
Stochastik 1

LEIF DÖRING¹

UNIVERSITÄT MANNHEIM

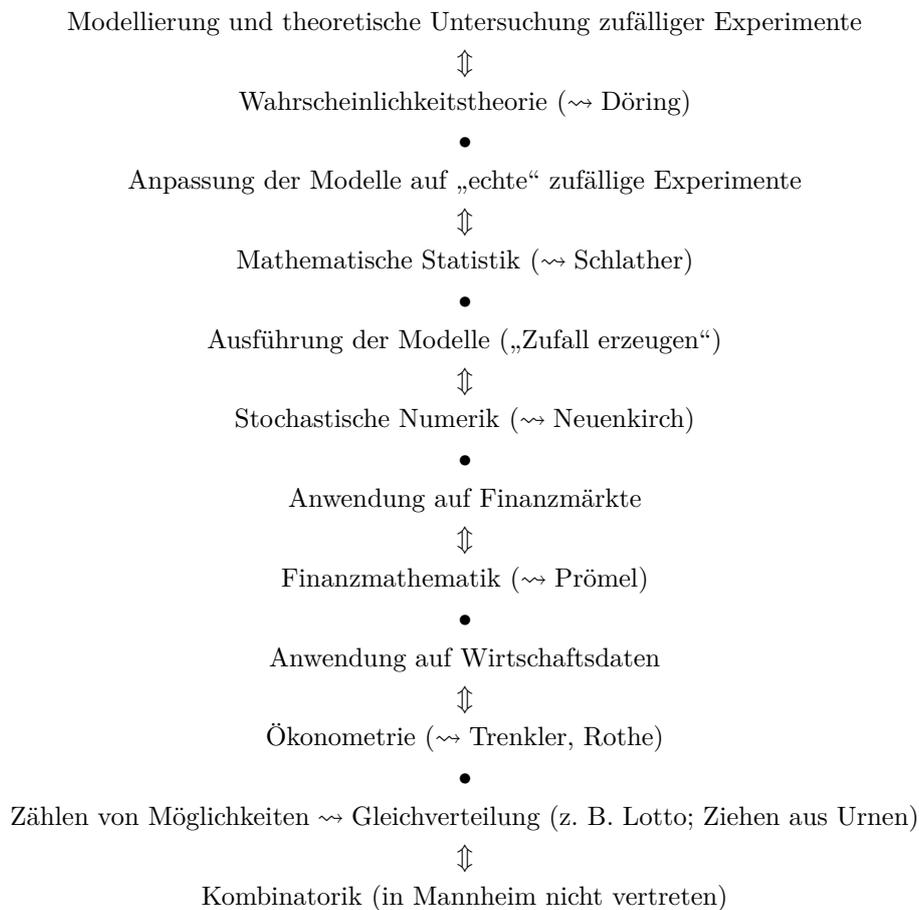
¹Besten Dank an den Skriptschreiber (aka Johannes Nägele) fürs Tippen und die Erstellung der Graphiken!
Auch vielen Dank für die Korrekturen (auch zukünftige) von Tippfehlern durch Teilnehmer der Vorlesung!

Inhaltsverzeichnis

1	Maßtheorie (Modellierung von Ereignissen und Wahrscheinlichkeiten)	3
1.1	σ -Algebren und Maße	3
1.2	Erzeuger von σ -Algebren und Dynkin-Systeme	8
1.3	Konstruktion von Maßen	14
1.4	Das Beispiel für Wahrscheinlichkeitsmaße	22
2	Abbildungen zwischen messbaren Räumen	32
2.1	Messbare Abbildungen	32
2.2	Bildmaße oder „push-forward“ eines Maßes	34
2.3	Messbare numerische Funktionen	35
3	Integrationstheorie	38
3.1	Das (allgemeine) Lebesgue Integral	38
3.2	Konvergenzsätze	51
3.3	Integrale für <u>das</u> Beispiel	56
3.4	Integralabschätzungen und L^p -Räume	63
3.5	Produktmaße und Satz von Fubini	68
4	Stochastik	75
4.1	Zufallsvariablen	75
4.2	Zufallsvektoren	85
4.3	Rechnen mit Zufallsvariablen	99
4.3.1	Inverse Transformations Methode	99
4.3.2	Ein paar konkrete Beispiele	101
4.3.3	Summen von unabhängigen Zufallsvariablen	103
4.4	Bedingte Wahrscheinlichkeiten und Unabhängigkeit	108
4.5	Konvergenz von Folgen von Zufallsvariablen	115
4.6	Starkes Gesetz der großen Zahlen	126
4.7	Zentraler Grenzwertsatz	134
5	Sammlung von Fakten und Beispielen zu Zufallsvariablen	139

Was machen wir, was nicht?

„Stochastik“ ist ein Oberbegriff für „Mathematik des Zufalls“. In Mannheim ist die Stochastik in Lehre und Forschung sehr ausgeprägt:



Das Ziel dieser Vorlesung ist es, die Grundlagen der Stochastik zu legen. Das ist anfangs etwas trocken, ihr werdet aber im Verlauf des Studiums davon profitieren, dass alle Begriffe auf stabilen Fundament stehen. In den Vorlesungen Stochastik 2, Monte Carlo Methoden, Finanzmathematik, Ökonometrie und Wahrscheinlichkeitstheorie 1 werden die Grundlagen noch im Bachelor angewandt und in diversen Spezialisierungsrichtungen im Master erweitert.

Teil 1: Maß- und Integrationstheorie

Kapitel 1

Maßtheorie (Modellierung von Ereignissen und Wahrscheinlichkeiten)

Vorlesung 1

Maß- und Integrationstheorie bildet die formale Grundlage um zufällige Experimente zu modellieren. In diesem ersten Teil der Vorlesungen beweisen wir alle notwendigen Theoreme. Nicht alles wird später uneingeschränkt wichtig sein, das Arbeiten mit den neuen Begriffen wird sich in zukünftigen Vorlesungen aber auszahlen!

Im Prinzip sind die kommenden fünf Vorlesungen total elementar, wir brauchen eigentlich nur Kenntnisse über Mengen, Folgen und Reihen. Die Vorlesung nutzt also nur Kenntnisse der Analysis 1. Dennoch wird euch der Inhalt schwer fallen weil wir Mengensysteme nicht visualisieren können und daher viel abstrakt denken müssen. Es wird sehr wichtig sein, die richtigen Beispiele im Kopf zu haben. Diese sollten nicht zu einfach sein, weil sonst der Großteil der Schwierigkeiten nicht erkannt werden kann. Für σ -Algebren sollten wir möglichst schnell die Borel- σ -Algebra als Standardbeispiel im Kopf halten, für Maße das Lebesgue Maß. Endliche Beispiele werden wir nur ganz kurz als Motivation der Maßtheorie für Stochastik betrachten (Würfeln, Münzwurf, etc.), solche Beispiele bringen leider nicht viel um die Konzepte der Wahrscheinlichkeitstheorie richtig zu verstehen.

1.1 σ -Algebren und Maße

$\Omega \neq \emptyset$ sei immer eine beliebige Grundmenge. Für $A \subseteq \Omega$ bezeichnet A^C immer das Komplement von A in Ω , d. h. $A^C = \{w \in \Omega \mid w \notin A\}$. $\mathcal{P}(\Omega)$ bezeichnet die Potenzmenge von Ω (inklusive \emptyset und Ω), eine Teilmenge von $\mathcal{P}(\Omega)$ ist also eine Menge von Mengen.

Definition 1.1.1.  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt σ -Algebra, falls

- (i) $\Omega \in \mathcal{A}$,
- (ii) $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$, das nennt man auch stabil (oder abgeschlossen) unter Komplementbildung,
- (iii) $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$, das nennt man auch stabil (oder abgeschlossen) unter abzählbarer Vereinigung.

Elemente von \mathcal{A} heißen **messbare Mengen**. Ist $\mathcal{A} \subseteq \mathcal{B}$ und \mathcal{A}, \mathcal{B} sind σ -Algebren, so nennt man \mathcal{A} Unter- σ -Algebra von \mathcal{B} .

Beispiel 1.1.2. 

- $\mathcal{A}_1 = \{\emptyset, \Omega\}$
- $\mathcal{A}_2 = \{\emptyset, \Omega, A, A^C\}$ für $A \subseteq \Omega$ beliebig
- $\mathcal{A}_3 = \{A \subseteq \Omega \mid A \text{ oder } A^C \text{ ist abzählbar}\}$

In allen Beispielen muss man nur die drei definierenden Eigenschaften testen. Bei den ersten zwei Beispielen ist das direkt, indem man alle Möglichkeiten ausprobiert. Im dritten Beispiel müssen wir nur bei der abzählbaren Vereinigung kurz nachdenken. Seien also A_1, A_2, \dots Mengen, die entweder abzählbar sind oder deren Komplemente abzählbar sind. Sind all diese Mengen abzählbar, so ist nach Analysis 1 auch die Vereinigung abzählbar, also ist die Vereinigung wieder in \mathcal{A}_3 . Ist eine der Mengen nicht abzählbar, sagen wir A_j , so ist das Komplement A_j^C abzählbar. Doch dann ist wegen

$$\left(\bigcup_{k=1}^{\infty} A_k\right)^C = \bigcap_{k=1}^{\infty} A_k^C \subseteq A_j^C$$

das Komplement der Vereinigung nach Analysis 1 abzählbar.

