Werte in

$$W = \{1, 2, 3, 4, 5, 6\}$$

fast sicher an. X ist somit eine diskrete Zufallsvariable und ihre Verteilung ist gegeben durch

$$\forall x \in W \quad p(x) = \mathbb{P}[X = x] = \frac{1}{6}.$$

Example 2: *Wetten auf ein Würfelauge*

Betrachte folgende Zufallsvariable

$$\forall \omega \in \Omega \quad X(\omega) := \begin{cases} -1 & \text{falls } \omega = 1, 2, 3, \\ 0 & \text{falls } \omega = 4, \\ 2 & \text{falls } \omega = 5, 6. \end{cases}$$

vergleiche dies mit Beispiel 1 auf Seite 22. Dann nimmt X folgende Werte

$$W = \{-1, 0, 2\}$$

fast sicher an. Die Verteilung ist gegeben durch

$$p(-1) = \frac{1}{2}, \quad p(0) = \frac{1}{6}, \quad p(2) = \frac{1}{3}.$$

Example 3: *Vielfaches der Zahl 3*

Betrachte die folgende Zufallsvariable,

$$\forall \omega \in \Omega \quad X(\omega) := \begin{cases} 1 & \text{falls } \omega \in \{3,6\}, \\ 0 & \text{sonst.}. \end{cases}$$

(X ist eine Indikatorfunktion auf dem Ergebnis eines Vielfachen der 3). Dann nimmt X die Werte

$$W = \{0,1\}$$

fast sicher an und die Verteilung ist gegeben durch

$$p(0) = \frac{2}{3}, \quad p(1) = \frac{1}{3}.$$

Nach Definition 3.4 ist X eine Bernoulli-Zufallsvariable mit dem Parameter 1/3.

Bemerkung 3.8. *Umgekehrt, wenn wir eine Folge von Zahlen $(p(x))_{x\in W}$ mit Werten in $[0,1]$ gegeben haben, sodass*

$$\sum_{x\in W} p(x) = 1, \tag{3.1}$$

dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Zufallsvariable X mit zugehöriger Verteilung $(p(x))_{x\in W}$. Dies gilt nach dem Existenzsatz 2.16 in Kapitel 2. Diese Beobachtung ist in der Praxis wichtig, denn sie erlaubt uns zu schreiben:

"Sei X eine diskrete Zufallsvariable mit Verteilung $(p(x))_{x\in W}$."

Verteilung p vs. Verteilungsfunktion F_X

Von der Verteilung p zur Verteilungsfunktion F_X

Satz 3.9. *Sei X eine diskrete Zufallsvariable, dessen Werte in einer endlichen oder abzählbaren Menge W liegen, und deren Verteilung p ist. Dann ist die Verteilungsfunktion von X gegeben durch*

$$\boxed{\forall x \in \mathbb{R} \quad F_X(x) = \sum_{\substack{y\le x\\ y\in W}} p(y)} \tag{3.2}$$

Beweis. Für jedes $x \in \mathbb{R}$ erhalten wir

$$\mathbb{P}[X \le x] = \mathbb{P}[X \in (-\infty, x] \cap W]] + \underbrace{\mathbb{P}[X \in (-\infty, x] \cap W^c]}_{\le \mathbb{P}[X\in W^c]=0} = \mathbb{P}[\bigcup_{\substack{y\le x\\ y\in W}} \{X = y\}] = \sum_{\substack{y\le x\\ y\in W}} \mathbb{P}[X = y].$$

□

Von der Verteilungsfunktion F_X zur Verteilung p Sei X eine diskrete Zufallsvariable. Gleichung (3.2) drückt die Verteilungsfunktion F_X in Bezug auf p als eine stückweise konstante Funktion aus. Umgekehrt ist eine Zufallsvariable mit einer stückweise konstanten Verteilungsfunktion F diskret und W und p sind gegeben durch

$$W = \{\text{Position des Sprungs von } F_X\},$$
$$p(x) = \text{``Höhe des Sprungs'' im Punkt } x \in W.$$

4 Beispiele diskreter Zufallsvariablen

Bernoulli Verteilung

Die einfachste (nicht konstante) Zufallsvariable ist die Bernoulli-Zufallsvariable. Sie wurde bereits im vorherigen Kapitel definiert. Wir erinnern hier an ihre Definition.

Definition 3.10 (Bernoulli Verteilung). *Es sei $0 \le p \le 1$. Eine Zufallsvariable X heisst **Bernoulli Zufallsvariable mit Parameter** p, wenn sie Werte in $W = \{0,1\}$ annimmt und folgendes gilt*

$$\mathbb{P}[X = 0] = 1 - p \quad \textit{und} \quad \mathbb{P}[X = 1] = p.$$

In diesem Fall schreiben wir $X \sim \mathrm{Ber}(p)$.

Binomialverteilung

Ein weiteres grundlegendes Beispiel ist die Binomialverteilung. Sie modelliert die Anzahl der Erfolge beim Wiederholen von Bernoulli-Experimenten.

Definition 3.11 (Binomialverteilung). *Sei $0 \le p \le 1$, sei $n \in \mathbb{N}$. Eine Zufallsvariable X heisst **binomiale Zufallsvariable mit Parametern** n **und** p, wenn sie Werte in $W = \{0,\dots,n\}$ annimmt und folgendes gilt*

$$\boxed{\forall k \in \{0,\dots,n\} \quad \mathbb{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}}.$$

Wir schreiben dann $X \sim \mathrm{Bin}(n,p)$.

Bemerkung 3.12. *Wenn wir $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ definieren, haben wir*

$$\sum_{k=0}^{n} p(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1,$$

daher ist die Gleichung (3.1) *erfüllt. Dies garantiert die Existenz von binomialverteilten Zufallsvariablen.*

Satz 3.13 (Summe von unabhängigen Bernoulli und Binomial Z.V.)**.** *Sei* $0 \le p \le 1$*, sei* $n \in \mathbb{N}$*. Seien* $X_1, \dots, X_n$ *unabhängige Bernoulli Z.V. mit Parameter* p*. Dann ist*

$$S_n := X_1 + \cdots + X_n$$

eine binomialverteilte Z.V. mit Parametern n *und* p*.*

Beweis. Man kann leicht überprüfen, dass S_n eine Zufallsvariable ist, welche Werte in $\{0, \dots, n\}$ annimmt. Außerdem gilt für jedes $k \in \{0, \dots, n\}$

$$\{S_n = k\} = \bigcup_{\substack{x_1,\dots,x_n \in \{0,1\} \\ x_1+\cdots+x_n=k}} \{X_1 = x_1, \cdots, X_n = x_n\}.$$

Da die Elemente der Vereinigung disjunkt sind, gilt

$$\begin{aligned}
\mathbb{P}[S_n = k] &= \sum_{\substack{x_1,\dots,x_n \in \{0,1\} \\ x_1+\cdots+x_n=k}} \mathbb{P}[X_1 = x_1, \cdots, X_n = x_n] \\
&= \sum_{\substack{x_1,\dots,x_n \in \{0,1\} \\ x_1+\cdots+x_n=k}} \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n] \\
&= \sum_{\substack{x_1,\dots,x_n \in \{0,1\} \\ x_1+\cdots+x_n=k}} p^k (1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

□

Bemerkung 3.14. *Insbesondere ist die Verteilung* $\mathrm{Bin}(1, p)$ *gerade* $\mathrm{Ber}(p)$ *verteilt. Es sei noch folgendes anzumerken: Falls* $X \sim \mathrm{Bin}(m, p)$*,* $Y \sim \mathrm{Bin}(n, p)$ *und* X, Y *unabhängig sind, dann ist* $X + Y \sim \mathrm{Bin}(m + n, p)$ *verteilt.*

Geometrische Verteilung

Definition 3.15 (Geometrische Verteilung)**.** *Es sei* $0 < p \le 1$*. Eine Zufallsvariable* X *heisst* ***geometrische Zufallsvariable mit Parameter*** p*, falls sie Werte in* $W = \mathbb{N} \backslash \{0\}$ *annimmt und folgendes gilt*

$$\boxed{\forall k \in \mathbb{N}\backslash\{0\} \quad \mathbb{P}[X = k] = (1-p)^{k-1} \cdot p}.$$

Wir schreiben dann $X \sim \mathrm{Geom}(p)$*.*

Bemerkung 3.16. *Für* $p = 1$ *und* $k = 1$ *erscheint in der obigen Gleichung ein Term* 0^0*, wir verwenden die Konvention* $0^0 = 1$ *und damit gilt* $\mathbb{P}[X = 1] = p$*.*

Bemerkung 3.17. *Falls wir $p(k) = (1-p)^{k-1}\,p$ definieren, haben wir*

$$\sum_{k=1}^{\infty} p(k) = p \sum_{k=1}^{\infty} (1-p)^{k-1} = p \cdot \frac{1}{p} = 1,$$

somit ist die Gleichung (3.1) *erfüllt. Dies garantiert die Existenz von geometrischen Zufallsvariablen.*

Die geometrische Zufallsvariable modelliert den ersten Erfolg in einer unendlichen Folge von Bernoulli-Experimenten mit Parameter p. Dies wird durch den folgenden Satz formalisiert.

Satz 3.18. *Sei $X_1, X_2, \ldots$ eine Folge von unendlich vielen unabhängigen Bernoulli-Z.V. mit Parameter p. Dann ist*

$$T := \min\{n \geq 1 \,:\, X_n = 1\}$$

eine geometrisch verteilte Zufallsvariable mit Parameter p.

Bemerkung 3.19. *Falls wir sagen, dass T eine geometrische Zufallsvariable ist, müssen wir folgendes präzisieren: Tatsächlich kann die Zufallsvariable T den Wert $+\infty$ annehmen, wenn alle Zufallsvariablen X_i gleich* 0 *sind. Dies ist jedoch kein Problem für den Bewei des Satzes. Man kann leicht überprüfen, dass $\mathbb{P}[T = \infty] = 0$ gilt.*

Beweis. Wir haben $T = k$, wenn die ersten $k-1$ Versuche fehlschlagen und der k-te Versuch ein Erfolg ist. Formal gesehen, haben wir

$$\{T = k\} = \{X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1\}.$$

Somit gilt per Unabhängigkeit

$$\begin{aligned} \mathbb{P}[T = k] &= \mathbb{P}[X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1] \\ &= \mathbb{P}[X_1 = 0] \cdots \mathbb{P}[X_{k-1} = 0]\ \mathbb{P}[X_k = 1] \\ &= (1-p)^{k-1} p. \end{aligned}$$

□

Der vorhergehende Satz gibt uns eine einfache Möglichkeit, die Definition der geometrischen Z.V. zu merken. Zudem gibt uns der Satz einige einfache Formeln, die mit der geometrischen Verteilung zusammenhängen. Sei zum Beispiel T eine geometrische Verteilung mit Parameter p. Dann ist $T > n$, wenn die ersten n Bernoulli-Experimente fehlschlagen. Daher gilt

$$\mathbb{P}[T > n] = (1-p)^n. \tag{3.3}$$

Außerdem gibt sie der Gleichung (3.4) im folgenden Satz eine wichtige Interpretation: Wenn wir auf einen ersten Erfolg in einer Folge von Experimenten warten und wissen, dass die ersten n Schritte Fehlschläge waren, dann ist die verbleibende Wartezeit wiederum eine geometrische Zufallsvariable mit dem Parameter p.

Satz 3.20 (Gedächnislosigkeit der Geometrischen Verteilung)**.** *Sei $T \sim Geom(p)$ für $0 < p < 1$. Dann gilt*

$$\forall n \geq 0 \quad \forall k \geq 1 \qquad \mathbb{P}[T \geq n+k \mid T > n] = \mathbb{P}[T \geq k]. \tag{3.4}$$

Beweis. Dies folgt direkt aus der folgenden Gleichung (3.3). □

Poisson Verteilung

Definition 3.21. *Sei $\lambda > 0$ eine positive reelle Zahl. Eine Zufallsvariable X heisst* ***Poisson-Zufallsvariable mit Parameter*** λ*, wenn sie Werte in $W = \mathbb{N}$ annimmt und folgendes gilt*

$$\boxed{\forall k \in \mathbb{N} \quad \mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}}.$$

Wir schreiben dann $X \sim \text{Poisson}(\lambda)$.

Bemerkung 3.22. *Alternativ definieren wir $p(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, haben wir*

$$\sum_{k=0}^{\infty} p(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1,$$

und somit ist die Gleichung (3.1) *erfüllt. Dies garantiert die Existenz der Poisson-Zufallsvariable.*

Die Poisson-Verteilung erscheint natürlich als eine Annäherung der Binomialverteilung, wenn der Parameter n groß und der Parameter p klein ist, wie im folgenden Satz formell festgestellt wird.

Satz 3.23 (Poisson-Approximation der Binomialverteilung)**.** *Sei $\lambda > 0$. Für jedes $n \geq 1$ seien $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$ Zufallsvariablen. Dann gilt*

$$\forall k \in \mathbb{N} \quad \lim_{n \to \infty} \mathbb{P}[X_n = k] = \mathbb{P}[N = k], \tag{3.5}$$

wobei N eine Poisson Zufallsvariable mit Parameter λ.