Lemma 1.1.3.  Für jede σ -Algebra \mathcal{A} gelten:

- (i) $\emptyset \in \mathcal{A}$
- (ii) $A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcap_{k=1}^{\infty} A_k \in \mathcal{A}$
- (iii) Aus $A, B \in \mathcal{A}$ folgt $A \setminus B := A \cap B^C \in \mathcal{A}$ sowie $A \Delta B := (A \cap B^C) \cup (B \cap A^C) \in \mathcal{A}$.

Beweis. Übung □

Bemerkung. 

Im Folgenden nutzen wir die erweiterte Zahlengerade $\overline{\mathbb{R}} = [-\infty, +\infty] := \mathbb{R} \cup \{-\infty, +\infty\}$. Wir definieren

- $-\infty < a < +\infty$,
- $+\infty + a = +\infty$ und $-\infty + a = -\infty$ für alle $a \in \mathbb{R}$,
- $x \cdot (+\infty) = +\infty$ und $x \cdot (-\infty) = -\infty$ für alle $x > 0$,
- $0 \cdot (+\infty) = 0$ und $0 \cdot (-\infty) = 0$,
- $+\infty + (+\infty) = +\infty$ und $-\infty + (-\infty) = -\infty$,
- $-\infty + (+\infty)$ wird nicht definiert.

Im Gegensatz zu \mathbb{R} können wir aus $\overline{\mathbb{R}}$ keine sinnvolle algebraische Struktur formen, das soll uns aber nicht weiter stören. Sehr oft schreibt man ∞ statt $+\infty$. Wenn wir in dieser Vorlesung von den natürlichen Zahlen sprechen, meinen wir $\mathbb{N} = \{0, 1, \dots\}$, die 0 soll also zu \mathbb{N} gehören.

Definition 1.1.4.  Für eine σ -Algebra \mathcal{A} heißt $\mu: \mathcal{A} \rightarrow [0, \infty]$ ein **Maß auf \mathcal{A}** , falls folgende Eigenschaften gelten:

- (i) $\mu(\emptyset) = 0$
- (ii) Sind $A_1, A_2, \dots \in \mathcal{A}$ paarweise disjunkte Mengen, so gilt $\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$. Wir nenne diese Eigenschaft σ -Additivität, wobei sich das σ auf die unendliche Anzahl von Mengen bezieht.

Ein Maß μ heißt **endlich**, falls $\mu(\Omega) < \infty$. μ heißt **Wahrscheinlichkeitsmaß**, falls $\mu(\Omega) = 1$.

Natürlich impliziert die σ -Additivität auch die endliche Additivität

$$\mu\left(\bigcup_{k=1}^N A_k\right) = \sum_{k=1}^N \mu(A_k).$$

Dazu wird einfach $A_{k+1} = A_{k+2} = \dots = \emptyset$ gewählt.

Bemerkung 1.1.5. 

Oft werden Wahrscheinlichkeitsmaße mit \mathbb{P} anstelle von μ geschrieben und **Verteilungen** genannt.

Definition 1.1.6. 

- (Ω, \mathcal{A}) heißt **messbarer Raum**
- $(\Omega, \mathcal{A}, \mu)$ heißt **Maßraum**
- $(\Omega, \mathcal{A}, \mathbb{P})$ heißt **Wahrscheinlichkeitsraum**
- $\mu(A)$ nennt man **Maß** von A

Bemerkung 1.1.7.  Bei einem Wahrscheinlichkeitsraum spricht man von **Ereignissen** A statt messbaren Mengen. $\mathbb{P}(A)$ heißt **Wahrscheinlichkeit** von A . Einelementige messbare Mengen $A = \{a\}$ heißen in Wahrscheinlichkeitsräumen **Elementarereignisse**.

Um langsam in die Denkweise der Stochastik einzusteigen, werden wir wieder und wieder diskutieren, warum unsere Modelle für die Modellierung zufälliger Experimente gut geeignet sind.

Diskussion 1.1.8.  **[Stochastische Modellierung, Nr. 1]**

Warum machen die Definitionen von Wahrscheinlichkeitsräumen $(\Omega, \mathcal{A}, \mathbb{P})$ für die Modellierung von zufälligen Experimenten Sinn? Wir interpretieren dazu

- Ereignisse = „Ereignisse, deren Eintreten (oder Nichteintreten) beobachtet werden kann.“ Die σ -Algebra besteht also aus den Ereignissen des Experiments, die wir beobachten können.
- $\mathbb{P}(A)$ = „Wahrscheinlichkeit des Eintretens des Ereignisses A .“
- A^C = „Gegenereignis“
- $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$ bedeutet „Eintreten von Ereignis beobachtbar, impliziert Eintreten von Gegenereignis beobachtbar.“
- $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$ bedeutet „Gegenereignis hat Gegenwahrscheinlichkeit.“

Zu dem letzten Punkt beachte man, dass $\mathbb{P}(A) + \mathbb{P}(A^C) = \mathbb{P}(\Omega) = 1$ wegen $A \cup A^C = \Omega$ gilt. Also ist die Forderung der Additivität von Maßen sinnvoll. Natürlich sollte die Vereinigung zweier Ereignisse „es passiert A oder B “ auch wieder beobachtbar (also messbar) sein, wenn A und B beobachtbar sind. Mit vollständiger Induktion sollten damit alle endlichen Vereinigungen messbarer Mengen wieder messbar sein. Damit haben wir den Sinn der Definitionen einer σ -Algebra und eines Maßes großteils eingesehen. Nur die Erweiterung von endlichen auf unendliche Vereinigungen ist nicht so einfach zu motivieren. Hier bleibt für den Moment nur zu sagen: Es würde nicht funktionieren.

Als Beispiel modellieren wir den Wurf eines Würfels gemäß obiger Interpretation. Sei $\Omega = \{1, \dots, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(\{1\}) = \dots = \mathbb{P}(\{6\}) = \frac{1}{6}$. Ein Ereignis $A \in \mathcal{A}$ bedeutet also „Eine der Zahlen in A ist gewürfelt worden“. Die Wahrscheinlichkeiten aller weiteren Ereignisse sind festgelegt, indem

das Ereignis in die disjunkten Elementarereignisse zerlegt wird, z. B. die Wahrscheinlichkeit eine gerade Zahl zu würfeln:

$$\mathbb{P}(\{2, 4, 6\}) \stackrel{\text{disj.}}{=} \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{3}{6} = \frac{1}{2}.$$

Ganz analog ist natürlich ein Maß auf der Potenzmenge einer endlichen Menge schon eindeutig durch die Werte auf einelementigen Mengen festgelegt.

Da wir das Maß einer Menge in einer abstrakten Weise als die „Größe“ interpretieren, sind folgende Eigenschaften wichtig. Überlegt euch mal, ob ihr das für eine naive Idee von „Größe“ von Mengen auch erwarten würdet (z. B. beim Zählmaß auf \mathbb{N}).

Lemma 1.1.9.  [Monotonie und Subadditivität]

Es sei μ ein Maß auf einer beliebigen σ -Algebra \mathcal{A} , dann gelten:

- (i) Sind $A, B \in \mathcal{A}$ mit $B \subseteq A$, so gilt $\mu(B) \leq \mu(A)$.
- (ii) Sind $A_1, A_2, \dots \in \mathcal{A}$, so gilt: $\mu(\cup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} \mu(A_k)$.

Beweis. (i) Mit den definierenden Eigenschaften des Maßes gilt:

$$\mu(B) \leq \mu(B) + \mu(A \setminus B) = \mu(B \cup A \setminus B) = \mu(A),$$

wobei wir beide definierenden Eigenschaften des Maßes genutzt haben.

(ii) wird in der großen Übung diskutiert. □

Zu beachten ist, dass bei der Subadditivität nicht gefordert wird, dass die Mengen disjunkt sind (sonst würde schließlich Gleichheit gelten!). Die Ungleichheit entsteht dadurch, dass bei nicht-disjunkten Mengen der Schnitt mehrfach gezählt wird. Am anschaulichsten wird es, wenn man zwei Mengen und diese in disjunkte Teile zerlegt:

$$\begin{aligned} \mu(A \cup B) &= \mu(A \setminus B \cup (A \cap B) \cup B \setminus A) \\ &\stackrel{\sigma\text{-Add.}}{=} \mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A) \\ &\leq \mu(A \setminus B) + 2\mu(A \cap B) + \mu(B \setminus A) \stackrel{\sigma\text{-Add.}}{=} \mu(A) + \mu(B). \end{aligned}$$

Um ein wenig mit der Definition des Maßes zu experimentieren, rechnet mal die folgende Bemerkung nach:

Bemerkung. Sind μ_1, μ_2 Maße auf \mathcal{A} und $a, b \geq 0$, so ist auch die Summe $A \mapsto a\mu_1(A) + b\mu_2(A)$ ein Maß auf \mathcal{A} . Summen von Maßen nennt man auch **Mischung**. Sind \mathbb{P}_1 und \mathbb{P}_2 Wahrscheinlichkeitsmaße und zusätzlich $a + b = 1$, so ist die Mischung $\mathbb{P} = a\mathbb{P}_1 + b\mathbb{P}_2$ wieder ein Wahrscheinlichkeitsmaß.