Bemerkung 3.24. *Die Konvergenz* (3.5) *wird Konvergenz in Verteilung genannt. Intuitiv besagt sie, dass X_n und N sehr ähnliche wahrscheinlichkeitstheoretische Eigenschaften für große n haben.*

Beweis. Man fixiere $k \in \mathbb{N}$, für jedes $n \geq 1$ haben wir

$$\begin{aligned}
\mathbb{P}[X_n = k] &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n \cdot (n-1) \cdots (n-k+1)}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{-k}}_{\xrightarrow[n\to\infty]{} 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\xrightarrow[n\to\infty]{} e^{-\lambda}},
\end{aligned}$$

Dies erfüllt das Gewünschte. □

Dieser Näherungswert kann in der Praxis nützlich sein. Betrachten wir zum Beispiel eine einzelne Seite der "Neuen Zürcher Zeitung" mit, sagen wir, $n = 10^4$ Zeichen, und nehmen wir an, dass die Druckmachine etwa 1/1000 der Zeichen falsch setzt. Mit anderen Worten, jedes Zeichen hat eine Wahrscheinlichkeit $p = 10/n$, falsch gesetzt zu werden. Die Anzahl M der Fehler auf der Seite entspricht einer binomialen Zufallsvariablen mit den Parametern n und $p = 10/n$. Durch die Poisson-Approximation ergibt sich also zum Beispiel

$$\mathbb{P}[M = 5] \simeq \frac{10^5}{5!}e^{-10} \simeq 0,0378.$$

5 Stetige Verteilungen

Definition 3.25 (Stetig verteilte Zufallsvariablen)**.** *Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heisst **stetig**, wenn ihre Verteilungsfunktion F_X wie folgt geschrieben werden kann*

$$F_X(a) = \int_{-\infty}^{a} f(x)dx \quad \text{für alle } a \text{ in } \mathbb{R}. \tag{3.6}$$

*wobei $f : \mathbb{R} \to \mathbb{R}_+$ eine nicht-negative Funktion ist. Wir nennen dann f **Dichte** von X.*

Intuition: $f(x)dx$ ist die Wahrscheinlichkeit, dass X Werte in $[x, x + dx]$ annimmt.

Um die Terminologie "stetig" zu verstehen, beachte , dass die Formel (3.6) impliziert, dass F_X eine stetige Funktion ist. Insbesondere erfüllt nach Proposition 3.1 die Z.V. X folgende Bedingung

$$\forall x \in \mathbb{R} \quad \mathbb{P}[X = x] = 0.$$

Dichte f vs. Verteilungsfunktion F_X

Von f zu F_X Sei X eine stetige Z.V. mit Dichte f. Per Definition, können wir F_X berechnen mittels folgendem Integral

$$F_X(x) = \int_{-\infty}^{x} f(y)dy.$$

Von F_X zu f Da wir von f zu F_X mittels Integration gelangen, ist es naheliegend zu erwarten, dass die umgekehrte Operation eine Ableitung ist. Dies ist im Allgemeinen der Fall, sofern F_X genug Regularität aufweist.

Das folgende Theorem wird bei der Berechnung von Dichten nützlich sein.

Theorem 3.26. *Sei X eine Zufallsvariable. Die Verteilungsfunktion F_X sei stetig und stückweise $\mathcal{C}^1$, d.h. es gibt $x_0 = -\infty < x_1 < \cdots < x_{n-1} < x_n = +\infty$, sodass F_X auf jedem*

Intervall (x_i, x_{i+1}) *Element von* $\mathcal{C}^1$ *ist. Dann ist* X *eine stetige Zufallsvariable und die Dichte* f *kann konstruiert werden, indem man folgendes festlegt*

$$\forall x \in (x_i, x_{i+1}) \quad f(x) = F_X'(x)$$

mit beliebigen Werten in $x_1, \dots, x_{n-1}$.

Beweis. Wir schreiben stets $F = F_X$. Es sei $0 \le i < n$. Falls $x_i < a < b < x_{i+1}$, impliziert der Fundamentalsatz der Analysis, dass

$$F(b) - F(a) = \int_a^b F'(y)dy = \int_a^b f(y)dy$$

Nun sei $x \in \mathbb{R}$ und wähle i, sodass $x \in [x_i, x_{i+1})$. Unter der Konvention $F(x_0) = 0$, können wir $F(x)$ als Teleskopsumme schreiben

$$F(x) = F(x) - F(x_0) = (F(x) - F(x_i)) + \cdots + (F(x_1) - F(x_0)). \tag{3.7}$$

Anwenden der Stetigkeit von F liefert das Folgende

$$(F(x) - F(x_i)) = \lim_{a \downarrow x_i} (F(x) - F(a)) = \lim_{a \downarrow x_i} \int_a^x f(y)dy = \int_{x_i}^x f(y)dy.$$

Analog gilt

$$F(x_i) - F(x_{i-1}) = \int_{x_{i-1}}^{x_i} f(y)dy.$$

Einsetzten der beiden Identitäten in (3.7), liefert das Gewünschte

$$\begin{aligned} F(x) &= \int_{x_i}^x f(y)dy + \int_{x_{i-1}}^{x_i} f(y)dy + \cdots + \int_{x_0}^{x_1} f(y)dy \\ &= \int_{-\infty}^x f(y)dy. \end{aligned}$$

□

6 Beispiele stetiger Zufallsvariablen

Gleichverteilung

Definition 3.27 (Gleichverteilung auf $[a,b]$, $a < b$.)**.** *Eine stetige Zufallsvariable X heisst* ***gleichverteilt auf*** $[a,b]$ *falls ihre Dichte gegeben ist durch*

$$f_{a,b}(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b], \\ 0 & x \notin [a,b]. \end{cases}$$

Wir schreiben dann stets $X \sim \mathcal{U}([a,b])$.

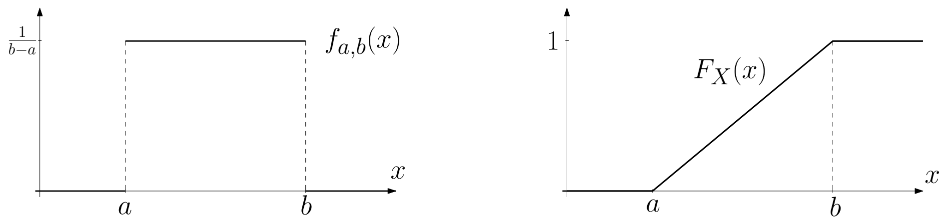


Abbildung 3.1: Dichte und Verteilungsfunktion einer gleichverteilten Zufallsvariable auf $[a,b]$.

Intuition: X modelliert das Auswählen eines Punktes in $[a,b]$ unter gleich verteilten Chancen.

Eigenschaften einer gleichverteilten Zufallsvariable X auf $[a,b]$:

- Die Wahrscheinlichkeit in ein Intervall $[c,c+\ell] \subset [a,b]$ zu fallen ist lediglich abhängig von dessen Länge ℓ:
$$\mathbb{P}[X \in [c, c+\ell]] = \frac{\ell}{b-a}.$$
- Die Verteilungsfunktion X ist gegeben durch

$$F_X(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \le x \le b, \\ 1 & x > b. \end{cases}$$

Beweis. □

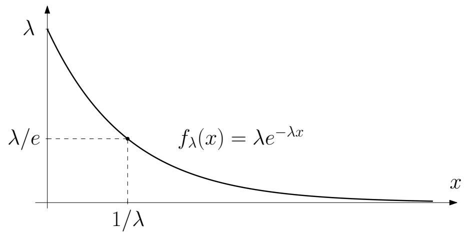
Exponentialverteilung

Die Exponentialverteilung ist das stetige Pendant zur diskreten geometrischen Verteilung.

Definition 3.28 (Exponentialverteilung mit Parameter $\lambda > 0$). *Eine stetige Zufallsvariable T heisst **exponentialverteilt mit Parameter** $\lambda > 0$ falls ihre Dichte gegeben ist durch*

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Wir schreiben dann stets $T \sim \exp(\lambda)$.



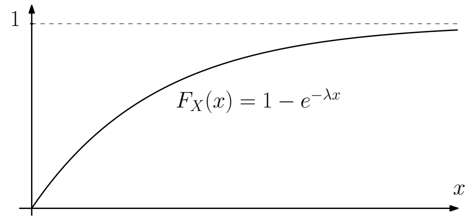


Abbildung 3.2: Dichte und Verteilungsfunktion einer exponentialverteilten Zufallsvariable mit Parameter λ.

Intuition/Anwendung: T modelliert häufig die Lebensdauer oder Wartezeit eines allgemeines Ereignisses. Beispiele dafür sind die folgenden Ereignisse: Die Wartezeit bis zum ersten Kunden eines neu eröffneten Ladens. Auftreten eines technischer Defekts eines Smartphones nach Kauf.

Eigenschaften einer exponentialverteilten Zufallsvariable T mit Parameter λ.

- Die Wahrscheinlichkeit des Wartens ist exponentiell klein:
$$\forall t \geq 0 \quad \boxed{\mathbb{P}[T > t] = e^{-\lambda t}}.$$
- T besitzt die Eigenschaft der Gedächnislosigkeit
$$\forall t, s \geq 0 \quad \mathbb{P}[T > t + s | T > t] = [T > s].$$

Die erste Eigenschaft folgt direkt aus der Definition

$$\mathbb{P}[T \geq t] = \int_t^\infty \lambda e^{-\lambda x} dx = e^{-\lambda t}.$$

Die zweite Eigenschaft folgt direkt aus dem Anwenden der bedingten Wahrscheinlichkeit

$$\mathbb{P}[T > t + s | T > t] = \frac{\mathbb{P}[T > t + s]}{\mathbb{P}[T > t]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s}.$$

Normalverteilung

Definition 3.29. *Eine stetige Zufallsvariable X heisst* ***normal verteilt mit Parametern*** m ***und*** $\sigma^2 > 0$ *falls ihre Dichte gegeben ist durch*

$$f_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Wir schreiben dann stets $X \sim \mathcal{N}(m, \sigma^2)$.

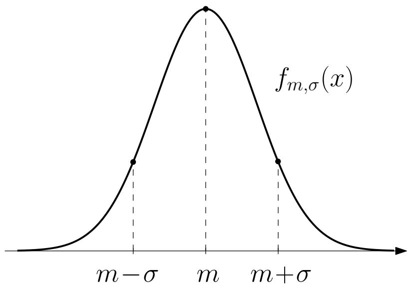


Abbildung 3.3: Dichte einer normalverteilten Zufallsvariable mit Parameter m und σ^2.

Intuition/Anwendung: Die Normalverteilung taucht in vielen Anwendungen auf. Man stelle sich zum Beispiel vor, dass man eine physikalische Größe misst: Der reale Wert ist m und der gemessene Wert wird im Allgemeinen sehr gut durch eine normale Zufallsvariable X mit den Parametern m und σ modelliert. Die Größe σ, die die Schwankungen von X darstellt, kann in diesem Zusammenhang auch als Qualität der Messung interpretiert werden. Ein kleines σ entspricht kleinen Schwankungen im Messen von X. Daraus kann man schliessen, dass der Messwert von X nahe bei m liegt. Im Gegensatz dazu entspricht ein großes σ großen Schwankungen und somit kann die Messung als ungenau interpretiert werden. Wir werden später eine mathematische Begründung sehen, warum die normale Zufallsvariable an vielen Stellen der Wahrscheinlichkeitstheorie präsent ist.

Eigenschaften der Normalverteilung

- Seien $X_1, \ldots, X_n$ unabhängige normalverteilte Zufallsvariablen mit Parametern $(m_1, \sigma_1^2), \ldots, (m_n, \sigma_n^2)$, dann ist

  $$Z = m_0 + \lambda_1 X_1 + \ldots + \lambda_n X_n$$

  eine normalverteilte Zufallsvariable mit Parametern $m = m_0 + \lambda_1 m_1 + \cdots + \lambda_n m_n$ und $\sigma^2 = \lambda_1^2 \sigma_1^2 + \cdots + \lambda_n^2 \sigma_n^2$.

- Wir sprechen im Fall von $X \sim \mathcal{N}(0,1)$, gerade von einer **standardnormalverteilten Zufallsvariable**). Man merke sich dann folgende Beziehung

$$Z = m + \sigma \cdot X,$$

wobei X eine normalverteilte Zufallsvariable mit Parametern m und σ^2 ist.

- Falls X normalverteilt mit Parametern m und σ^2 ist, dann liegt die 'meiste' Wahrscheinlichkeitsmasse der Z.V. im Intervall $[m - 3\sigma, m + 3\sigma]$. Präzise gilt gerade

$$\mathbb{P}[|X - m| \geq 3\sigma] \leq 0.0027.$$

Auf den ersten Blick mag es überraschen, dass die rechte Seite $(0,0027)$ nicht von den Parametern σ und m abhängt... Das erklärt sich wie folgt: Betrachte die Zufallsvariable $Z = \frac{X-m}{\sigma}$. Die erste Eigenschaft einer normalverteilten Z.V. impliziert, dass $Z = \frac{1}{\sigma}X - \frac{m}{\sigma}$ eine standardnormalverteilte Zufallsvariable ist. Die linke Seite kann dann umgeschrieben werden als

$$\mathbb{P}[|X - m| \geq 3\sigma] = \mathbb{P}[|\frac{X - m}{\sigma}| \geq 3] = \mathbb{P}[|Z| \geq 3]$$

Die Wahrscheinlichkeit $\mathbb{P}[|Z| \geq 3] \leq 0.0027$ kann direkt aus Tabellen zur standardnormalverteilung nachgeschaut werden.

Kapitel 4

Der Erwartungswert

Ziele

- Definition des Erwartungswerts und Intuition.
- Rechnen mit Zufallsvariablen (Summe/Produkte von Zufallsvariablen).
- Definition der Varianz und Intuition.
- Ungleichungen, Beziehungen zwischen Erwartungswert und Wahrscheinlichkeit von Ereignissen.

Rahmenbedingung In diesem Kapitel fixieren wir stets einen W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$.

In diesem Kurs konzentrieren wir uns auf diskrete und kontinuierliche Zufallsvariablen und definieren den Erwartungswert für diese beiden verschiedenen Arten von Zufallsvariablen mit zwei verschiedenen Formeln. Es gibt eine einheitliche Theorie (basierend auf der Masstheorie) des Erwartungswerts, die den Erwartungswert für allgemeine Zufallsvariablen definiert. In dieser Vorlesung werden wir uns auf die wichtigen Ergebnisse dieser Theorie und ihre Anwendungen konzentrieren, ohne die allgemeine Theorie zu beweisen.

1 Der allgemeine Erwartungswert

Definition 4.1. *Sei $X : \Omega \to \mathbb{R}_+$ eine Zufallsvariable mit nicht-negativen Werten. Dann heisst*

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

der Erwartungswert von X.

Bemerkung 4.2. *Der Erwartungswert kann sowohl endliche also auch nicht endliche Werte annehmen.*

Satz 4.3. *Sei X eine nicht-negative Zufallsvariable. Dann gilt*

$$\mathbb{E}[X] \geq 0.$$

Gleichheit gilt genau dann wenn $X = 0$ fast sicher hält.

Beweis. Der Erwartungswert $\mathbb{E}[X]$ ist definiert als das Integral einer nichtnegativen Funktion $G(x) = 1 - F_X(x) \geq 0$. Folglich ist $\mathbb{E}[X] \geq 0$. Nehmen wir nun an, dass $\mathbb{E}[X] = 0$. Dies impliziert, dass $G(x) = 0$ für jedes $x > 0$ (Per Widerspruch, falls $G(x) = \alpha > 0$ für ein $x > 0$, dann $G(y) \geq G(x) = \alpha$ für alle $y \in [0, x]$ aufgrund von Monotonie. Dann gilt aber $\int_0^x G(y)dy \geq x\alpha > 0$ und somit ist $\mathbb{E}[X] \neq 0$). Anwendung der Stetigkeit von W-Massen liefert

$$\mathbb{P}[X > 0] = \lim_{x \downarrow 0} \mathbb{P}[X > x] = \lim_{x \downarrow 0} G(x) = 0.$$

Somit gilt schliesslich,

$$\mathbb{P}[X \leq 0] = 1 - \mathbb{P}[X > 0] = 1.$$

Schlussendlich gilt somit $X \geq 0$ und $X \leq 0$ fast sicher. Dies liefert aber gerade $X = 0$ fast sicher und somit ist das Gewünschte bewiesen. □

Für allgemeine Zufallsvariablen (nicht unbedingt mit konstantem Vorzeichen) definieren wir den Erwartungswert durch Zerlegung in einen positiven und einen negativen Teil. Der positive und der negative Teil von X sind die Zufallsvariablen X_-, X_+, definiert durch

$$X_+(\omega) = \begin{cases} X(\omega) & \text{falls } X(\omega) \geq 0, \\ 0 & \text{falls } X(\omega) < 0, \end{cases} \quad \text{and} \quad X_-(\omega) = \begin{cases} -X(\omega) & \text{falls } X(\omega) \leq 0, \\ 0 & \text{falls } X(\omega) > 0. \end{cases}$$

Bemerke, dass sowohl X_+ und X_- nicht-negative Zufallsvariablen sind. Zudem gilt sowohl $X = X_+ - X_-$ als auch $|X| = X_+ + X_-$.

Definition 4.4. *Sei X eine Zufallsvariable. Falls $\mathbb{E}[|X|] < \infty$, dann heisst*

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]. \tag{4.1}$$

Erwartungswert von X.

Bemerkung 4.5. $\mathbb{E}[|X|] < \infty$ *impliziert* $\mathbb{E}[X_-], \mathbb{E}[X_+] < \infty$ *(da $|X| = X_+ + X_-$). Somit ist Gleichung.* (4.1) *erst wohldefiniert.*

Für $X \geq 0$ ist der Erwartungswert von X immer definiert. Er kann endlich oder unendlich sein.

Wenn X kein konstantes Vorzeichen hat, ist der Erwartungswert von X wohldefiniert, falls $\mathbb{E}[|X|] < \infty$. Falls diese Bedingung nicht erfüllt ist, sagen wir, dass der Erwartungswert von X **undefiniert** ist.

2 Erwartungswert einer diskreten Zufallsvariable

Der Erwartungswert einer Zufallsvariablen ist ein grundlegendes Konzept, das intuitiv dem Begriff des "Durchschnitts" entspricht. Bevor wir ihn für allgemeine Zufallsvariablen definieren, wollen wir mit unserem Lieblingsbeispiel beginnen: dem Wurf eines Würfels!

Sei $X : \Omega \to E = \{1, \cdots, 6\}$ eine Z.V. mit gleichverteilten Werten in $\{1, 2, 3, 4, 5, 6\}$, d.h. $\mathbb{P}[X = x] = 1/6$ für jedes $x = 1, \ldots, 6$. Natürlich ist der Durchschnittswert gegeben durch

$$\frac{1+2+3+4+5+6}{6}.$$

in der Sprache der Wahrscheinlichkeitstheorie ist dies gerade

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \sum_{x \in E} x \cdot \mathbb{P}[X = x],$$

dies ist zugleich die allgemeine Formel des Erwartungswerts einer diskreten Zufallsvariable.

Satz 4.6. *Sei $X : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable dessen Werte in W (endlich oder abzählbar) fast sicher liegen. Sei $\phi : \mathbb{R} \to \mathbb{R}$ eine Abbildung. Dann gilt*

$$\boxed{\mathbb{E}[\phi(X)] = \sum_{x \in W} \phi(x) \cdot \mathbb{P}[X = x]},$$

solange der Erwartungswert wohldefiniert ist.

Beweis. Wir führen zuerst den Beweis für den Fall $X \geq 0 a.s.$. Dann gilt für jedes $x \in \mathbb{R}$

$$1 - F_X(x) = \sum_{y > x, y \in W} p(y) = \sum_{y \in W} \mathbf{1}_{y > x} \cdot p(y).$$

Ausnutzen der Identität des Erwartungswerts liefert

$$\mathbb{E}[X] = \int (\mathbf{1}_{y>x} \cdot p(y))\, dx = \sum_{y \in W} \left(\int_0^\infty \mathbf{1}_{y>x} \right) = \sum_{y \in W} y \cdot p(y).$$

Die Behauptung folgt somit fpr $X \geq 0 a.s.$. Der allgemeine Fall folgt aus der Zerlegung von $X = X_+ - X_-$ und wir können gerade den obigen Fall anwenden, dies liefert

$$\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-] = \sum_{y \in W} y\mathbb{P}[X_+ = y] - \sum_{y \in W} y\mathbb{P}[X_- = y].$$

Die Definitionen von X_+ und X_- implizieren $\{X = y\} = \{X_+ = y\} \cup \{X_- = -y\}$. Und somit ist der Beweis abgeschlossen. □

Example 1: *Bernoulli Z.V.*
Sei X eine Bernoulli Zufallsvariable mit Parameter p. Dann gilt

$$\boxed{\mathbb{E}[X] = p}.$$

Dies folgt direkt aus der Definition

$$\mathbb{E}[X] = 0 \cdot \mathbb{P}[X = 0] + 1 \cdot \mathbb{P}[X = 1] = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Example 2: *Wetten auf einen Würfel*
Sei die Zufallsvariable $X : \Omega \to \{-1, 0, +2\}$ gegeben durch Gleichung. (2.1) auf Seite 22. Dann gilt

$$\mathbb{E}[X] = -1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} = \frac{1}{6}.$$

Example 3: *Poisson Z.V.*
Sei X Poisson-verteilt mit Parameter $\lambda > 0$, dann gilt

$$\boxed{\mathbb{E}[X] = \lambda}.$$

Dies folgt direkt,

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = \lambda \cdot (\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}) \cdot e^{-\lambda} = \lambda.$$

Example 4: *Indikator auf einem Ereignis*
Sei $A \in \mathcal{F}$ ein Ereignis. Sei $\mathbb{1}_A$ die **Indikator Funktion** auf A, welche definiert ist durch

$$\forall \omega \in \Omega \quad \mathbb{1}_A(\omega) = \begin{cases} 0 & \text{falls } \omega \notin A, \\ 1 & \text{falls } \omega \in A. \end{cases}$$

Dann ist $\mathbb{1}_A$ eine Zufallsvariable. Die Messbarkeit folgt gerade aus

$$\{\mathbb{1}_A \leq a\} = \begin{cases} \varnothing & \text{if } a < 0, \\ A^c & \text{if } 0 \leq a < 1, \\ \Omega & \text{if } a \geq 1, \end{cases}$$

wobei $\varnothing$, A^c und Ω drei Element aus $\mathcal{F}$ sind. Dann definieren wir gerade die diskrete Z.V. $X = 1_A$. Per Definition gilt

$$\mathbb{P}[X = 0] = 1 - \mathbb{P}[A] \quad \text{und} \quad \mathbb{P}[X = 1] = \mathbb{P}[A].$$

Somit ist $\mathbb{1}_A$ eine Bernoulli Zufallsvariable mit Parameter $\mathbb{P}[A]$. Schlussendlich gilt

$$\boxed{\mathbb{E}[\mathbb{1}_A] = \mathbb{P}[A]}.$$

Satz 4.7. *Sei $X : \Omega \to \mathbb{R}$ eine diskrete Zufallsvariable mit Werten in W (endlich oder abzählbar). Für jedes $\phi : \mathbb{R} \to \mathbb{R}$ gilt*

$$\mathbb{E}[\phi(X)] = \sum_{x \in W} \phi(x)\mathbb{P}[\phi(X = x)],$$

solange die Summe wohldefiniert ist.

Beweis. Ausgelassen. □

3 Erwartungswert stetiger Zufallsvariablen

Satz 4.8. *Sei X eine stetige Zufallsvariable mit Dichte f. Dann gilt*

$$\boxed{\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx}, \tag{4.2}$$

solange das Integral wohldefiniert ist.

Beweis. Es genügt die Behauptung für $X \geq 0$ *f.s.* zu zeigen. Der allgemeine Fall folgt nach der Zerlegung $X = X_+ - X_-$. Mittels Definition der Dichte und Anwendung von ??? liefert für jedes $x \in \mathbb{R}$

$$1 - F_X(x) = \int_x^{\infty} f(y)dy = \int_{-\infty}^{\infty} \mathbf{1}_{x<y} \cdot f(y)dy.$$

Anwendung dieser Identität in die Definition des Erwartungswerts liefert

$$\mathbb{E}[X] = \int_0^{\infty} \left(\int_{-\infty}^{\infty} \mathbf{1}_{x<y} dx \right) \cdot f(y)dy = \int_{-\infty}^{\infty} y \cdot f(y)dy.$$

□

Example 1: *Gleichverteilung auf* $[a, b]$, $a < b$

$$\mathbb{E}[X] = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \cdot (\frac{1}{2} b^2 - \frac{1}{2} a^2).$$

Somit gilt,

$$\boxed{\mathbb{E}[X] = \frac{a+b}{2}.}$$

Example 2: *Die Exponentialverteilung mit* $\lambda > 0$
Durch partielle Integration erhalten wir gerade

$$\mathbb{E}[X] = \int_0^\infty x \lambda e^{-\lambda x} dx = \left[-x e^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x} dx.$$

Somit gilt,

$$\boxed{\mathbb{E}[X] = \frac{1}{\lambda}.}$$

Theorem 4.9. *Sei X eine stetige Zufallsvariable mit Dichte f. Sei $\phi : \mathbb{R} \to \mathbb{R}$ eine Abbildung, sodass $\phi(X)$ eine Zufallsvariable ist. Dann gilt*

$$\boxed{E[\phi(X)] = \int_{-\infty}^\infty \phi(x) f(x) dx}, \tag{4.3}$$

solange das Integral wohldefiniert ist).

4 Rechnen mit Zufallsvariablen

Einer der Gründe, warum der Erwartungswert ein so mächtiges Werkzeug in der Wahrscheinlichkeitstheorie ist, liegt darin, dass wir damit vernünftig rechnen können. Zum Beispiel kann man den Erwartungswert von $X + Y$ auswerten, wenn man den Erwartungswert von X und Y kennt. In diesem Abschnitt geben wir die Rechenregeln für die Grundoperationen von Zufallsvariablen an.

Linearität

Theorem 4.10 (Linearität des Erwartungswert)**.** *Seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen, sei $\lambda \in \mathbb{R}$. Falls die Erwartungswerte wohldefiniert sind gilt*

1. $\boxed{\mathbb{E}[\lambda \cdot X] = \lambda \cdot \mathbb{E}[X]}$.

2. $\boxed{\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]}$.