Kommen wir nun zu ein paar Beispielen:

Beispiel 1.1.10.  [endliche Gleichverteilung]

Sei $\#\Omega < \infty$ und $\mathcal{A} = \mathcal{P}(\Omega)$. Dann heißt das Maß $\mu(A) = \frac{\#A}{\#\Omega}$ Gleichverteilung auf Ω . Weil $\mu(\Omega) = 1$, würde man \mathbb{P} statt μ schreiben. Der Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ ist ein Modell für das zufällige Experiment, in dem aus $\#\Omega$ vielen Elementen jedes Element mit der selben Wahrscheinlichkeit gezogen wird, zum Beispiel Lotto.

Beispiel 1.1.11.  [abzählbare Verteilungen, Zählmaß]

Sei Ω abzählbar, z. B. $\Omega = \mathbb{N}$. Wir wählen $\mathcal{A} = \mathcal{P}(\Omega)$ und eine Folge $(p_k)_{k \in \Omega}$ nicht-negativer Zahlen. Definieren wir

$$\mu(A) := \sum_{k \in A} p_k, \quad A \in \mathcal{A},$$

so ist μ ein Maß. Weil ein Maß per Definition nicht-negativ ist, muss natürlich $p_k \geq 0$ gelten für alle $k \in \mathbb{N}$ (wähle dazu $A = \{k\}$). Zwei Spezialfälle:

- Damit μ ein Wahrscheinlichkeitsmaß ist, muss $\sum_{k \in \Omega} p_k = \mu(\Omega) = 1$ gelten. In dem Fall würden wir wieder \mathbb{P} statt μ schreiben.
- Ist $p_k = 1$ für alle $k \in \Omega$, so heißt μ **Zählmaß** weil $\mu(A) = \#A$ die Anzahl der Elemente von A zählt.

Beispiel 1.1.12. ▶ [Poissonverteilung]

Hier ist ein konkretes Beispiel zu der vorherigen Klasse von Beispielen, die Poissonverteilung. Für ein $\lambda > 0$ (der Parameter der Verteilung) sei $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ für $k \in \mathbb{N}$. Es gelten dann

- $p_k \geq 0$ für alle $k \in \mathbb{N}$,
- $\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = e^{-\lambda+\lambda} = 1$.

Also definiert $\mathbb{P}(A) = e^{-\lambda} \sum_{k \in A} \frac{\lambda^k}{k!}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Beispiel 1.1.13. ▶ [Diracmaß] Sei \mathcal{A} eine σ -Algebra auf Ω und $x \in \Omega$, so heißt

$$\delta_x(A) := \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}, \quad A \in \mathcal{A},$$

Diracmaß an der Stelle x . Die Eigenschaften eines Maßes kann man ganz einfach checken:

- Aufgrund der Definition gilt natürlich $\delta_x(\emptyset) = 0$.
- Für paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{A}$ gilt

$$\begin{aligned} \delta_x\left(\bigcup_{k=1}^{\infty} A_k\right) &= \begin{cases} 1, & x \in \bigcup_{k=1}^{\infty} A_k \\ 0, & x \notin \bigcup_{k=1}^{\infty} A_k \end{cases} \\ &= \sum_{k=1}^{\infty} \delta_x(A_k), \end{aligned}$$

weil in der unendlichen Summe nur der Summand 1 sein kann, in dem x liegt.

Also ist das Diracmaß ein Maß auf \mathcal{A} .

Weitere wichtige Beispiele wie die geometrische Verteilung und die Binominalverteilung kommen auf dem Übungsblatt zum ausprobieren. An dieser Stelle legen wir die Begrifflichkeiten der Stochastik wieder beiseite und beschäftigen uns für die nächsten Wochen nur mit allgemeinen Maßen. Zum Gewöhnen für später denkt immer daran, dass endliche Maße und Wahrscheinlichkeitsmaße sehr eng beieinander liegen: Durch $\mathbb{P}(A) := \frac{\mu(A)}{\mu(\Omega)}$ kann ein endliches Maß immer zu einem Wahrscheinlichkeitsmaß „normiert“ werden.

Um uns mit den definierenden Eigenschaften weiter vertraut zu machen, beweisen wir eine wichtige Eigenschaft von Maßen:

Vorlesung 2

Satz 1.1.14. ▶ [Stetigkeit von Maßen]

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $(A_n)_{n \in \mathbb{N}}$ eine Folge messbarer Mengen, so gelten:

- Aus $A_n \uparrow A$ (d. h. $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$, $\bigcup_{n=1}^{\infty} A_n = A$) folgt $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$.
- Aus μ endlich und $A_n \downarrow A$ (d. h. $A_1 \supseteq A_2 \supseteq \dots$, $\bigcap_{n=1}^{\infty} A_n = A$) folgt $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$.

Beweis.

(i) Definiere

$$\begin{aligned} A'_1 &:= A_1 \\ A'_2 &:= A_2 \setminus A_1 \\ A'_n &:= A_n \setminus A_{n-1}, \quad n \geq 3. \end{aligned}$$

Weil die A'_n paarweise disjunkt sind und $A_n = \bigcup_{k=1}^n A'_k$ gilt, folgt

$$\begin{aligned} \lim_{n \rightarrow \infty} \mu(A_n) &= \lim_{n \rightarrow \infty} \mu\left(\bigcup_{k=1}^n A'_k\right) \stackrel{\text{Def. Ma\ss}}{=} \lim_{n \rightarrow \infty} \sum_{k=1}^n \mu(A'_k) \stackrel{\text{Def.}}{=} \sum_{k=1}^{\infty} \mu(A'_k) \\ &\stackrel{\text{Def. Ma\ss und disj.}}{=} \mu\left(\bigcup_{k=1}^{\infty} A'_k\right) = \mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \mu(A). \end{aligned}$$

(ii) Folgt sofort aus (i) weil

$$A_n \downarrow A \Leftrightarrow A_n^C \uparrow A^C, \quad n \rightarrow \infty.$$

Weil μ endlich ist gilt für alle messbaren Mengen

$$\mu(\Omega) = \mu(A \cup A^C) \stackrel{\text{Ma\ss}}{=} \mu(A) + \mu(A^C)$$

und damit auch $\mu(A) = \mu(\Omega) - \mu(A^C)$. Damit folgt mit (i)

$$\lim_{n \rightarrow \infty} \mu(A_n) = \lim_{n \rightarrow \infty} (\mu(\Omega) - \mu(A_n^C)) = \mu(\Omega) - \lim_{n \rightarrow \infty} \mu(A_n^C) = \mu(\Omega) - \mu(A^C) = \mu(A).$$

□

Beispiel 1.1.15.  [Gegenbeispiel zu (ii) mit $\mu(\Omega) = \infty$]

Sei $\Omega = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$, $p_k = 1$ für alle $k \in \mathbb{N}$ und μ das Zählmaß:

$$\mu(A) = \sum_{k \in A} p_k.$$

Mit $A_n = \{n, n+1, \dots\}$ gilt $\mu(A_n) = +\infty$ für alle $n \in \mathbb{N}$ und $A_n \downarrow A = \emptyset$. Weil $\mu(\emptyset) = 0$ aber $\mu(A_n) = +\infty$ für alle $n \in \mathbb{N}$, gilt die Aussage von Satz 1.1.14 hier nicht.

1.2 Erzeuger von σ -Algebren und Dynkin-Systeme

Wir wollen Maße auf komplizierten σ -Algebren studieren, indem wir die Maße nur auf sehr wenigen Mengen der σ -Algebra betrachten (auf Erzeugern). Das ist ein wenig wie in der linearen Algebra, dort müssen lineare Abbildungen auch nur auf einer Basis definiert werden. Ganz konkret halten wir folgendes Ziel im Kopf: Wir werden zeigen, dass Maße auf $\mathcal{B}(\mathbb{R}^d)$ schon durch die Werte auf den Quadern eindeutig definiert werden können.

Satz 1.2.1.  Der Durchschnitt einer beliebigen Menge von σ -Algebren ist eine σ -Algebra.

Beweis. Sei \mathcal{A}_i , $i \in I$, eine Menge von σ -Algebren und $\mathcal{A} := \bigcap_{i \in I} \mathcal{A}_i$. Wir checken die drei Eigenschaften einer σ -Algebra für \mathcal{A} :

- (i) $\Omega \in \mathcal{A}$ ist klar weil $\Omega \in \mathcal{A}_i$ für alle $i \in I$ und damit ist Ω auch im Durchschnitt.
- (ii) Sei $A \in \mathcal{A}$, also ist $A \in \mathcal{A}_i$ für alle $i \in I$. Weil alle \mathcal{A}_i σ -Algebren sind, ist auch $A^C \in \mathcal{A}_i$ für alle $i \in I$ und damit ist $A^C \in \mathcal{A}$. Folglich ist \mathcal{A} abgeschlossen bezüglich Komplementbildung.

(iii) Sei (A_n) eine Folge paarweiser disjunkter Mengen in \mathcal{A} , also $A_n \in \mathcal{A}_i$ für alle i und n . Weil das alles σ -Algebren sind, gilt

$$\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_i$$

für alle $i \in I$. Damit ist die Vereinigung auch im Durchschnitt aller \mathcal{A}_i , also $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Folglich ist \mathcal{A} auch abgeschlossen bezüglich beliebigen Vereinigungen. \square

Bemerkung 1.2.2. ▶ Die Vereinigung von σ -Algebren ist nicht immer eine σ -Algebra. Für das Übungsblatt sollt ihr euch dazu Beispiele überlegen.