Bemerkung 4.11. *Die Zufallsvariablen X und Y müssen dabei nicht unabhängig sein.*

Bemerkung 4.12. *Unter Anwendung der Induktion für $n \geq 1$ ergibt sich direkt*

$$\mathbb{E}[\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_n X_n] = \lambda_1 \mathbb{E}[X_1] + \lambda_2 \mathbb{E}[X_2] + \cdots + \lambda_n \mathbb{E}[X_n],$$

für jede Z.V. $X_1, X_2, \ldots, X_n : \Omega \to E$, und für jedes $\lambda_1, \lambda_2, \cdots, \lambda_n \in \mathbb{R}$, unter der Annahme, dass die Erwartungswerte wohldefiniert sind.

Beweis. Der Beweis der Eigenschaft (i) folgt direkt aus der Definition. Der Beweis der Eigenschaft (ii) für allgemeine Zufallsvariablen gehört in die Masstheorie und wir lassen ihn daher aus. Stattdessen geben wir hier einen Beweis für einen diskreten Grundraum Ω. Die Hauptidee des allgemeinen Falls wird dadurch illustriert. Wir nehmen an, dass Ω endlich oder abzählbar ist. In diesem Fall sind die beiden Zufallsvariablen X und Y notwendigerweise diskret (siehe Bemerkung 3.5). Nehmen wir an, dass Ω endlich ist. In diesem Fall haben wir für jedes $x \in X$

$$\mathbb{P}[X = x] = \mathbb{P}[\bigcup_{\omega \in \Omega \,:\, X(\omega) = x} \{\omega\}] = \sum_{\omega \in \Omega} \mathbf{1}_{X(\omega)=x} \cdot \mathbb{P}[\omega],$$

wobei wir gerade $\mathbb{P}[\omega]$ anstelle von $\mathbb{P}[\{\omega\}]$ schreiben. Dann gilt

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x \in X(\Omega)} x \cdot \Big(\sum_{\omega \in \Omega} \mathbf{1}_{X(\omega)=x} \cdot \mathbb{P}[\omega] \Big) \\
&= \sum_{x \in X(\Omega)} \sum_{\omega \in \Omega} x \cdot 1_{X(\omega)=x} \cdot \mathbb{P}[\omega] \\
&= \sum_{\omega \in \Omega} \mathbb{P}[\omega] \cdot \underbrace{\sum_{x \in X(\Omega)} x \cdot 1_{X(\omega)=x}}_{=X(\omega)} \\
&= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}[\omega].
\end{aligned}$$

Unter Ausnutzung dieser Formel für $X + Y$ und Y, erhalten wir gerade das Gewünschte

$$\mathbb{E}[X + Y] = \sum_{\omega \in \Omega} \big(X(\omega) + Y(\omega) \big) \cdot \mathbb{P}[\omega] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}[\omega] + \sum_{\omega \in \Omega} Y(\omega) \cdot \mathbb{P}[\omega] = \mathbb{E}[X] + \mathbb{E}[Y].$$

□

Anwendung 1: Der Erwartungswert einer binomialverteilten Zufallsvariable.
Sei $n \geq 1$ und $0 \leq p \leq 1$. Sei S eine Binomialverteilte Zufallsvariable mit Parametern n und p. Was ist der Erwartungswert von S?

Per Definition erhalten wir

$$\mathbb{E}[S] = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1 - p)^{n-k}$$

wobei diese Summe sieht nicht handlich aussieht... Wir können jedoch davon ausgehen, dass S die gleiche Verteilung hat wie $S_n = X_1 + \cdots + X_n$, wobei $X_1, \ldots, X_n$ u.i.v. Bernoulli-Z.V. mit dem Parameter p sind. Anwendung der Linearität liefert

$$\mathbb{E}[S_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n].$$

Wir wissen bereits, dass $\mathbb{E}[X_i] = p$ für jedes i, gilt. Dies liefert das Gewünschte

$$\boxed{\mathbb{E}[S] = \mathbb{E}[S_n] = np.}$$

Anwendung 2: Rechnen mit Normalverteilungen mit Parametern m und σ^2

Wenn X eine Normalverteilung mit Parametern m und σ^2 ist, dann hat sie die gleiche Verteilung wie $m+\sigma \cdot Y$, wobei Y eine standardnormalverteilte Z.V. ist. Unter Verwendung der Definition 6.3, erhalten wir, dass

$$\mathbb{E}[X] = \mathbb{E}[m + \sigma \cdot Y] = m + \sigma \mathbb{E}[Y],$$

und somit genügt es gerade den Erwartungswert von Y auszurechnen. Wir schreiben $f_{0,1}$ für die Dichtefunktion von Y und erhalten

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} x \cdot f_{0,1}(x) dx = 0$$

Wobei wir gerade ausnutzen, dass $x \cdot f_{0,1}(x)$ eine ungerade Funktion ist. Wir erhalten schließlich

$$\boxed{\mathbb{E}[X] = m.}$$

Theorem 4.13. *Seien X, Y zwei Zufallsvariablen. Falls X und Y unabhängig sind, dann ist*

$$\boxed{\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].}$$

Beweis. Ausgelassen. □

5 Extremwert Formel

Satz 4.14 (Stetige Extremwertformel)**.** *Sei X eine Zufallsvariable, sodass $X \geq 0$ fast sicher gilt. Dann gilt*

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}[X \geq x] dx.$$

Beweis. Dies folgt direkt aus Eq 4.1 und der Identität

$$1 - F_X(x) = 1 - \mathbb{P}[X \leq x] = \mathbb{P}[X > x].$$

□

Anwendung: Ausrechnen des E-Werts einer exponential-verteilten Zufallsvariable. Sei T eine exponential-verteilte Zufallsvariable mit Parameter $\lambda \geq 0$. Anwendung der obigen Proposition liefert

$$\mathbb{E}[T] = \int_0^\infty \mathbb{P}[X \geq x]dx = \int_0^\infty e^{-\lambda x}dx = \frac{1}{\lambda}.$$

Satz 4.15 (Diskrete Extremwertformel)**.** *Sei X eine diskrete Zufallsvariable mit Werten in $\mathbb{N} = \{0, 1, 2, \dots\}$. Dann gilt folgende Identität*

$$\mathbb{E}[X] = \sum_{n=1}^\infty \mathbb{P}[X \geq n]. \tag{4.4}$$

Beweis. Wir schreiben gerade $p_n = \mathbb{P}[X = n]$ für die Verteilung von X. Da X nur Werte in $\mathbb{N}$, annimmt erhalten wir gerade für $n \geq 1$

$$\mathbb{P}[X \geq n] = \sum_{k \geq n} \mathbb{P}[X = k] = \sum_{k=1}^\infty \mathbb{1}_{k \geq n} \cdot p_k.$$

Austauschen mit der rechten Seite aus (4.4) und Anwendung des Satzes von Fubini (für positive Summen) tauscht gerade die Summen. Somit erhalten wir schließlich das Gewünschte

$$\begin{aligned}
\sum_{n=1}^\infty \mathbb{P}[X \geq n] &= \sum_{n=1}^\infty \Big(\sum_{k=1}^\infty \mathbb{1}_{k \geq n} \cdot p_k \Big) \\
&= \sum_{k=1}^\infty \Big(\sum_{n=1}^\infty \mathbb{1}_{k \geq n} \Big) \cdot p_k \\
&= \sum_{k=1}^\infty k \cdot p_k \\
&= \mathbb{E}[X].
\end{aligned}$$

□

Anwendung: Ausrechnen des E-Werts einer geometrischverteilten Zufallsvariable.

Sei T eine geometrisch verteilte Zufallsvariable mit Parameter $0 < p \leq 1$. Dann gilt

$$\boxed{\mathbb{E}[T] = \frac{1}{p}.}$$

Bemerke, dass T nur Werte in $\mathbb{N}$ annimmt, dann liefert der obige Satz gerade das Gewünschte

$$\mathbb{E}[T] = \sum_{n \geq 1} \mathbb{P}[T \geq n] = \sum_{n \geq 1} (1-p)^{n-1} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

6 Charakterisierung der Eigenschaften von Z.V.

Dichte von stetigen Z.V.

Für eine Zufallsvariable X mit Dichte f, können wir den Erwartungswert von X mittels Formel (4.3) bestimmen. Es sei anzumerken, dass unterschiedliche Zufallsvariablen den gleichen Erwartungswert haben kann. Nehme X als gleichverteilte Zufallsvariable auf $[-1, 1]$ und Y normalverteilt mit Parametern $m = 0$ und $\sigma^2 > 0$. Dann haben X und Y den gleichen Erwartungswert aber unterschiedliche Dichten. Somit charakterisiert der Erwartungswert **nicht** die Dichte einer Zufallsvariable.

Allerdings ist es möglich eine Charakterisierung der Dichte einer Zufallsvariable mittels stärkerem Kriterium möglich. Hierfür muss den man den Erwartungswert von X unter einer grossen Klasse von Funktionen ϕ testen. Der nachfolgende Satz beschreibt dies.

In diesem Kurs beschränken wir uns auf stückweise stetige und beschränkte Abbildungen ϕ. Dabei heisst $\phi : \mathbb{R} \to \mathbb{R}$ **stückweise** stetig, falls $a_1 < a_2 < \cdots < a_n$ existieren, sodass ϕ stetig auf jedem Intervall (a_i, a_{i+1}) ist. Eine Funktion $\phi : \mathbb{R} \to \mathbb{R}$ heisst **beschränkt** falls ein $C > 0$ existiert, sodass

$$\forall x \in \mathbb{R} \quad |\phi(x)| \leq C$$

gilt.

Satz 4.16. *Sei X eine Zufallsvariable. Sei $f : \mathbb{R} \to \mathbb{R}_+$ eine Abbildung, sodass $\int_{-\infty}^{+\infty} f(x)dx = 1$. Dann sind folgende Aussagen äquivalent*

(i) X ist stetig mit Dichte f,

(ii) Für jede Abbildung stückweise stetige, beschränkte Abbildung $\phi : \mathbb{R} \to \mathbb{R}$ gilt

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

Beweis.

$\boxed{(i) \Rightarrow (ii)}$ Dies folgt direkt mittels Theorem. 4.9.

$\boxed{(ii) \Rightarrow (i)}$ Nehme $a \in \mathbb{R}$ and setzte $\phi(x) = \mathbf{1}_{x \leq a}$. Dies liefert das Gewünschte. □

Unabhängigkeit

Die Unabhängigkeit von zwei Zufallsvariablen X und Y impliziert nach Theorem 4.13 folgende Faktorisierung"des Erwartungswerts

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

Umgekehrt impliziert obige Gleichheit **nicht** die Unabhängigkeit von X und Y (Vergleiche mit Serie 6.6). Wir brauchen also eine stärkere Bedingung. Hier stellt sich heraus, dass falls obige Gleichung auch unter **jedem Bild** von stetigen, messbaren und beschränken Funktionen gilt, dann ist dies hinreichend für die Unabhängigkeit von X und Y. Wir formulieren dies im nächsten Satz.

Theorem 4.17. *Seien X, Y zwei Zufallsvariablen. Die folgenden Aussagen sind äquivalent*

(i) X, Y *sind unabhängig,*

(ii) Für jede Abbildungen $\phi : \mathbb{R} \to \mathbb{R}, \psi : \mathbb{R} \to \mathbb{R}$ *stückweise stetig, beschränkt gilt*

$$\boxed{\mathbb{E}[\phi(X)\psi(Y)] = \mathbb{E}[\phi(X)]\mathbb{E}[\psi(Y)]}. \tag{4.5}$$

Beweis.

(i)⇒ (ii) Ausgelassen.

(ii)⇒ (i) Sei $a, b \in \mathbb{R}$. Ausnutzen der Gleichung. (4.5) auf die Funktionen $\phi_a(x) = \mathbf{1}_{x \le a}$ und $\psi_b(y) = \mathbf{1}_{y \le b}$ liefert

$$\mathbb{E}[\mathbf{1}_{X \le a, Y \le b}] = \mathbb{E}[\mathbf{1}_{X \le a}]\mathbb{E}[\mathbf{1}_{X \le b}].$$

Anwendung von $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}[A]$ zeigt, dass X und Y unabhängig sind.

□

Oben haben wir nur ein Paar (X, Y) von Zufallsvariablen betrachtet, aber die gleiche Aussage gilt ebenfalls für n Zufallsvariablen $X_1, \dots, X_n$. Der nachfolge Satz demonstriert dies.

Theorem 4.18. *Seien* $X_1, \dots, X_n$ *Zufallsvariablen. Die folgenden Aussagen sind äquivalent*

(i) $X_1, \dots, X_n$ *sind unabhängig,*

(ii) Für jede $\phi_1 : \mathbb{R} \to \mathbb{R}, \dots, \phi_n : \mathbb{R} \to \mathbb{R}$ *stückweise stetig, beschränkt gilt*

$$\boxed{\mathbb{E}[\phi_1(X_1)\cdots\phi_n(X_n)] = \mathbb{E}[\phi_1(X_1)]\cdots\mathbb{E}[\phi_n(X_n)].}$$

7 Ungleichungen

Monotonie

Satz 4.19. *Seien X, Y zwei Zufallsvariablen, sodass*

$$X \leq Y f.s.$$

gilt. Falls beide Erwartungswerte wohldefiniert sind folgt dann

$$\mathbb{E}[X] \leq \mathbb{E}[Y].\ f.s.$$

Beweis. Sei $Z = Y - X$. Per Annahme, gilt $Z \geq 0$. Dies liefert direkt $\mathbb{E}[Z] \geq 0$ (mittels Satz. 4.3). Anwendung der Linearität liefert das Gewünschte

$$\mathbb{E}[Y] - \mathbb{E}[X] = \mathbb{E}[Z] \geq 0.$$

□

Markow Ungleichung

Theorem 4.20 (Markow-Ungleichung)**.** *Sei X eine* ***nicht-negative*** *Zufallsvariable. Für jedes $a > 0$ gilt dann*

$$\boxed{\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.} \tag{4.6}$$

Beweis. Sei $a > 0$. Anwendung der Monotonie des Erwartungswert liefert das Gewünschte

$$E[X] \geq E[X \cdot \mathbb{1}_{X \geq a}] \geq E[a \cdot \mathbb{1}_{X \geq a}] = a\mathbb{P}[X \geq a].$$

□

Jensen Ungleichung

Theorem 4.21 (Jensen Ungleichung)**.** *Sei X eine Zufallsvariable. Sei $\phi : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion. Falls $\mathbb{E}[\phi(X)]$ und $\mathbb{E}[X]$ wohldefiniert sind, gilt*

$$\boxed{\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)].}$$

Die Jensen-Ungleichung hat mehrere wichtige Konsequenzen. Erstens erhalten wir durch ihre Anwendung auf $\phi(x) = |x|$ die Dreiecksungleichung. Für jede diskrete Zufallsvariable X gilt

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|].$$

Eine weitere wichtige Folge ist der Zusammenhang zwischen dem Erwartungswert von $|X|$ und dem Erwartungswert von X^2. Anwendung der Jensen Ungleichung auf die konvexe Funktion $\phi(x) = x^2$ liefert für jede diskrete Zufallsvariable

$$\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}[X^2]}. \tag{4.7}$$

8 Varianz

Definition 4.22. *Sei X eine Zufallsvariable, sodass $\mathbb{E}[X^2] < \infty$. Wir definieren die **Varianz von** X durch*

$$\boxed{\sigma_X^2 = \mathbb{E}[(X-m)^2], \quad \text{wobei } m = \mathbb{E}[X].}$$

*Die Wurzel aus σ_X^2 nennen wir gerade die **Standardabweichung von** X.*

Bemerkung 4.23. *Falls $\mathbb{E}[X^2] < \infty$, dann gilt gerade $\mathbb{E}[|X|] < \infty$ durch Gleichung.* (4.7) *und somit ist $m = \mathbb{E}[X]$ wohldefiniert.*

Die Standardabweichung ist ein Indikator für die Fluktuation von X um den Mittelwert $m = \mathbb{E}[X]$ herum. Wir veranschaulichen dies gerade anhand von zwei einfachen Beispielen.

Example 1: *Deterministische Zufallsvariablen*

Sei $a \in \mathbb{R}$. Sei X eine deterministische Zufallsvariable mit Wert a, also $X(\omega) = a$ für jedes ω. Dann gilt gerade $m = \mathbb{E}[X] = a$ und $\sigma_X^2 = \mathbb{E}[(X-m)^2] = 0$.

Example 2: *Gleichverteilung in zwei Punkten*

Seien $a < b$ zwei reelle Zahlen. Sei X eine gleichverteilte Zufallsvariable in zwei Punkten. Die Verteilung von X ist dann gerade gegeben durch $\mathbb{P}[X = a] = \mathbb{P}[X = b] = 1/2$. Somit gilt $m = \mathbb{E}[X] = (a+b)/2$ und

$$\sigma_X = \sqrt{\mathbb{E}[(X-m)^2]} = \frac{a-b}{2}.$$

Allgemein ist eine Zufallsvariable mit geringer Varianz konzentriert um ihren Erwartungswert $m = \mathbb{E}[X]$. Die Tschebyscheffsche Ungleichung formalisiert diese Beobachtung.

Satz 4.24. *Sei X eine Zufallsvarable mit $\mathbb{E}[X^2] < \infty$. Dann gilt für jedes $a \geq 0$*

$$\boxed{\mathbb{P}[|X-m| \geq a] \leq \frac{\sigma_X^2}{a^2}, \quad \text{wobei } m = \mathbb{E}[X]}.$$

Beweis. Wir setzten $Y = (X-m)^2$. Beachte gerade, dass $Y \geq 0$ und $\mathbb{E}[Y] = \mathbb{E}[(X-m)^2] = \sigma_X^2$ gilt. Zudem gilt für jedes $a \geq 0$

$$\mathbb{P}[|X-m| \geq a] = P[Y \geq a^2].$$

Anwendung der Markov-Ungleichung auf Y liefert das Gewünschte

$$\mathbb{P}[|X-m| \geq a] \leq \frac{\mathbb{E}[Y]}{a^2} \leq \frac{\sigma_X^2}{a^2}.$$

□

Satz 4.25 (Grundlegende Eigenschaften der Varianz).

1. *Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$. Dann gilt*

$$\sigma_X^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

2. *Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$ und sei $\lambda \in \mathbb{R}$. Dann gilt*

$$\sigma_{\lambda X}^2 = \lambda^2 \cdot \sigma_X^2.$$

3. *Seien $X_1, \ldots, X_n$ n-viele paarweise unabhängigen Zufallsvariablen und $S = X_1 + \cdots + X_n$. Dann gilt*

$$\sigma_S^2 = \sigma_{X_1}^2 + \cdots + \sigma_{X_n}^2.$$

Beweis.

1. Sei $m = \mathbb{E}[X]$. Ausnutzung der Linearität des Erwartungswerts liefert die erste Behauptung

$$\mathbb{E}[(X-m)^2] = [X^2 - 2mX + m^2] = \mathbb{E}[X^2] - 2m\mathbb{E}[X] + m^2 = \mathbb{E}[X^2] - m^2.$$

2. Ausnutzen der Behauptung 1 und Anwendung der Linearität des Erwartungswert liefert das Gewünschte

$$\sigma_{\lambda X}^2 = \mathbb{E}[(\lambda X)^2] - (\mathbb{E}[\lambda X])^2 = \lambda^2 \cdot \mathbb{E}[X^2] - \lambda^2 \cdot \mathbb{E}[X]^2 = \lambda^2 \cdot \sigma_X^2.$$

3. Wir schreiben gerade $m_i = \mathbb{E}[X_i]$, dann erhalten wir

$$S - \mathbb{E}[S] = \sum_{i=1}^{n} (X_i - m_i),$$

und somit gilt

$$\sigma_S^2 = \sum_{1 \leq i,j \leq n} \mathbb{E}[(X_i - m_i)(X_j - m_j)].$$

Für $i \neq j$ impliziert die Unabhängigkeit jedoch $\mathbb{E}[(X_i - m_i)(X_j - m_j)] = \mathbb{E}[(X_i - m_i)]\mathbb{E}[(X_j - m_j)] = 0$. Folglich bleiben nur die diagonalen Terme (für die $i = j$) in der Summe erhalten und wir erhalten

$$\sigma_S^2 = \sum_{i=1}^{n} \mathbb{E}[(X_i - m_i)^2] = \sum_{i=1}^{n} \sigma_{X_i}^2.$$

□

Anwendung: Sei S eine binomialverteilte Zufallsvariable mit Parametern n und p. Was ist die Varianz von S?

Wir benutzen gerade wiedermal, dass S dieselbe Verteilung wie $S_n = X_1 + \cdots + X_n$, mit $X_1, \ldots, X_n$ u.i.v. Bernoulli Zufallsvariablen mit Parameter p, hat. Dann erhalten wir

$$\begin{aligned} \sigma_S^2 = \sigma_{S_n}^2 &\overset{\text{Unabhängigkeit}}{=} \sigma_{X_1}^2 + \cdots + \sigma_{X_n}^2 \\ &\overset{\text{ident. vert.}}{=} n \cdot \sigma_{X_1}^2. \end{aligned}$$

Zudem gilt $\sigma_{X_1}^2 = \mathbb{E}[X_i^2] - p^2 = p - p^2 = p(1-p)$. Einsetzen von oben liefert das Gewünschte

$$\boxed{\sigma_S^2 = n \cdot p(1-p)}.$$

Beachte, dass wir hier einen wichtigen Aspekt der Summen von u.i.v. Zufallsvariablen entdeckt haben. Wir erhalten stets

$$\mathbb{E}[S] = n \cdot p \quad \text{and} \quad \sigma_S = \sqrt{n} \cdot \sqrt{p(1-p)}$$

Somit wächst der Erwartungswert von S wie n, während die Fluktuation von S_n dank der Annullierungen in der Summe $S_n = X_1 + \cdots + X_n$ wie $\sqrt{n}$ wächst.

9 Kovarianz

In diesem Abschnitt führen wir die Kovarianz ein. Die Kovarianz wird häufig eingesetzt um eine Messung eines Unterschieds zwischen zwei Zufallsvariablen X und Y zu messen.

Definition 4.26. *Seien X, Y zwei Zufallsvariablen mit endlichen zweiten Momenten $\mathbb{E}[X^2] < \infty$ und $\mathbb{E}[Y^2] < \infty$. Wir definieren die **Kovarianz zwischen** X **und** Y durch*

$$\boxed{\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].}$$

Bemerkung: Die endlichen zweiten Momente von X und Y ermöglichen die Wohldefiniertheit der Kovarianz von X und Y. Hierfür wende man die elementare Ungleichung $|XY| \leq \frac{1}{2}X^2 + \frac{1}{2}Y^2$ in Kombination mit Monotonie und Linearität des Erwartungswerts an, um folgende Ungleichung zu erhalten

$$\mathbb{E}[|XY|] \leq \frac{1}{2}\mathbb{E}[X^2] + \frac{1}{2}\mathbb{E}[Y^2] < \infty.$$

Analog zum Abschnitt 4, verschwindet die Kovarianz von unabhängigen Zufallsvariablen X und Y, somit gilt

$$X, Y \text{ unabhängig} \quad \Longrightarrow \quad \operatorname{Cov}(X, Y) = 0.$$

Die **Umgekehrte Implikation** ist **falsch** (Siehe Serie 6.6). Nichtsdestotrotz sehen wir in Abschnitt 6, dass wir eine Charakterisierung mittels beschränkten und stetigen Testfunktionen erhalten. Mittels Theorem 4.17, erhalten wir folgende Charakterisierung

$$X, Y \text{ Unabhängig} \quad \Longleftrightarrow \quad \forall \phi, \psi \text{ stückweise stetig, beschränkt} \quad \operatorname{Cov}(\phi(X), \psi(Y)) = 0.$$

Kapitel 5

Gemeinsame Verteilung

Ziele

- Definition der gemeinsamen Verteilung für diskrete/kontinuierliche Zufallsvariablen.
- Rechnen mit Randverteilungen und gemeinsamen Dichten
- Interpretation der Abhängigkeit/Unabhängigkeit von Zufallsvariablen.

Vorbemerkung Alle Zufallsvariablen in diesem Kapitel beziehen sich auf einen fixierten W-Raum $(\Omega, \mathcal{F}, \mathbb{P})$.

1 Gemeinsame diskrete Verteilung

1.1 Definition

Definition 5.1. *Seien $X_1, \dots, X_n$ n diskrete Zufallsvariablen, sei $W_i \subset \mathbb{R}$ endlich oder abzählbar, wobei $X_i \in W_i$ fast sicher gilt. Die gemeinsame Verteilung von $(X_1, \dots, X_n)$ ist eine Familie $p = (p(x_1, \dots, x_n))_{x_1 \in W_1, \dots, x_n \in W_n}$, wobei jedes Mitglied definiert ist durch*

$$\boxed{p(x_1, \dots, x_n) = \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]}.$$

Example:

Seien X, Y zwei unabhängige Bernoulli Zufallsvariablen mit Parameter 1/2. Die gemeinsame Verteilung von (X, Y) ist gegeben durch

$$\forall x, y \in \{0, 1\} \qquad p(x, y) = \frac{1}{4}.$$

Die gemeinsame Verteilung von (X, X) ist gegeben durch

$$\forall x, y \in \{0, 1\} \qquad p(x, y) = \begin{cases} \frac{1}{2} & x = y, \\ 0 & x \neq y. \end{cases}$$

Sei $Z = X + Y$, dann ist die gemeinsame Verteilung p von (X, Z) gegeben durch folgende Tabelle:

x	y	$p(x,y)$
0	0	1/4
0	1	1/4
0	2	0
1	0	0
1	1	1/4
1	2	1/4

Satz 5.2. *Eine gemeinsame Verteilung von Zufallsvariablen $X_1, \dots, X_n$ erfüllt*

$$\boxed{\sum_{x_1 \in W_1, \dots, x_n \in W_n} p(x_1, \dots, x_n) = 1.} \tag{5.1}$$

Beweis. Sei $A = \{X_1 \in W_1, \dots, X_n \in W_n\}$. Da A eine endlicher Schnitt von fast sicheren Ereignissen ist, gilt $\mathbb{P}[A] = 1$ (Nach Aufgabe 2.1, Serie 2). Zudem haben wir noch folgende Darstellung für A

$$A = \bigcup_{x_1 \in W_1, \dots, x_n \in W_n} \{X_1 = x_1, \dots, X_n = x_n\}.$$

Dies liefert nun das Gewünschte

$$1 = \mathbb{P}[A] = \sum_{x_1 \in W_1, \dots, x_n \in W_n} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n].$$

□

Umgekehrt gibt es bei endlichen oder abzählbaren Familien von Mengen $W_1, \dots, W_n$ und einer gegebenen Funktion $p : W_1 \times \dots \times W_n \to [0,1]$, die (5.1) erfüllt, einen Wahrscheinlichkeitsraum und diskrete Zufallsvariablen mit Verteilung p (siehe Serie).