Korollar 1.2.3. ▶ Sei $\mathcal{E} \subseteq \mathcal{P}(\Omega)$, so existiert genau eine σ -Algebra \mathcal{A} mit

- (i) $\mathcal{E} \subseteq \mathcal{A}$
- (ii) Ist $\mathcal{E} \subseteq \mathcal{B}$ und \mathcal{B} ist eine σ -Algebra, so gilt $\mathcal{A} \subseteq \mathcal{B}$.

Dabei bedeutet (ii), dass \mathcal{A} die kleinste σ -Algebra ist, die \mathcal{E} enthält.

Beweis. Existenz:

$$\mathcal{A} := \bigcap_{\substack{\mathcal{E} \subseteq \mathcal{B}, \\ \mathcal{B} \text{ } \sigma\text{-Alg.}}} \mathcal{B}$$

erfüllt die geforderten Eigenschaften.

Eindeutigkeit: Sei \mathcal{A}' eine weitere solche σ -Algebra. Dann gilt $\mathcal{A} \subseteq \mathcal{A}'$ weil \mathcal{A} der Schnitt über alle solche \mathcal{A}' ist. Wegen (ii) ist auch $\mathcal{A}' \subseteq \mathcal{A}$. Also ist $\mathcal{A} = \mathcal{A}'$. \square

Definition 1.2.4. ▶ Für $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ heißt

$$\sigma(\mathcal{E}) = \bigcap_{\substack{\mathcal{E} \subseteq \mathcal{B}, \\ \mathcal{B} \text{ } \sigma\text{-Alg.}}} \mathcal{B}$$

die von \mathcal{E} **erzeugte σ -Algebra**. Ist $\mathcal{A} = \sigma(\mathcal{E})$, so nennt man \mathcal{E} einen Erzeuger von \mathcal{A} .

Warnung: Der Erzeuger einer σ -Algebra ist nicht eindeutig.

Beispiel 1.2.5. ▶ Sei $\Omega \neq \emptyset$ und $\mathcal{E} = \{\{x\} : x \in \Omega\}$. Dann ist

$$\sigma(\mathcal{E}) = \{A \subseteq \Omega : A \text{ abzählbar oder } A^C \text{ abzählbar}\} =: \mathcal{B}.$$

Warum? Es gilt offensichtlich $\sigma(\mathcal{E}) \subseteq \mathcal{B}$ weil $\sigma(\mathcal{E})$ die kleinste σ -Algebra ist, die \mathcal{E} enthält und \mathcal{B} auch eine σ -Algebra ist, die \mathcal{E} enthält (siehe Beispiel 1.1.2). Es gilt aber auch $\mathcal{B} \subseteq \sigma(\mathcal{E})$ weil jede abzählbare Menge als abzählbare Vereinigung von einelementigen Mengen wieder zu $\sigma(\mathcal{E})$ gehört und auch Komplemente abzählbarer Mengen wieder in $\sigma(\mathcal{E})$ enthalten sind (Definition σ -Algebra).

Beispiel 1.2.6. ▶ **[Das Beispiel - die Borel- σ -Algebra]**

Wir kommen jetzt zu dem mit Abstand wichtigstem Beispiel einer σ -Algebra der Stochastik. Sei $\Omega = \mathbb{R}^d$, $\mathcal{E} = \{O \subseteq \mathbb{R}^d : O \text{ offen}\}$. Dann heißt $\sigma(\mathcal{E}) =: \mathcal{B}(\mathbb{R}^d)$ **Borel- σ -Algebra auf \mathbb{R}^d** . In $\mathcal{B}(\mathbb{R}^d)$ messbare Mengen heißen **Borelmengen**.

Übung: Die Borel- σ -Algebra hat viele verschiedene Erzeuger, z. B.

$$\begin{aligned} \mathcal{E}_2 &= \{K \subseteq \mathbb{R}^d : K \text{ kompakt}\}, \\ \mathcal{E}_3 &= \{Q \subseteq \mathbb{R}^d : Q \text{ Quader}\}, \\ \mathcal{E}_4 &= \{(a_1, b_1) \times \dots \times (a_d, b_d) : a_i, b_i \in \mathbb{R}\}, \\ \mathcal{E}_5 &= \{(-\infty, b_1] \times \dots \times (-\infty, b_n] : b_i \in \mathbb{R}\}, \\ \mathcal{E}_6 &= \{A \subseteq \mathbb{R}^d : A \text{ abgeschlossen}\}. \end{aligned}$$

Wir zeigen hier nur, dass $\sigma(\mathcal{E}_4) = \mathcal{B}(\mathbb{R}^d)$ gilt. Wie zeigt man nun allgemein für zwei Mengensysteme $\mathcal{E}, \mathcal{E}' \subseteq \mathcal{P}(\Omega)$, dass $\sigma(\mathcal{E}) = \sigma(\mathcal{E}')$ gilt? Indem man zeigt, dass

$$\mathcal{E} \subseteq \sigma(\mathcal{E}') \quad \text{sowie} \quad \mathcal{E}' \subseteq \sigma(\mathcal{E}) \quad (1.1)$$

gelten. Warum reicht das? Dazu nutzen wir zwei Eigenschaften, die direkt aus der Definition der erzeugten σ -Algebra folgen:

- (i) $\sigma(\sigma(\mathcal{E})) = \sigma(\mathcal{E})$
- (ii) $\mathcal{A} \subseteq \mathcal{A}' \Rightarrow \sigma(\mathcal{A}) \subseteq \sigma(\mathcal{A}')$

Aus (1.1) und (i), (ii) folgt

$$\sigma(\mathcal{E}) \subseteq \sigma(\sigma(\mathcal{E}')) = \sigma(\mathcal{E}')$$

sowie

$$\sigma(\mathcal{E}') \subseteq \sigma(\sigma(\mathcal{E})) = \sigma(\mathcal{E}),$$

also zusammen $\sigma(\mathcal{E}) = \sigma(\mathcal{E}')$. Nun zurück zu $\sigma(\mathcal{E}_4)$ und $\mathcal{B}(\mathbb{R}^d)$. Es ist klar, dass $\mathcal{E}_4 \subseteq \sigma(\{O : O \text{ offen}\})$, denn \mathcal{E}_4 enthält nur offene Mengen und die von einer Menge von Mengen erzeugte σ -Algebra enthält auch all die Mengen selbst. Umgekehrt existieren für jedes Element x einer offenen Menge O irgendwelche $a_1, \dots, a_d, b_1, \dots, b_d \in \mathbb{Q}$ mit $x \in (a_1, b_1) \times \dots \times (a_d, b_d) \subseteq O$. Damit gilt:

$$O = \bigcup_{\text{abz. viele}} (a_1, b_1) \times \dots \times (a_d, b_d) \in \sigma(\mathcal{E}_4)$$

und damit gilt $\{O : O \text{ offen}\} \subseteq \sigma(\mathcal{E}_4)$ weil abzählbar viele Vereinigungen von Mengen einer σ -Algebra wieder drin sind.

Definition 1.2.7. ▶ $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ heißt **Dynkin-System**, falls

- (i) $\Omega \in \mathcal{D}$
- (ii) $A \in \mathcal{D} \Rightarrow A^C \in \mathcal{D}$
- (iii) $A_1, A_2, \dots \in \mathcal{D}$ **paarweise disjunkt** $\Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{D}$

Im Gegensatz zu einer σ -Algebra ist ein Dynkin-System also nur abgeschlossen bezüglich paarweise disjunkter Vereinigungen. Das erinnert natürlich an die σ -Additivität von Maßen, woraus sich auch das wichtigste Beispiel eines Dynkin-Systems ergibt.

Beispiel 1.2.8. ▶

- Jede σ -Algebra ist ein Dynkin-System, die Definition einer σ -Algebra fordert mehr.
- Sind μ_1, μ_2 endliche Maße auf einem messbaren Raum (Ω, \mathcal{A}) mit $\mu_1(\Omega) = \mu_2(\Omega)$, so ist $\mathcal{M} = \{A \in \mathcal{A} : \mu_1(A) = \mu_2(A)\}$ ein Dynkin-System. Warum?
 - (i) Klar, wegen der Definition von Maßen.
 - (ii) Ist $A \in \mathcal{M}$, so gilt $\mu_1(A^C) = \mu_1(\Omega) - \mu_1(A) = \mu_2(\Omega) - \mu_2(A) = \mu_2(A^C)$ wegen der Rechenregel für Maße und der Annahme an μ_1, μ_2 . Damit ist auch $A^C \in \mathcal{M}$.
 - (iii) Seien $A_1, A_2, \dots \in \mathcal{M}$ paarweise disjunkt. Es gilt also $\mu_1(A_n) = \mu_2(A_n)$ für alle $n \geq 1$. Damit folgt wegen der σ -Additivität für Maße

$$\mu_1\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_1(A_n) = \sum_{n=1}^{\infty} \mu_2(A_n) = \mu_2\left(\bigcup_{n=1}^{\infty} A_n\right),$$

also ist $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$.

Proposition 1.2.9. ▶ Ein Dynkin-System \mathcal{D} ist eine σ -Algebra genau dann, wenn \mathcal{D} \cap -stabil ist (d. h. $A, B \in \mathcal{D} \Rightarrow A \cap B \in \mathcal{D}$).

Beweis.