1.2 Verteilung des Bildes

Ein Vorteil des Arbeitens mit Zufallsvariablen ist, dass wir sie als Zahlen "manipulieren" können. Seien $X_1, X_2, \dots, X_n$ n Zufallsvariablen. Dann können wir sie intuitiv als n "zufällige" Zahlen betrachten und sie mittels mathematischen Operationen manipulieren.

Der folgende Satz gibt die Verteilung von Zufallsvariablen als Bilder von diskreten Zufallsvariablen an.

Satz 5.3. *Sei $n \geq 1$ und seien $\phi : \mathbb{R}^n \to \mathbb{R}$ Abbildungen. Seien $X_1, \dots, X_n$ n diskrete Zufallsvariablen in $(\Omega, \mathcal{F}, \mathbb{P})$, welche fast sicher Werten in endlichen oder abzählbaren Mengen $W_1, \dots, W_n$ annehmen. Dann ist $Z = \phi(X_1, \dots, X_n)$ eine diskrete Zufallsvariable, welche fast sicher Werte in $W = \phi(W_1 \times \dots \times W_n)$ annimmt. Zudem ist die Verteilung von Z gegeben durch*

$$\forall z \in W \quad \mathbb{P}[Z = z] = \sum_{\substack{x_1 \in W_1, \dots, x_n \in W_n \\ \phi(x_1, \dots, x_n) = z}} \mathbb{P}[X_1 = x_1, \dots X_n = x_n].$$

Beweis. W ist als Bild von endlichen oder abzählbaren Mengen endlich oder abzählbar. Um zu zeigen, dass Z eine diskrete Zufallsvariable ist reicht es folgendes zu zeigen

- nimmt Werte W fast sicher an, und
- für jedes z in W, haben wir $\{Z = z\} \in \mathcal{F}$.

Beachte, dass die zweite Eigenschaft gerade für beliebiges $a \in \mathbb{R}$ impliziert, dass

$$\{Z \leq a\} = \bigcup_{\substack{z \in W \\ z \leq a}} \{Z = z\}$$

ebenfalls ein Ereignis ist. (Als abzählbare Vereinigung von Ereignissen).

Die erste Eigenschaft folgt aus der Inklusion $\{X_1 \in W_1, \dots, X_n \in W_n\} \subset \{Z \in W\}$. Für die zweite Eigenschaft sei $z \in W$ und beachte

$$\{Z = z\} = \bigcup_{\substack{x_1 \in W_1, \dots, x_n \in W_n \\ \phi(x_1, \dots, x_n) = z}} \{X_1 = x_1, \dots X_n = x_n\}. \tag{5.2}$$

Somit ist $\{Z = z\} \in \mathcal{F}$ (abzählbare Vereinigung von Ereignissen). Somit ist Z eine diskrete Zufallsvariable. Um die Verteilung zu bestimmen, beachte, dass die Vereinigung in (5.2) disjunkt und höchstens abzählbar ist. Dies liefert das Gewünschte

$$\mathbb{P}[Z = z] = \sum_{\substack{x_1 \in W_1, \dots, x_n \in W_n \\ \phi(x_1, \dots, x_n) = z}} \mathbb{P}[X_1 = x_1, \dots X_n = x_n].$$

□

Example:

Sei $Z = X + Y$ analog zum Abschnitt 1.1) eine Summe von zwei Zufallsvariablen mit Parameter 1/2. Anwendung des Satzes mittels $\phi(x, y) = x + y$ liefert

$$\begin{aligned}
\mathbb{P}[Z = 0] &= \sum_{\substack{x,y \in \{0,1\} \\ x+y=0}} \mathbb{P}[X = x, Y = y] = \mathbb{P}[X = 0, Y = 0] = 1/4 \\
\mathbb{P}[Z = 1] &= \sum_{\substack{x,y \in \{0,1\} \\ x+y=1}} \mathbb{P}[X = x, Y = y] = \mathbb{P}[X = 0, Y = 1] + \mathbb{P}[X = 1, Y = 0] = 1/2 \\
\mathbb{P}[Z = 2] &= \sum_{\substack{x,y \in \{0,1\} \\ x+y=2}} \mathbb{P}[X = x, Y = y] = \mathbb{P}[X = 1, Y = 1] = 1/4.
\end{aligned}$$

1.3 Randverteilung

Unter Kenntnis der Verteilung von $X_1, \dots, X_n$, kann man die Verteilung der einzelnen X_i separat ermitteln. In diesem Zusammenhang wird die Verteilung von X_i als i-te Randverteilung bezeichnet.

Satz 5.4. *Seien $X_1, \dots, X_n$ n diskrete Zufallsvariablen mit gemeinsamer Verteilung $p = (p(x_1, \dots, x_n))_{x_1 \in W_1, \dots, x_n \in W_n}$. Für jedes i gilt*

$$\boxed{\forall z \in W_i \qquad \mathbb{P}[X_i = z] = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} p(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n)}.$$

Beweis. Anwenden des Satzes 5.3 mittels $\phi(x_1, \dots, x_n) = x_i$ liefert das Gewünschte. □

Nach dem obigen Satz kann man die Randverteilung der Zufallsvariablen X und Y unter Kenntnis der gemeinsamen Verteilung p von X, Y bestimmen. Der Umkehrschluss ist nicht möglich. Die Kenntnis der Randverteilungen ist nicht ausreichend, um die gemeinsame Verteilung zu berechnen. Man nehme zwei unabhängige Zufallsvariablen X, Y. Dann haben (X, Y) und (X, X) die gleichen Randverteilung (beide Bernoulli (1/2)), allerdings sind die gemeinsamen Verteilungen unterschiedlich.

Der Begriff der Randverteilungen kann daher nur helfen, gemeinsame Verteilungen zu verstehen: Sie geben Fingerzeig auf die gemeinsame Verteilung von $X_1, \dots, X_n$ mittels Verteilung der einzelnen X_i. Zudem geben sie Aufschluss auf die Abhängigkeit der Zufallsvariablen untereinander.

1.4 Erwartungswert des Bildes

Satz 5.5. *Seien $X_1, \dots, X_n$ n diskrete Zufallsvariablen mit gemeinsamer Verteilung $p = (p(x_1, \dots, x_n))_{x_1 \in W_1, \dots, x_n \in W_n}$. Sei $\phi : \mathbb{R}^n \to \mathbb{R}$, dann gilt*

$$\boxed{\mathbb{E}[\phi(X_1, \dots, X_n)] = \sum_{x_1, \dots, x_n} \phi(x_1, \dots, x_n) p(x_1, \dots, x_n)},$$

solange die Summe wohldefiniert ist.

Beweis. Sei $W = \phi(W_1, \dots, W_n)$ unter Verwendung von Satz. 3.9 erhalten wir

$$\begin{aligned}\sum_{z \in F} z \cdot \mathbb{P}[Z = z] &= \sum_{z \in F} \sum_{x_1, \dots, x_n \in E} z \cdot \mathbf{1}_{\phi(x_1, \dots, x_n) = z} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] \\ &= \sum_{x_1, \dots, x_n \in E} \underbrace{\Big(\sum_{z \in F} z \cdot \mathbf{1}_{\phi(x_1, \dots, x_n) = z}\Big)}_{=\phi(x_1, \dots, x_n)} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n].\end{aligned}$$

(Die Vertauschung der Summen wird durch den Satz von Fubini gerechtfertigt, solange die Summen wohldefiniert sind). □

1.5 Unabhängigkeit

Satz 5.6. *Seien $X_1, \dots, X_n$ diskrete Zufallsvariablen mit gemeinsamer Verteilung $p = (p(x_1, \dots, x_n))_{x_1 \in W_1, \dots, x_n \in W_n}$. Die folgenden Aussagen sind äquivalent*

(i) $X_1, \dots, X_n$ sind unabhängig,

(ii) $p(x_1, \dots, x_n) = \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n]$ für jedes $x_1 \in W_1, \dots, x_n \in W_n$.

Beweis.

(i)⇒ (ii) Wähle die Abbildungen $\phi_1, \dots, \phi_n$ gegeben durch $\phi_i(z) = \mathbf{1}_{z = x_i}$. Dies liefert

$$\begin{aligned}\mathbb{P}[X = x_1, \dots, X_n = x_n] &= \mathbb{E}[\mathbf{1}_{X_1 = x_1, \dots, X_n = x_n}] \\ &= \mathbb{E}[\phi_1(X_1) \cdots \phi_n(X_n)] \\ &\overset{(i)}{=} \mathbb{E}[\phi_1(X_1)] \cdots \mathbb{E}[\phi_n(X_n)] \\ &= \mathbb{P}[X_1 = x_1] \cdots \mathbb{P}[X_n = x_n]\end{aligned}$$

(ii)⇒ (i) Sei $\phi_1 : W_1 \to \mathbb{R}, \dots, \phi_n : W_n \to \mathbb{R}$. Anwendung des Satzes. 5.3 mittels $\phi(x_1, \dots, x_n) =$

$\phi_1(x_1)\cdots\phi_n(x_n)$ liefert

$$\begin{aligned}
\mathbb{E}[\phi_1(X_1)\cdots\phi_n(X_n)] &= \sum_{x_1,\dots,x_n} \phi_1(x_1)\cdots\phi_n(x_n)\cdot p(x_1,\dots,x_n) \\
&\overset{(ii)}{=} \sum_{x_1,\dots,x_n} \phi_1(x_1)\cdots\phi_n(x_n)\cdot \mathbb{P}[X_1 = x_x]\cdots\mathbb{P}[X_n = x_n] \\
&= \Big(\sum_{x_1}\phi_1(x_1)\cdot\mathbb{P}[X = x]\Big)\cdots\Big(\sum_{x_n}\phi_n(x_n)\cdot\mathbb{P}[X_n = x_n]\Big) \\
&= \mathbb{E}[\phi_1(X_1)]\cdots\mathbb{E}[\phi_n(X_n)].
\end{aligned}$$

Anwendung des Satzes 4.18, $X_1,\cdots,X_n$ liefert die Unabhängigkeit.

□

2 Stetige Gemeinsame Verteilung

2.1 Definition

Definition 5.7. *Sei $n \geq 1$. Wir sagen, dass die Zufallsvariablen $X_1,\dots,X_n : \Omega \to \mathbb{R}$ eine **stetige gemeinsame Verteilung** besitzen, falls eine Abbildung $f : \mathbb{R}^n \to \mathbb{R}_+$ existiert, sodass*

$$\boxed{\mathbb{P}[X_1 \leq a_1,\dots,X_n \leq a_n] = \int_{-\infty}^{a_1}\cdots\int_{-\infty}^{a_n} f(x_1,\dots,x_n)dx_n\dots dx_1}$$

*für jedes $a_1,\dots,a_n \in \mathbb{R}$ gilt. Obige Abbildung f nennen wir gerade **gemeinsame Dichte von** $(X_1,\dots,X_n)$.*

Bemerkung 5.8. *In ähnlicher Weise können wir die gemeinsame Dichte $f(x_1,\dots,x_n)$ für n Zufallsvariablen definieren. In diesem Abschnitt konzentrieren wir uns auf den Fall $n = 2$, aber alle vorgestellten Definitionen/Ergebnisse gelten natürlich auch für allgemeine n.*

Satz 5.9. *Sei f die gemeinsame Dichte der Zufallsvariablen $(X_1,\dots,X_n)$. Dann gilt*

$$\boxed{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,\dots,x_n)dx_n\dots dx_1 = 1.}$$

Umgekehrt ist es auch stets möglich für eine gegebene Dichte f welche Gleichung 5.9 erfüllt einen W-Raum $(\Omega,\mathcal{F},\mathbb{P})$ und n Zufallsvariablen $X_1,\dots,X_n$ zu finden, sodass f die gemeinsame Dichte der stetigen gemeinsamen Verteilung von $X_1,\dots,X_n$ ist. Der Beweis ist ausgelassen.

Beweis. Anwendung der Stetigkeit des Wahrscheinlichkeitmasses (Proposition 1.11) liefert gerade das Gewünschte

$$\begin{aligned}1 &= \lim_{a_1,\dots,a_n\to\infty} \mathbb{P}[X_1 \le a_1,\dots,X_n \le a_n] \\ &= \lim_{a_1,\dots,a_n\to\infty} \int_{-\infty}^{a_1}\cdots\int_{-\infty}^{a_n} f(x_1\dots x_n)dx_n\dots dx_1 = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1\dots x_n)dx_n\dots dx_1 = 1.\end{aligned}$$

□

Interpretation: Nehme zum Beispiel zwei Zufallsvariablen X, Y. Intuitiv beschreibt $f(x,y)dxdy$ dabei die Wahrscheinlichkeit, dass ein Zufallspunkt (X,Y) einem Rechteck $[x, x+dx]\times[y, y+dy]$ liegt. Wie lautet die Interpretation für $n > 2$?

Example 1: *Gleichverteilung auf einem Quadrat*

Seien X und Y zwei Zufallsvariablen mit gemeinsamer stetiger Dichte $f(x,y) = \mathbf{1}_{0\le x,y\le 1}$, d.h.

$$f(x,y) = \begin{cases} 1 & (x,y)\in[0,1]^2 \\ 0 & (x,y)\notin[0,1]^2. \end{cases}$$

Example 2: *Gleichverteilung auf dem Einheitskreis*

Sei $D = \{(x,y) : x^2+y^2 \le 1\}$ der Einheitskreis zentriert um 0. Seien X und Y Zufallsvariablen mit gemeinsamer Dichte $f(x,y) = \frac{1}{\pi}\mathbf{1}_{x^2+y^2\le 1}$, d.h.

$$f(x,y) = \begin{cases} \frac{1}{\pi} & x^2+y^2 \le 1 \\ 0 & x^2+y^2 > 1. \end{cases}$$

2.2 Erwartungswert unter Abbildungen

Satz 5.10. *Sei $\phi : \mathbb{R}^n \to \mathbb{R}$ eine Abbildung. Falls $X_1,\dots,X_n$ eine gemeinsame Dichte f besitzen, dann lässt sich der Erwartungswert der Zufallsvariablen $Z = \phi(X_1,\dots,X_n)$ mittels*

$$\boxed{\mathbb{E}[\phi(X_1,\dots,X_n)] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \phi(x_1,\dots,x_n)\cdot f(x_1,\dots,x_n)dx_n\dots dx_1,} \tag{5.3}$$

berechnen (solange das Integral wohldefiniert ist).

Beweis. Ausgelassen. □

Anwendungen: Betrachten wir das Paar (X,Y) analog zum obigen ersten Beispiel. Falls wir die Funktion $\phi(x,y) = \mathbf{1}_{(x,y)\in R}$ betrachten, gilt für jedes Rechteck $R = (a,a') \times (b,b') \subseteq [0,1]^2$

$$\mathbb{P}[(X,Y) \in R] = \mathbb{E}[\phi(X,Y)] = \int_a^{a'} \int_b^{b'} dxdy = (a'-a)(b'-b) = \text{Flaeche}(R).$$

Die gemeinsame Verteilung (X,Y) repräsentiert dabei das uniforme Setzen eines Punktes auf dem Rechteck $[0,1]^2$.

Analog gilt mit (X,Y) (analog nach Beispiel 2) für jedes Rechteck $R = (a,a') \times (b,b') \subset D$

$$\mathbb{P}[(X,Y) \in R] = \frac{1}{\pi}(a'-a)(b'-b) = \frac{\text{Flaeche}(R)}{\text{Flaeche}(D)}.$$

Dabei repräsentiert (X,Y) gerade das uniforme Setzten eines Punktes auf $[0,1]^2$.

2.3 Randverteilungen

Falls X, Y eine gemeinsame Dichte $f_{X,Y}$ besitzt, dann gilt

$$\begin{aligned}\mathbb{P}[X \le a] &= \mathbb{P}[X \in [-\infty, a], Y \in [-\infty, \infty]] \\ &= \int_{-\infty}^{a} (\int_{-\infty}^{\infty} f(x,y)dy)dx.\end{aligned}$$

Somit ist X stetig mit folgender Dichte

$$\boxed{f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy.}$$

Analog ist Y stetig mit folgender Dichte

$$\boxed{f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.}$$

Wir wollen nun die Randverteilung der beiden obigen Beispielen unter gegebenen gemeinsamen Dichten ausrechnen.

Example 1: *Gleichverteilng eines Punktes auf dem Quadrat*

Unter gemeinsamer Dichte $f_{X,Y}(x,y) = \mathbf{1}_{0 \le x,y \le 1}$ hat X folgende Dichte

$$f_X(x) = \int_{0,1} \mathbf{1}_{0\le x \le 1} \mathbf{1}_{0 \le y \le 1} dy = \mathbf{1}_{0 \le x \le 1}.$$

Analog ist $f_Y(y) = \mathbf{1}_{0\le y\le 1}$. Somit sind sowohl X, als auch Y gleichverteilte Zufallsvariablen auf $[0,1]$.

Example 2: *Gleichverteilung auf dem Einheitskreis*

Unter gemeinsamer Dichte $f_{X,Y}(x,y) = \frac{1}{\pi}\mathbf{1}_{x^2+y^2\le 1}$ hat X folgende Dichte

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{\pi} \mathbf{1}_{x^2+y^2 \le 1} dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2}{\pi}\sqrt{1-x^2}.$$

und analog gilt ebenfalls $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$.

2.4 Unabhängigkeit stetiger Zufallsvariablen

Das folgende Theorem gibt eine nützliche Charakterisierung für die Unabhängigkeit von kontinuierlichen Zufallsvariablen.

Theorem 5.11. *Seien $X_1, \ldots, X_n$ Zufallsvariablen mit Dichten $f_1, \ldots f_n$. Dann ist folgende Aussagen äquivalent*

(i) $X_1, \ldots, X_n$ sind unabhängig,

(ii) $X_1, \ldots, X_n$ sind insgesamt stetig mit gemeinsamer Dichte

$$\boxed{f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n)}.$$

Bemerkung 5.12. *Somit sind zwei unabhängige stetige Zufallsvariablen automatisch gemeinsam stetig.*

Beweis. Wir führen den Beweis für $n = 2$. Für $n > 2$ erfolgt der Beweis analog. Seien X und Y zwei stetige Z.V. mit Dichte f_X und f_Y. $\boxed{\text{(i)} \Rightarrow \text{(ii)}}$ Seien X und Y unabhängig. Für jedes $a, b \in \mathbb{R}$ erhalten wir

$$\begin{aligned} \mathbb{P}[X \le a, Y \le b] &= \mathbb{P}[X \le a] \cdot \mathbb{P}[Y \le b] \\ &= \int_{-\infty}^{a} f_X(x)dx \cdot \int_{-\infty}^{b} f_Y(y)dy \\ &= \int_{-\infty}^{a} \int_{-\infty}^{b} f_X(x) f_Y(y) dx dy. \end{aligned}$$

Somit haben X und Y die gemeinsame Dichte $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$.

$\boxed{\text{(ii)} \Rightarrow \text{(i)}}$ Unter Anwendung der Formel (5.3) gilt für jedes $a, b \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}[\phi(X)\psi(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x)\psi(y) f_{X,Y}(x,y) dx dy \\ &\overset{(ii)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x)\psi(y) f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} \phi(x) f_X(x) dx \cdot \int_{-\infty}^{\infty} \psi(y) f_Y(y) dy \\ &= \mathbb{E}[\phi(X)] \cdot \mathbb{E}[\psi(Y)]. \end{aligned}$$

Anwendung der Charakterisierung von unabhängigen Zufallsvariablen (Theorem. 4.17) liefert die Unabhängigkeit von X und Y. □

Example 1: *Gleichverteiltung auf dem Quadrat*
Seien X und Y Zufallsvariablen mit gemeinsamer Dichte $f_{X,Y}(x,y) = \mathbf{1}_{0 \le x, y \le 1}$, dann gilt

$$f_{X,Y}(x,y) = \mathbf{1}_{0 \le x \le 1} \mathbf{1}_{0 \le y \le 1} = f_X(x) \cdot f_Y(y).$$

Mit anderen Worten: Die beiden Koordinaten eines gleichverteilt gewählten Zufallspunktes in $[0,1]^2$ sind unabhängig.

Example 2: *Gleichverteilung auf dem Einheitskreis*

Seien X und Y Zufallsvariablen mit gemeinsamer Dichte $f_{X,Y} = \frac{1}{\pi}\mathbf{1}_D$. Wir haben bereits gesehen, dass $f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}$ und $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$. Somit gilt, dass

$$f_{X,Y}(x,y) \neq f_X(x)f_Y(y).$$

Somit sind die Koordinaten von X und Y eines gleichmäßig verteilten Punktes in D nicht unabhängig!Die kann man leicht nachvollziehen für den Fall, dass X grösser ist als zum Beispiel $\sqrt{3}/2$. Unter diesem Fall muss Y bereits in $[-1/2, 1/2]$ erhalten sein.

Kapitel 6

Grenzwertsätze

Ziele

- Verständnis von Grenzwerten empirischer Durchschnitte.
- Rechnen und Reskalieren von Summen von u.i.v. Zufallsvariablen.
- Verstehen und Anwendung der beiden Grenzwertsätze ZGWS und SGGZ.

Vorbemerkung In diesem Kapitel fixieren wir einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ und eine Folge von u.i.v.-Zufallsvariablen $X_1, X_2, \ldots$. Mit anderen Worten, wir erhalten Zufallsvariablen $X_i : \Omega \to \mathbb{R}$, so dass

$$\forall i_1 < \cdots < i_k \ \forall x_1, \ldots x_k \in \mathbb{R} \quad \mathbb{P}[X_{i_1} \leq x_1, \ldots, X_{i_k} \leq x_k] = F(x_1) \cdots F(x_k).$$

wobei F die allgemeine Verteilungsfunktion ist. Für jedes n betrachten wir die Partialsumme

$$S_n = X_1 + \cdots + X_n,$$

und wir interessieren uns für das Verhalten (wenn n groß ist) der folgender Zufallsvariablen

$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}. \tag{6.1}$$

Dies wird manchmal auch **empirischer Durchschnitt** genannt.

1 Gesetz der grossen Zahlen (GGZ)

Theorem 6.1. *Sei* $\mathbb{E}[|X_1|] < \infty$*. Setzte* $m = \mathbb{E}[X_1]$ *dann gilt*

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = m \ a.s. \tag{6.2}$$

Was bedeutet. (6.2)**?**
Wenn wir ω in Ω fixieren, definieren die Terme $\frac{X_1(\omega)}{1}, \frac{X_1(\omega)+X_2(\omega)}{2}, \cdots$ einfach eine Folge von reellen Zahlen. Die Eigenschaften dieser Folge hängen nun von dem Elementarereigniss ω ab. Somit sind für uns die Elementarereignisse ω's, für welche die Folge $\frac{X_1(\omega)+\cdots+X_n(\omega)}{n}$ zu m konvergiert interessant. Präzise gesagt, definieren wir die Teilmenge $E \subset \Omega$ durch

$$E = \{\omega : \lim_{n\to\infty} \frac{X_1(\omega) + \cdots + X_n(\omega)}{n} = m\}. \tag{6.3}$$

Man kann überprüfen, dass E ein Ereignis ist und Gleichung. (6.2) besagt gerade, dass $\mathbb{P}[E] = 1$ gilt.

Bemerkung 6.2. *In der Aussage des Satzes mag es überraschen, dass die Annahme und die Definition von* m *sich nur auf* X_1 *beziehen. Da die Zufallsvariablen aber u.i.v. sind, haben wir auch* $\mathbb{E}[|X_i|] < \infty$ *und* $m = \mathbb{E}[X_i]$ *für jedes* i*.*

Example 1: *Bernoulli Zufallsvariablen*

Sei $X_1, X_2, \ldots$ eine Folge von u.i.v. Bernoulli Z.V. mit Parameter p. Dann ist

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = p \quad a.s.$$

Example 2: *Exponentialverteilte Zufallsvariablen*

Sei $T_1, T_2, \ldots$ eine u.i.v. Folge von exponential verteilten Zufallsvariablen mit Parameter λ. Dann gilt

$$\lim_{n\to\infty} \frac{T_1 + \cdots + T_n}{n} = \lambda \quad a.s.$$

Beweis. Wir beweisen das Gesetz der grossen Zahlen mittels einer stärkeren Annahme. Wir nehmen folgendes an

$$C := \mathbb{E}[X_1^4] < \infty.$$

Ohne Beschränkung der Allgemeinheit gilt zudem

$$\mathbb{E}[X_1] = 0. \tag{6.4}$$

In der Tat, falls die Behauptung für $m = 0$ gilt, wähle gerade die Zufallsvariable $Y_i = X_i - m$, $i \geq 1$. Somit ist die Behauptung auch für $m \neq 0$ gültig.