(i) Es gilt $A, B \in \mathcal{D}$, $B \subseteq A \Rightarrow A \setminus B \in \mathcal{D}$, weil

$$A \setminus B = A \cap B^C = \underbrace{(A^C \cup B)}_{\substack{\in \mathcal{D}, \text{ weil disj.} \\ \in \mathcal{D}, \text{ weil Kompl.}}}^C.$$

(ii) Beliebige endliche Vereinigungen von Mengen aus \mathcal{D} sind in \mathcal{D} : Seien dazu $A, B \in \mathcal{D}$. Da $A \cap B \in \mathcal{D}$ per Annahme, gilt

$$A \cup B = A \cup (B \setminus (A \cap B)) \in \mathcal{D}.$$

Per Induktion bekommt man aus der Vereinigung zweier Mengen auch die Vereinigung endlich vieler Mengen.

(iii) Seien $A_1, A_2, \dots \in \mathcal{D}$. Definiere

$$B_n = \bigcup_{k=1}^n A_k \quad \text{sowie} \quad C_n = B_n \setminus B_{n-1}.$$

Weil nach (ii) $B_n \in \mathcal{D}$ und dann nach (i) $C_n \in \mathcal{D}$ gilt, folgt mit der Definition der Dynkin-Systeme

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} C_n \in \mathcal{D}.$$

Also ist \mathcal{D} abgeschlossen bezüglich abzählbarer Vereinigungen und damit ist \mathcal{D} eine σ -Algebra.

□

Vorlesung 3

Definition 1.2.10. ▶ Für ein Mengensystem $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ heißt

$$d(\mathcal{E}) = \bigcap_{\substack{\mathcal{E} \subseteq \mathcal{D}, \\ \mathcal{D} \text{ Dynk.-S.}}} \mathcal{D}$$

das \mathcal{E} erzeugtes Dynkin-System. Dass $d(\mathcal{E})$ das kleinste Dynkin-System ist das \mathcal{E} enthält, zeigt man genauso wie für σ -Algebren.

Satz 1.2.11. ▶ [Hauptsatz für Dynkin-Systeme]

Ist $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ \cap -stabil, so gilt $d(\mathcal{E}) = \sigma(\mathcal{E})$.

Beweis. Die Richtung $d(\mathcal{E}) \subseteq \sigma(\mathcal{E})$ folgt sofort, denn jede σ -Algebra ist auch immer ein Dynkin-System, folglich gilt

$$d(\mathcal{E}) = \bigcap_{\substack{\mathcal{E} \subseteq \mathcal{D}, \\ \mathcal{D} \text{ Dynk.-S.}}} \mathcal{D} \subseteq \bigcap_{\substack{\mathcal{E} \subseteq \mathcal{B}, \\ \mathcal{B} \sigma\text{-Alg.}}} \mathcal{B} = \sigma(\mathcal{E}).$$

Für die Richtung $\sigma(\mathcal{E}) \subseteq d(\mathcal{E})$ nehmen wir an, dass $d(\mathcal{E})$ \cap -stabil ist. Denn in diesem Fall folgt nach Proposition 1.2.9, dass $d(\mathcal{E})$ eine σ -Algebra ist. Weil dann aber $\mathcal{E} \subseteq d(\mathcal{E})$ gilt, muss die kleinste σ -Algebra (was gerade $\sigma(\mathcal{E})$ ist) Teilmenge von $d(\mathcal{E})$ sein.

Wir müssen also nur noch zeigen, dass auch $d(\mathcal{E})$ \cap -stabil, wenn \mathcal{E} \cap -stabil ist:

(a) Definiere dazu zunächst

$$\mathcal{D}_D = \{A \in \mathcal{P}(\Omega) : A \cap D \in d(\mathcal{E})\}$$

für beliebige $D \in d(\mathcal{E})$. Wir zeigen zunächst, dass \mathcal{D}_D ein Dynkin-System ist:

- (i) Weil per Annahme $D \in d(\mathcal{E})$ und $D = \Omega \cap D$ gilt, ist $\Omega \in \mathcal{D}_D$.
- (ii) Sei $A \in \mathcal{D}_D$. Damit auch $A^C \in \mathcal{D}_D$ gilt, zeigen wir $A^C \cap D \in d(\mathcal{E})$. Da Dynkin-Systeme abgeschlossen bezüglich disjunkter Vereinigung sind, folgt

$$A^C \cap D = (A \cup D^C)^C = \underbrace{((A \cap D) \cup D^C)^C}_{\in d(\mathcal{E})} \in d(\mathcal{E}).$$

(iii) Seien $A_1, A_2, \dots \in \mathcal{D}_D$ paarweise disjunkt, dann gilt

$$\bigcup_{n=1}^{\infty} A_n \cap D = \bigcup_{n=1}^{\infty} \underbrace{(A_n \cap D)}_{\in d(\mathcal{E})} \in d(\mathcal{E}).$$

- (b) Es gilt $d(\mathcal{E}) \subseteq \mathcal{D}_D$ für alle $D \in \mathcal{E}$. Warum? Sei $A \in \mathcal{E}$, so ist $A \cap D \in \mathcal{E}$, weil \mathcal{E} nach Annahme \cap -stabil ist. Damit ist $A \in \mathcal{D}_D$ und folglich $\mathcal{E} \subseteq \mathcal{D}_D$. Dann gilt aber auch $d(\mathcal{E}) \subseteq d(\mathcal{D}_D) \stackrel{(a)}{=} \mathcal{D}_D$ weil das von einem Dynkin-System \mathcal{D} erzeugte Dynkin-System gerade \mathcal{D} ist.
- (c) Des Weiteren gilt wegen (b) auch $\mathcal{E} \subseteq \mathcal{D}_D$ für alle $D \in d(\mathcal{E})$, d.h. $D \cap E \in d(\mathcal{E})$ für alle $E \in \mathcal{E}$.
- (d) Aus (c) und (a) folgt $d(\mathcal{E}) \subseteq d(\mathcal{D}_D) = \mathcal{D}_D$ für alle $D \in d(\mathcal{E})$. Das ist wegen Definition von \mathcal{D}_D gerade die \cap -Stabilität von $d(\mathcal{E})$.

□

Nun kommen wir zu der wesentlichen Anwendung von Dynkin-Systemen. Mit Dynkin-Systemen können wir ganz einfach zeigen, dass die Gleichheit von Maßen schon aus der Gleichheit auf einem schnittstabilen Erzeuger folgt. Wenn wir an die wahnsinnig große Borel- σ -Algebra denken, macht das ganze schnell Sinn. Es reicht nämlich die Gleichheit auf allen Intervallen zu zeigen, statt auf allen Borelmengen.

Satz 1.2.12. ▶ Sei (Ω, \mathcal{A}) ein messbarer Raum und \mathcal{E} ein \cap -stabiler Erzeuger von \mathcal{A} . Sind μ_1, μ_2 endliche Maße auf \mathcal{A} und es gelten

- $\mu_1(\Omega) = \mu_2(\Omega)$,
- $\mu_1(A) = \mu_2(A)$ für alle $A \in \mathcal{E}$,

so gilt $\mu_1(A) = \mu_2(A)$ für alle $A \in \mathcal{A}$, d. h. $\mu_1 = \mu_2$.

Beweis. Wir nutzten den **Trick der guten Mengen**. Dazu schreiben wir die Menge \mathcal{M} der Mengen hin, für die die Aussage gelten soll. Das sind die guten Mengen. Das Ziel ist also $\mathcal{M} = \mathcal{A}$ zu zeigen. Hierfür nutzen wir einen Dynkin-System Trick, den wir noch mehrmals sehen werden. Gezeigt haben wir schon, dass

$$\mathcal{M} = \{A \in \mathcal{A} : \mu_1(A) = \mu_2(A)\}$$

ein Dynkin-System ist. Nach Annahme ist $\mathcal{E} \subseteq \mathcal{M}$. Weil \mathcal{M} ein Dynkin-System ist, gilt $d(\mathcal{E}) \subseteq d(\mathcal{M}) = \mathcal{M} \subseteq \mathcal{A}$. Weil nach Annahme \mathcal{E} \cap -stabil ist, gilt nach dem Hauptsatz über Dynkin-Systeme $\sigma(\mathcal{E}) = d(\mathcal{E})$. Nach Annahme ist aber $\sigma(\mathcal{E}) = \mathcal{A}$. Alles zusammen ergibt

$$\mathcal{A} = \sigma(\mathcal{E}) = d(\mathcal{E}) \subseteq d(\mathcal{M}) = \mathcal{M} \subseteq \mathcal{A}.$$

Also gilt $\mathcal{M} = \mathcal{A}$ und das ist die Aussage des Korollars.

□

Wir schauen uns noch einen Trick an, die Endlichkeitsannahme aus Satz 1.2.12 abzuschwächen.

Satz 1.2.13. ▶ **[Eindeutigkeitsatz]** Es sei (Ω, \mathcal{A}) ein messbarer Raum, \mathcal{E} ein \cap -stabiler Erzeuger von \mathcal{A} und μ_1, μ_2 seien Maße auf \mathcal{A} . Zudem gelten:

- (i) Es gibt eine Folge $(E_n) \subseteq \mathcal{E}$ mit $E_n \uparrow \Omega$, $n \rightarrow \infty$, und $\mu_i(E_n) < \infty$ für alle $n \in \mathbb{N}$, $i = 1, 2$.
- (ii) $\mu_1(A) = \mu_2(A)$ für alle $A \in \mathcal{E}$.

Dann gilt $\mu_1 = \mu_2$, d. h. $\mu_1(A) = \mu_2(A)$ für alle $A \in \mathcal{A}$.

Geben wir der genutzten Erweiterung endlicher Maße einen Namen:

Definition 1.2.14. ▶ Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und es gibt eine Folge $(E_n) \subseteq \mathcal{A}$ mit $E_n \uparrow \Omega$, $n \rightarrow \infty$, und $\mu(E_n) < \infty$ für alle $n \in \mathbb{N}$, so nennt man μ ein **σ -endliches Maß**.

Die meisten Sätze für endliche Maße lassen sich mit dem Trick des folgenden Beweises auf σ -endliche Maße ausdehnen. Beispiele folgen noch.

Beweis von Satz 1.2.13. Definiere dazu für $A \in \mathcal{A}$ und $n \in \mathbb{N}$

$$\begin{aligned}\mu_1^n(A) &:= \mu_1(A \cap E_n), \\ \mu_2^n(A) &:= \mu_2(A \cap E_n).\end{aligned}$$

Man rechnet sofort nach, dass auch die μ_i^n wieder Maße auf \mathcal{A} sind. Des Weiteren sind μ_1^n, μ_2^n endlich, weil $\mu_i^n(\Omega) = \mu_i(\Omega \cap E_n) = \mu_i(E_n) \stackrel{\text{Ann.}}{<} \infty$. Nach Satz 1.2.12 gilt $\mu_1^n = \mu_2^n$ für alle $n \in \mathbb{N}$. Nun gilt wegen Stetigkeit von Maßen

$$\begin{aligned}\mu_1(A) &= \mu_1(A \cap \Omega) = \mu_1\left(A \cap \bigcup_{n=1}^{\infty} E_n\right) = \mu_1\left(\bigcup_{n=1}^{\infty} (A \cap E_n)\right) \stackrel{1.1.14}{=} \lim_{n \rightarrow \infty} \mu_1(A \cap E_n) \\ &\stackrel{\text{Def.}}{=} \lim_{n \rightarrow \infty} \mu_1^n(A) = \lim_{n \rightarrow \infty} \mu_2^n(A) \stackrel{\text{Def.}}{=} \lim_{n \rightarrow \infty} \mu_2(A \cap E_n) \stackrel{1.1.14}{=} \mu_2\left(A \cap \bigcup_{n=1}^{\infty} E_n\right) = \mu_2(A).\end{aligned}$$

□

So, endlich ein richtig konkretes Beispiel!

Beispiel 1.2.15. ▶ Sei \mathbb{P} ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R})$. Dann heißt

$$F_{\mathbb{P}}(t) := \mathbb{P}((-\infty, t]), \quad t \in \mathbb{R},$$

die **Verteilungsfunktion** von \mathbb{P} . $F_{\mathbb{P}}$ erfüllt folgende Eigenschaften:

- $0 \leq F_{\mathbb{P}} \leq 1$,
- $F_{\mathbb{P}}$ ist nicht fallend,
- $\lim_{t \rightarrow +\infty} F_{\mathbb{P}}(t) = 1$,
- $\lim_{t \rightarrow -\infty} F_{\mathbb{P}}(t) = 0$.

Die ersten beiden Eigenschaften folgen aus der Definition von Wahrscheinlichkeitsmaßen und der Monotonie von Maßen. Die weiteren Eigenschaften folgen aus der Stetigkeit von Maßen. Um die gerade bewiesenen Sätze anzuwenden zeigen wir folgende Behauptung:

$$F_{\mathbb{P}_1}(t) = F_{\mathbb{P}_2}(t) \quad \text{für alle } t \in \mathbb{R} \quad \implies \quad \mathbb{P}_1 = \mathbb{P}_2.$$

Das folgt aus Satz 1.2.12 mit $\mathcal{E} = \{(-\infty, t] : t \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R})$. Checken wir dazu die benötigten Eigenschaften:

- $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$ ist aus den Übungen bekannt,
- \mathcal{E} ist \cap -stabil, denn $(-\infty, s] \cap (-\infty, t] = (-\infty, \min\{s, t\}]$, $s, t \in \mathbb{R}$,
- $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ für alle $A \in \mathcal{E}$ weil das gerade die Verteilungsfunktionen sind.

Genauso beweist man auch die Aussage des nächsten Beispiels:

Beispiel 1.2.16. ▶ Seien μ_1, μ_2 σ -endliche Maße auf $\mathcal{B}(\mathbb{R}^d)$ mit einer der folgenden Eigenschaften:

$$\begin{aligned} \mu_1(Q) &= \mu_2(Q) \text{ für alle Quader } Q, \\ \mu_1(K) &= \mu_2(K) \text{ für alle kompakten Mengen } K, \\ \mu_1(O) &= \mu_2(O) \text{ für alle offenen Mengen } O, \\ \mu_1(A) &= \mu_2(A) \text{ für alle abgeschlossenen Mengen } A. \end{aligned}$$

Dann gilt $\mu_1 = \mu_2$, die Maße stimmen also auf allen Borelmengen überein.

1.3 Konstruktion von Maßen

Gerade haben wir gesehen, dass zwei endliche Maße auf $\mathcal{B}(\mathbb{R})$ schon auf den Intervallen eindeutig festgelegt sind (sind μ_1, μ_2 gleich auf Intervallen, sind μ_1, μ_2 gleich auf $\mathcal{B}(\mathbb{R})$). Das ist eine Eindeutigkeitsaussage. Jetzt drehen wir das ganze um und untersuchen Existenzaussagen. Am Ende soll folgendes rauskommen: Wenn wir eine „geeignete Mengenfunktion“ auf den Intervallen definieren, dann gibt es auch ein passendes Maß auf $\mathcal{B}(\mathbb{R})$. Dazu brauchen wir einiges an Handwerkszeugs.

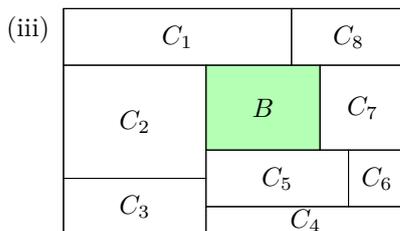
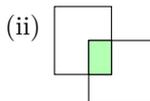
Definition 1.3.1. ▶ $\mathcal{S} \subseteq \mathcal{P}(\Omega)$ heißt **Semiring**, falls

- (i) $\emptyset \in \mathcal{S}$
- (ii) $A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S}$, also ist \mathcal{S} „ \cap -stabil“
- (iii) $A, B \in \mathcal{S} \Rightarrow$ es gibt paarweise disjunkte Mengen $C_1, \dots, C_m \in \mathcal{S}$ mit $A \setminus B = \bigcup_{k=1}^m C_k$.

Die Definition ist etwas komisch, verallgemeinert aber einfach nur das folgendes Beispiel, dass wir immer im Kopf halten und später auch hauptsächlich nutzen:

Beispiel. $\mathcal{Q} := \{Q \subseteq \mathbb{R}^d : Q \text{ Quader}\}$ ist ein Semiring. Checken wir anschaulich die definierenden Eigenschaften, vernünftige Argumente haben wir in Analysis II gesehen (Stichwort Quaderwahnsinn):

- (i) $\emptyset \in \mathcal{Q}$ weil $\emptyset = (a_1, a_1) \times \dots \times (a_d, a_d)$ für beliebige $a_1, \dots, a_d \in \mathbb{R}$.



Aus der zweiten Eigenschaft eines Semirings kann man durch Induktion sofort Abgeschlossenheit bezüglich endlich vielen Schnitten zeigen. Probiert's einfach aus, ist zu einfach als Übungsaufgabe. Auch die letzte Eigenschaft kann man verallgemeinern. Bei Quadern ist das anschaulich klar: Wenn man aus einem Quader endlich viele Quader entfernt, bleibt eine disjunkte Vereinigung von Quadern übrig. Mit Induktion kriegen wir das auch für allgemeine Semiringe hin:

Lemma 1.3.2.  [Eine kleine Indexschlacht]

Es gilt in einem Semiring auch (iii)': Sind $B_1, \dots, B_r, A \in \mathcal{S}$ mit B_1, B_2, \dots, B_r paarweise disjunkt, so existieren $C_1, \dots, C_n \in \mathcal{S}$ paarweise disjunkt mit

$$A \setminus (B_1 \cup \dots \cup B_r) = \bigcup_{k=1}^n C_k.$$

Beweis. Vollständige Induktion bezüglich r .

IA: Für $A \setminus B_1$ folgt das direkt aus der Definition des Rings.

IV: Die Behauptung gelte für ein beliebiges aber festes $r \in \mathbb{N}$.