Sei $n \geq 1$ fix und setze

$$S_n = X_1 + \cdots + X_n.$$

Erweitere S_n^4 und Ausnutzen der Linearität des Erwartungswerts liefert

$$\mathbb{E}[S_n^4] = \sum_{1 \le i,j,k,\ell \le n} \mathbb{E}[X_i X_j X_k X_\ell].$$

Sobald ein Faktor X_α ein einziges Mal im Produkt $X_i X_j X_k X_\ell$ vorkommt, dann implizieren Unabhängigkeit und die Hypothese (6.4), dass der Erwartungswert des Terms verschwindet. Daher sind die einzigen nicht verschwindenden Terme von der Form $\mathbb{E}[X_i^4]$ und $\mathbb{E}[X_i^2 Y_j^2]$, für $i \neq j$. Anwenden der Unabhängigkeit und der Jensen-Ungleichung liefert für $i \neq j$ $\mathbb{E}[X_i^2 Y_j^2] = \mathbb{E}[X_1^2]^2 \le C$. Da es höchstens $n^2 + n$ nicht verschwindende Terme in der obigen Summe gibt und jeder dieser Terme kleiner als C ist, erhalten wir

$$\mathbb{E}[S_n^4] \le C(n^2 + n) \le 2Cn^2.$$

Für jedes n, betrachte folgendes Ereignis

$$F_n = \left\{ \omega \in \Omega \,:\, \tfrac{|S_n(\omega)|}{n} < n^{-1/8} \right\}.$$

Per Markov Ungleichung gilt

$$\mathbb{P}[F_n^c] = \mathbb{P}[S_n^4 \ge n^{7/2}] \le \frac{\mathbb{E}[S_n^4]}{n^{7/2}} \le \frac{2C}{n^{3/2}}.$$

Nun für endliches $N \ge 1$, betrachte folgendes Ereignis

$$E_N = \bigcap_{n \ge N} F_n = \{\forall n \ge N \quad \frac{|S_n|}{n} \le n^{-1/8}\}.$$

Anwenden der union bound liefert die folgende finale Ungleichung

$$\mathbb{P}[E_N^c] = \mathbb{P}[\bigcup_{n \ge N} F_n^c] \le \sum_{n \ge N} \mathbb{P}[F_n^c] \le \sum_{n \ge N} \frac{C}{n^{3/2}}.$$

Somit gilt $\lim_{N \to \infty} \mathbb{P}[E_N] = 1$. Dies ist jedoch ausreichend. Für jedes $N \in \mathbb{N}$ gilt $E_N \subset E$, wobei E definiert ist nach Gleichung. (6.3) (mit $m = 0$). Nach Monotonie des Wahrscheinlichkeitmasses gilt $\mathbb{P}[E_N] \le \mathbb{P}[E]$, und somit folgt die Behauptung unter Anwendung des Grenzwerts $N \to \infty$. Der Beweis ist somit abgeschlossen. □

2 Anwendung: Monte-Carlo Integration

Eine nützliche Anwendung des Gesetztes der grossen Zahlen ist die Approximation von schwer exakt zu bestimmenden Integralen. Sei $d \ge 1$ eine ganze Zahl. Sei $g : [0,1] \to \mathbb{R}$ sodass,

$$\int_0^1 |g(x)| dx < \infty.$$

Unser Ziel ist es das folgende Integral

$$I = \int_0^1 g(x)dx$$

numerisch zu bestimmen. Allgemein kann das exakte Ausrechnen von I ein sehr schwieriges Problem darstellen. Wir suchen somit nach einer Methode I zufriedenstellend zu approximieren. Die Idee: Wir betrachten I als Erwartungswert und verwenden das Gesetzt der grossen Zahlen, um I zu approximieren. Sei U eine gleichverteilte Zufallsvariable auf $[0,1]$. Dann gilt,

$$\mathbb{E}[g(U)] = \int_0^1 g(x)dx = I.$$

Somit finden wir eine gute Approximation von I , falls wir obigen Erwartungswert $g(U)$ zufriedenstellend bestimmen können. Nun kommt das Gesetzt der grossen Zahlen ins Spiel. Sei $U_1, U_2, \dots$ eine u.i.v. Folge von gleichverteilten Zufallsvariablen auf $[0,1]$ und setzte $X_n = g(U_n)$ für jedes n. Somit sind die Folgenglieder $X_1, X_2, \dots$ u.i.v. und es gilt

$$\mathbb{E}[|X_1|] = \int_0^1 |g(x)|dx < \infty,$$

und $\mathbb{E}[X_1] = I$. Anwendung des Gesetztes der grossen Zahlen liefert

$$\lim_{n\to\infty} \frac{g(U_1) + \cdots + g(U_n)}{n} = I.$$

Somit erhalten wir eine Approximation von I mittels Ausrechnen von $\frac{g(U_1)+\cdots+g(U_n)}{n}$ für hinreichend grosses n. Dabei ist der letzte Term mittels Simulation von gleichverteilten Zufallsvariablen $U_1, \dots U_n$ einfach auszurechnen.

Die Methode kann einfach verallgemeinert werden. Man kann zum Beispiel weitere Dichten und Verteilungen nutzten um Integrale über $\mathbb{R}$ zu bestimmen.

3 Konvergenz in Verteilung

Deterministische Zahlen in $\mathbb{R}$ können wir stets einen Abstand zwischen ihnen zuordnen: Somit ist der Abstand zwischen x und y durch $|x-y|$ gegeben. Daraus ergibt sich ein natürlicher Begriff der Konvergenz. Eine Folge von reellen Zahlen $(x_n)_{n\in\mathbb{N}}$ konvergiert zu x, wenn

$$\lim_{n\to\infty} |x_n - x| = 0.$$

Eine Methode den "Abstand" zwischen X und Y zu messen, ist der Vergleich der Verteilungsfunktionen von X und Y. Die Idee: X und Y haben ähnliches stochastisches Verhalten, falls ihre Verteilungsfunktionen F_X und F_Y nahe beieinander liegen. Dies motiviert den Begriff der "Konvergenz in Verteilung" und wir geben die folgende Definition.

Definition 6.3. *Seien $(X_n)_{n\in\mathbb{N}}$ und X Zufallsvariablen. Wir schreiben*

$$X_n \overset{\text{Approx}}{\approx} X \ \textit{as}\ n \to \infty$$

falls für jedes $x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}[X_n \le x] = \mathbb{P}[X \le x].$$

Example 1: *Bernoulli Zufallsvariablen*

Für jedes n, sei X_n eine Bernoulli Zufallsvariable mit Parameter $p_n \in [0,1]$. Falls $\lim_{n\to\infty} p_n = p$ gilt, erhalten wir

$$X_n \overset{\text{Approx}}{\approx} X \text{ für } n \to \infty,$$

wobei X eine Bernoulli Zufallsvariable mit Parameter p ist.

Example 2: *Approximieren der Gleichverteilung*

Im ersten Beispiel handelt es sich um diskrete Zufallsvariablen, die gegen eine andere diskrete Zufallsvariable konvergieren. Es ist auch möglich, dass eine Folge von diskreten Zufallsvariablen gegen eine stetige Zufallsvariable konvergiert. Das folgende Beispiel illustriert dies.

Für jedes n, sei X_n eine gleichverteilte Zufallsvariable auf $\{0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\}$ (d.h. $\mathbb{P}[X_n = \frac{k}{n}] = \frac{1}{n+1}$ für $k = 0,1,2.\dots,n$). Dann erhalten wir

$$X_n \overset{\text{Approx}}{\approx} X \text{ as } n \to \infty,$$

wobei X eine gleichverteilte Zufallsvariable auf $[0,1]$ ist.

Tatsächlich erhalten wir für jedes $x \in [0,1]$

$$\mathbb{P}[X_n \le x] = \frac{\lfloor xn \rfloor}{n} \xrightarrow[n\to\infty]{} x = \mathbb{P}[X \le x],$$

wobei die Konvergenz für $x \notin [0,1]$ analog auszurechnen ist.

4 Zentraler Grenzwertsatz

Eine Frage der Fluktation?

Das Gesetz der großen Zahlen besagt, dass für große n der empirische Durchschnitt (6.1) nahe dem Erwartungswert $m = \mathbb{E}[X_1]$ ist. Eine zweite sehr natürliche Frage, die man stellen kann, ist:

$$\text{Wie weit ist } \frac{X_1 + \dots + X_n}{n} \text{ typischerweise von } m \text{ entfernt?}$$

Fluktuationen von Normalverteilten Zufallsvariablen

Betrachten wir zunächst den sehr lehrreichen Fall, dass $X_1, X_2, \dots$ eine Folge von i.i.d. normalen Zufallsvariablen mit den Parametern m und σ^2 ist. Dann sagen uns die Ergebnisse, die wir für normale Zufallsvariablen gesehen haben, dass

$$Z = \frac{X_1 + \dots + X_n}{n} - m$$

wiederum eine normale Zufallsvariable mit den Parametern $\bar{m} = 0$ und $\bar{\sigma}^2 = \frac{1}{n}\sigma^2$. Die Standardabweichung $\bar{\sigma} = \frac{1}{\sqrt{n}}\sigma$ stellt die typischen Schwankungen von Z dar. Grob kann man sagen, dass der typische Abstand zwischen $\frac{X_1+\dots+X_n}{n}$ und m von der Ordnung $\frac{\sigma}{\sqrt{n}}$ ist.

In diesem Zusammenhang ist eine ist eine Skalierung der Zufallsvariable Z um einen Faktor $\frac{\sqrt{n}}{\sigma}$ natürlich. Nach Skalierung erhalten wir gerade Fluktuationen der Ordnung 1: Nochmalige Verwendung der Eigenschaften von normalverteilten Zufallsvariablen sehen wir, dass

$$\frac{\sqrt{n}}{\sigma} Z = \frac{X_1 + \dots + X_n - n \cdot m}{\sqrt{\sigma^2 n}}$$

standard-normalverteilt ist.

Betrachtet man also i.i.d.-Normalverteilungen mit Erwartungswert m und Varianz σ^2, dann ist die Zufallsvariable

$$\frac{X_1 + \dots + X_n - n \cdot m}{\sqrt{\sigma^2 n}}$$

eine neu skalierte standard-normalverteilte Zufallsvariable der Fluktuation $\frac{X_1+\dots+X_n}{n}$.

Allgemeine Fall: Der zentrale Grenzwertsatz

Seien $X_1, X_2, \dots$ nicht normalverteilt, dann ist die Berechnung der Verteilung

$$\frac{X_1 + \dots + X_n - n \cdot m}{\sqrt{\sigma^2 n}}$$

nicht immer einfach. Hier setzt der zentrale Grenzwertsatz gerade an. Er besagt, dass für immer grösser werdende n die Verteilung der obige Zufallsvariablen sich der Verteilung einer Standard normalverteilten Zufallsvariable annähert.

Theorem 6.4 (Zentraler Grenzwertsatz (ZGWS))**.** *Nehme an, dass der Erwartungswert $\mathbb{E}[X_1^2]$ wohldefiniert und endlich ist. Setzte $m = \mathbb{E}[X_1]$ und $\sigma^2 = \mathrm{Var}(X_1)$, dann gilt folgender Grenzwert*

$$\mathbb{P}[\frac{S_n - n \cdot m}{\sqrt{\sigma^2 n}} \le a] \xrightarrow[n\to\infty]{} \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2} dx.$$

für jedes $a \in \mathbb{R}$.

Beachte gerade, dass Φ gerade die Verteilungsfunktion einer Zufallsvariable $Z \sim \mathcal{N}(0,1)$ ist. Der Satz besagt somit, dass für grosse $n \in \mathbb{N}$ die Zufallsvariable

$$Z_n = \frac{S_n - n \cdot m}{\sqrt{\sigma^2 n}}$$

in Verteilung $Z \sim \mathcal{N}(0,1)$ ähnelt.

Bemerkung 1:

Für jedes n können wir die Linearitätseigenschaften des Erwartungswerts und der Varianz verwenden, um zu zeigen, dass

$$\mathbb{E}[Z_n] = 0 \quad \text{und} \quad \mathrm{Var}(Z_n) = 1.$$

Bemerkung 2:

Der zentrale Grenzwertsatz hilft uns bei der Vorhersage des Verhaltens von S_n für große n. Betrachten wir zum Beispiel $p := \mathbb{P}[Z \in [-2,2]]$, wobei Z eine standardnormale Zufallsvariable ist.

Es ist bekannt, dass $p \simeq 0,95$ ist (dies entspricht dem blauen und braunen Bereich in der Abbildung unten).

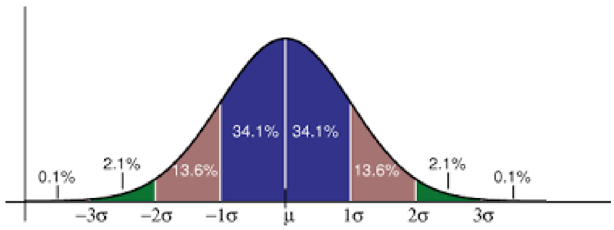


Abbildung 6.1: Quantile einer Normalverteilung für μ und σ^2

Verwendung des Zentralen Grenzwertsatzes liefert

$$\lim_{n\to\infty} \mathbb{P}[mn - 2\sqrt{\sigma^2 n} \leq S_n \leq mn + 2\sqrt{\sigma^2 n}] = p \simeq 95\%.$$

Example 1: *Berechnung von Binomialverteilungen*

Sei $X \sim \mathrm{Bin}(20, \frac{1}{2})$ verteilt. Wie können wir $\mathbb{P}(X \leq 14)$ ausrechnen? Natürlich können wir dies per Hand mittels $\sum_{k=1}^{14} p_k$, wobei $p_k = \binom{20}{k}\frac{1}{2^{20}}$ mühsam ausrechnen. Alternativ eignet sich hier gerade das Anwenden des ZGWS, denn wir können X gerade also die Summe von 20 unabhängigen gleichverteilten Bernoulli-Zufallsvariablen mit Parameter $p = \frac{1}{2}$ schreiben. Diese haben $\mu = \frac{1}{2}$ und $\sigma^2 = \frac{1}{4}$ und somit erhalten wir eine Schätzung für die obige Wahrscheinlichkeit

$$\mathbb{P}(X \leq 14) = \mathbb{P}(\sum_{i=1}^{14} X_i \leq 14) = \mathbb{P}\left(\frac{\sum_{i=1}^{14} X_i - 20 \times \frac{1}{2}}{\sqrt{20} \times \frac{1}{4}} \leq \frac{14 - 20 \times \frac{1}{2}}{\sqrt{20} \times \frac{1}{2}}\right) \approx \Phi(1.788).$$

Literaturverzeichnis

[LSW21] J. Lengler, A. Steger, and E. Welzl, **Algorithmen und Wahrscheinlichkeit**, 2021.

[Sch10] M. Schweizer, **Wahrscheinlichkeit und Statistik**, 2010.

[Wil01] David Williams, **Weighing the odds**, Cambridge University Press, Cambridge, 2001, A course in probability and statistics. MR 1854128