IS: Seien nun $B_1, B_2, \dots, B_{r+1} \subseteq A$ paarweise disjunkt. Nach Induktionsvoraussetzung und Rechenregeln mit Schnitten von Mengen gilt mit Umklammern von Schnitten und Vereinigungen

$$\begin{aligned} A \setminus \bigcup_{i=1}^{r+1} B_i &= A \cap \bigcap_{i=1}^{r+1} B_i^C \\ &= \left(A \setminus \bigcup_{i=1}^r B_i \right) \cap B_{r+1}^C \\ &\stackrel{\text{IV}}{=} \left(\bigcup_{k=1}^{n_r} C_{r,k} \right) \cap B_{r+1}^C = \bigcup_{k=1}^{n_r} [C_{r,k} \setminus B_{r+1}]. \end{aligned}$$

Nun existieren für alle $1 \leq k \leq n_r$ jeweils endlich viele paarweise disjunkte Mengen $C_{r,k,1}, \dots, C_{r,k,l_{r,k}} \in \mathcal{S}$, mit

$$C_{r,k} \setminus B_{r+1} = \bigcup_{m=1}^{l_{r,k}} C_{r,k,m}.$$

Also gilt

$$A \setminus \bigcup_{i=1}^{r+1} B_i = \bigcup_{k=1}^{n_r} \bigcup_{m=1}^{l_{r,k}} C_{r,k,m}.$$

Da die endlich vielen Mengen $C_{r,k,m}$ paarweise disjunkt sind, haben wir die gewünschte Darstellung für $A \setminus (B_1 \cup \dots \cup B_{r+1})$ gefunden. □

Vorlesung 4

Es kommen jetzt ein paar schmerzhaft vorbereitungen für den wichtigsten Satz dieses Abschnitts, den Fortsetzungssatz von Carathéodory. Wir werden zunächst zeigen, dass die Monotonie und Subadditivität (siehe Lemma 1.1.9) auch für „Maße“ auf Semiringen gilt. Warum steht hier Maß in Anführungsstrichen? Per Definition ist ein Maß immer auf einer σ -Algebra definiert, der Begriff macht also auf Semiringen gar keinen Sinn. Die genaue Aussage ist also, dass eine Mengenfunktion auf Semiringen mit Eigenschaften die einem Maß ähneln, ebenfalls Monotonie und Subadditivität erfüllen.

Lemma 1.3.3.  [Eine ziemlich große Indexschlacht]

Sei \mathcal{S} ein Semiring und $\mu: \mathcal{S} \rightarrow [0, \infty]$ eine Mengenfunktion mit

- $\mu(\emptyset) = 0$

- μ ist σ -**additiv** (d.h. sind $A_1, A_2, \dots \in \mathcal{S}$ paarweise disjunkt mit $A := \bigcup_{k=1}^{\infty} A_k \in \mathcal{S}$, so gilt $\mu(A) = \sum_{k=1}^{\infty} \mu(A_k)$).

Dann gilt

- (i) Monotonie: $\mu(A) \leq \mu(B)$ für alle $A, B \in \mathcal{S}$ mit $A \subseteq B$.
- (ii) „**Subadditivität**“: Sind $A, A_1, A_2, \dots \in \mathcal{S}$ und $A \subseteq \bigcup_{k=1}^{\infty} A_k$, so gilt

$$\mu(A) \leq \sum_{k=1}^{\infty} \mu(A_k).$$

Man beachte, dass die Eigenschaften der Additivität etwas komisch sind. Da wir nicht fordern, dass Semiringe abgeschlossen bezüglich Vereinigungen sind (sie sind es auch meistens nicht, man denke nur an den Semiring der Quader), muss immer gefordert werden, dass die Vereinigungen wieder in \mathcal{S} liegen. Sonst wäre $\mu(A)$ schließlich gar nicht definiert! Auch zu beachten ist, dass Vereinigungen und Komplemente nicht automatisch in \mathcal{S} liegen. Daher sind einfache Eigenschaften für Maße auf σ -Algebren, nicht so einfach für Mengenfunktionen auf Semiringen.

Beweis.

- (i) Es gibt wegen der Eigenschaften eines Semirings Mengen $C_1, \dots, C_m \in \mathcal{S}$ mit

$$B \setminus A = \bigcup_{k=1}^m C_k.$$

Damit gilt wegen der geforderten Additivität von μ

$$\begin{aligned} \mu(B) &= \mu(A \cup (B \setminus A)) = \mu(A \cup C_1 \cup \dots \cup C_m) = \mu(A) + \mu(C_1) + \dots + \mu(C_m) \\ &\geq \mu(A). \end{aligned}$$

- (ii) Erst machen wir (wie immer!) die A_n disjunkt:

$$\begin{aligned} A'_1 &:= A_1, \\ A'_2 &:= A_2 \setminus A'_1, \\ A'_n &:= A_n \setminus (A'_1 \cup \dots \cup A'_{n-1}), \quad n \geq 3. \end{aligned}$$

Beachte: Die A'_n müssen nicht in \mathcal{S} sein. Weil die A'_n die Form $A_n \setminus \dots$ haben, gibt es wegen Lemma 1.3.2 allerdings paarweise disjunkte $C_{n,j} \in \mathcal{S}$ mit

$$A'_n = \bigcup_{j=1}^{l_n} C_{n,j} \tag{1.2}$$

für alle $n \in \mathbb{N}$. Wieder gibt es wegen Lemma 1.3.2 Mengen $D_{n,k} \in \mathcal{S}$ mit

$$A_n \setminus A'_n = \bigcup_{k=1}^{m_n} D_{n,k}. \tag{1.3}$$

Damit gelten

- $A \cap C_{n,j} \in \mathcal{S}$ weil \mathcal{S} \cap -stabil ist,
-

$$A = \bigcup_{n=1}^{\infty} A'_n \cap A = \bigcup_{n=1}^{\infty} \bigcup_{j=1}^{l_n} A \cap C_{n,j}$$

weil $A \subseteq \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A'_n$ angenommen wurde,

•

$$A_n = A'_n \cup (A_n \setminus A'_n) \stackrel{(1.2)}{\stackrel{(1.3)}{=}} \bigcup_{j=1}^{l_n} C_{n,j} \cup \bigcup_{k=1}^{m_n} D_{n,k}.$$

Alles zusammen ergibt die Subadditivität:

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_{n=1}^{\infty} \bigcup_{j=1}^{l_n} \underbrace{A \cap C_{n,j}}_{\substack{\in \mathcal{S}, \text{ weil } \mathcal{S} \\ \cap\text{-stabil}}}\right) \\ &\stackrel{\sigma\text{-add.}}{=} \sum_{n=1}^{\infty} \sum_{j=1}^{l_n} \mu(A \cap C_{n,j}) \\ &\stackrel{\text{Monotonie}}{\leq} \sum_{n=1}^{\infty} \sum_{j=1}^{l_n} \mu(C_{n,j}) \\ &\stackrel{\mu \geq 0}{\leq} \sum_{n=1}^{\infty} \left(\sum_{j=1}^{l_n} \mu(C_{n,j}) + \sum_{k=1}^{m_n} \mu(D_{n,k}) \right) \\ &\stackrel{\sigma\text{-add.}}{=} \sum_{n=1}^{\infty} \mu\left(\bigcup_{j=1}^{l_n} C_{n,j} \cup \bigcup_{k=1}^{m_n} D_{n,k}\right) \\ &= \sum_{n=1}^{\infty} \mu(A_n). \end{aligned}$$

□

Definition 1.3.4. ▶ $\mu^* : \mathcal{P}(\Omega) \rightarrow [0, \infty]$ heißt **äußeres Maß**, falls

- (i) $\mu^*(\emptyset) = 0$
- (ii) $A \subseteq B \subseteq \Omega \Rightarrow \mu^*(A) \leq \mu^*(B)$
- (iii) $A_1, A_2, \dots \subseteq \Omega \Rightarrow \mu^*\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mu^*(A_k)$

Für den Moment bleibt es unklar, weshalb wir dieser abstrakten Definition den Namen „äußeres Maß“ geben. Das wird aber in dem Beweis des Carathéodory Fortsetzungssatzes klar werden. Das dort definierte äußere Maß hat eine klare Interpretation.

Definition 1.3.5. ▶ Sei μ^* ein äußeres Maß auf Ω . Dann heißt $A \subseteq \Omega$ **μ^* -messbare Menge**, falls für alle $Z \subseteq \Omega$

$$\mu^*(Z) = \mu^*(Z \cap A) + \mu^*(Z \cap A^C)$$

gilt. Die Menge der μ^* -messbaren Mengen heißt \mathcal{A}_{μ^*} .

Proposition 1.3.6. ▶

- (i) \mathcal{A}_{μ^*} ist eine σ -Algebra.
- (ii) μ^* eingeschränkt auf \mathcal{A}_{μ^*} ist ein Maß.

Beweis. Wir zeigen nacheinander (a) \mathcal{A}_{μ^*} ist eine σ -Algebra und (b) μ^* ist ein Maß auf \mathcal{A}_{μ^*} .

- (a) Zu zeigen sind die definierenden Eigenschaften einer σ -Algebra:

(i) Sei $Z \subseteq \Omega$. Dann gilt

$$\mu^*(Z) = \mu^*(Z) + 0 = \mu^*(Z \cap \Omega) + \underbrace{\mu^*(Z \cap \Omega^C)}_{= \emptyset},$$

also ist $\emptyset \in \mathcal{A}_{\mu^*}$ gezeigt.

- (ii) Sei $A \in \mathcal{A}_{\mu^*}$, dann erfüllt (+ kommutieren und $(A^C)^C = A$ nutzen) auch A^C die definierende Eigenschaft von \mathcal{A}_{μ^*} . Also ist \mathcal{A}_{μ^*} abgeschlossen unter Komplementbildung.
- (iii) Wir zeigen zunächst die Abgeschlossenheit bezüglich Vereinigungen von zwei Mengen. Seien also $A_1, A_2 \in \mathcal{A}_{\mu^*}$. Sei $Z \subseteq \Omega$ beliebig, dann folgt mit $Z' := Z \cap A_2$

$$\begin{aligned} \mu^*(Z \cap (A_1 \cup A_2)) &\stackrel{A_1 \in \mathcal{A}_{\mu^*}}{=} \mu^*(Z \cap (A_1 \cup A_2) \cap A_1) + \mu^*(Z \cap (A_1 \cup A_2) \cap A_1^C) \\ &= \mu^*(Z \cap A_1) + \mu^*(Z \cap A_2 \cap A_1^C). \end{aligned}$$

Folglich gilt auch

$$\begin{aligned} &\mu^*(Z \cap (A_1 \cup A_2)) + \mu^*(Z \cap (A_1 \cup A_2)^C) \\ &= \mu^*(Z \cap A_1) + \mu^*(Z \cap A_2 \cap A_1^C) + \mu^*(Z \cap (A_1 \cup A_2)^C) \\ &= \mu^*(Z \cap A_1) + \underbrace{\mu^*(Z \cap A_1^C \cap A_2)}_{=: Z''} + \underbrace{\mu^*(Z \cap A_1^C \cap A_2^C)}_{=: Z''} \\ &\stackrel{A_2 \in \mathcal{A}_{\mu^*}}{=} \mu^*(Z \cap A_1) + \mu^*(Z \cap A_1^C) \\ &\stackrel{A_1 \in \mathcal{A}_{\mu^*}}{=} \mu^*(Z) \end{aligned}$$

und damit ist $A_1 \cup A_2 \in \mathcal{A}_{\mu^*}$.

- (iv) Per Induktion folgt aus (iii) die Abgeschlossenheit bezüglich Vereinigungen endlich vieler Mengen.
- (v) Es fehlt jetzt noch die Abgeschlossenheit bezüglich abzählbar unendlicher Vereinigungen. Seien also $A_1, A_2, \dots \in \mathcal{A}_{\mu^*}$, zu zeigen ist $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_{\mu^*}$. Zuerst nutzen wir den schon bekannten Trick, der uns erlaubt, ohne Einschränkung der Allgemeinheit anzunehmen, dass die Mengen diskjunkt sind. Dazu definieren wir die paarweise diskjunkten Mengen

$$A'_1 = A_1 \quad \text{und} \quad A'_n = A_n \setminus (A'_1 \cup \dots \cup A'_{n-1}), \quad n \geq 2,$$

und beachten, dass damit $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A'_n$ gilt. Wenn also die Vereinigung der diskjunkten A'_n wieder in \mathcal{A}_{μ^*} ist, ist auch die Vereinigung über die A_n in \mathcal{A}_{μ^*} . Es reicht also die Aussage für disjunkte Mengen zu beweisen. Damit die Rechnungen lesbarer bleiben, nehmen wir also ohne Beschränkung der Allgemeinheit an, dass die Mengen A_1, A_2, \dots paarweise disjunkt sind, so sparen wir uns die ' in den Gleichungen. Aufgrund der Definition von \mathcal{A}_{μ^*} wählen wir ein $Z \subseteq \Omega$ beliebig. Wir zeigen erstmal induktiv

$$\mu^*(Z) = \sum_{k=1}^n \mu^*(Z \cap A_k) + \mu^*\left(Z \cap \bigcap_{k=1}^n A_k^C\right), \quad \forall n \in \mathbb{N}. \quad (1.4)$$

IA: Für $n = 1$ gilt die Behauptung, weil $A_1 \in \mathcal{A}_{\mu^*}$.

IV: Es gelte (1.4) für ein beliebiges, aber festes $n \in \mathbb{N}$.

IS: Eine kleine Runde Kampfrechnen mit Mengen. Weil nach Annahme $A_{n+1} \in \mathcal{A}_{\mu^*}$ und die A_n paarweise disjunkt sind, gilt

$$\begin{aligned} \mu^*\left(Z \cap \bigcap_{k=1}^n A_k^C\right) &= \mu^*\left(Z \cap \bigcap_{k=1}^n A_k^C \cap A_{n+1}\right) + \mu^*\left(Z \cap \bigcap_{k=1}^n A_k^C \cap A_{n+1}^C\right) \\ &= \mu^*(Z \cap A_{n+1}) + \mu^*\left(Z \cap \bigcap_{k=1}^{n+1} A_k^C\right). \end{aligned}$$

Einsetzen in die Induktionsvoraussetzung gibt

$$\mu^*(Z) = \sum_{k=1}^{n+1} \mu^*(Z \cap A_k) + \mu^*\left(Z \cap \bigcap_{k=1}^{n+1} A_k^C\right).$$

Damit ist der Induktionsschritt gezeigt.

Zurück zur Vereinigung: Wegen der angenommenen Monotonie von μ^* folgt aus (1.4)

$$\mu^*(Z) \geq \sum_{k=1}^n \mu^*(Z \cap A_k) + \mu^*\left(Z \cap \bigcap_{k=1}^{\infty} A_k^C\right), \quad \forall n \in \mathbb{N}.$$

Mit $n \rightarrow \infty$ folgt (Monotonie von Grenzwertbildung aus Analysis 1)

$$\begin{aligned} \mu^*(Z) &\geq \sum_{k=1}^{\infty} \mu^*(Z \cap A_k) + \mu^*\left(Z \cap \bigcap_{k=1}^{\infty} A_k^C\right) \\ &\geq \mu^*\left(\bigcup_{k=1}^{\infty} A_k \cap Z\right) + \mu^*\left(Z \cap \left(\bigcup_{k=1}^{\infty} A_k\right)^C\right) \\ &\geq \mu^*\left(\left(Z \cap \bigcup_{k=1}^{\infty} A_k\right) \cup \left(Z \cap \left(\bigcup_{k=1}^{\infty} A_k\right)^C\right)\right) \\ &= \mu^*\left(Z \cap \underbrace{\left(\bigcup_{k=1}^{\infty} A_k \cup \left(\bigcup_{k=1}^{\infty} A_k\right)^C\right)}_{\Omega}\right) = \mu^*(Z). \end{aligned} \tag{1.5}$$

Für die letzten beiden Ungleichungen haben wir die angenommene Subadditivität (Ungleichung andersrum als üblich) genutzt. Weil die linke und rechte Seite der Kette von Ungleichungen identisch sind, sind die Ungleichungen alles Gleichungen, also gilt

$$\mu^*(Z) = \mu^*\left(Z \cap \bigcup_{k=1}^{\infty} A_k\right) + \mu^*\left(Z \cap \left(\bigcup_{k=1}^{\infty} A_k\right)^C\right).$$

Weil Z beliebig war, ist damit

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{A}_{\mu^*}$$

aufgrund der Definition von \mathcal{A}_{μ^*} . Damit ist die Abgeschlossenheit bezüglich abzählbarer Vereinigungen gezeigt und folglich ist \mathcal{A}_{μ^*} eine σ -Algebra.

(b) Aufgrund der Gleichheiten in (1.5) gilt auch

$$\mu^*(Z) = \sum_{k=1}^{\infty} \mu^*(Z \cap A_k) + \mu^*\left(Z \cap \bigcap_{k=1}^{\infty} A_k^C\right).$$

Wählen wir $Z = \bigcup_{k=1}^{\infty} A_k$, so gilt wegen $\mu(\emptyset) = 0$

$$\mu^*\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu^*(A_k) + 0$$

und das ist gerade die σ -Additivität von μ^* . Also ist μ^* auch ein Maß auf der σ -Algebra \mathcal{A}_{μ^*} .

□

Kommen wir endlich zum Höhepunkt der ersten Wochen. Nach dem Beweis sind wir auch endlich mitten in der Stochastik angelangt! Der Beweis enthält tatsächlich viele Informationen. Insbesondere wird klar, wo der Begriff äußeres Maß herkommt.

Satz 1.3.7.  [Fortsetzungssatz von Carathéodory]

Sei \mathcal{S} ein Semiring und $\mu: \mathcal{S} \rightarrow [0, \infty]$ eine Mengenfunktion mit

- $\mu(\emptyset) = 0$,
- μ ist σ -**additiv** (d.h. sind $A_1, A_2, \dots \in \mathcal{S}$ paarweise disjunkt mit $A := \bigcup_{k=1}^{\infty} A_k \in \mathcal{S}$, so gilt $\mu(A) = \sum_{k=1}^{\infty} \mu(A_k)$).

Dann existiert ein Maß $\bar{\mu}$ auf $\sigma(\mathcal{S})$ mit $\mu(A) = \bar{\mu}(A)$ für alle $A \in \mathcal{S}$.

Man sagt, dass die Mengenfunktion μ von \mathcal{S} nach $\sigma(\mathcal{S})$ „fortgesetzt“ wird.

Beweis. Für $A \in \mathcal{P}(\Omega)$ definieren wir

$$\mu^*(A) = \inf \left\{ \sum_{k=1}^{\infty} \mu(A_k) : A_1, A_2, \dots \in \mathcal{S} \text{ mit } A \subseteq \bigcup_{k=1}^{\infty} A_k \right\}.$$

Wir zeigen nacheinander (a) μ^* ist ein äußeres Maß, (b) $\mathcal{S} \subseteq \mathcal{A}_{\mu^*}$ und (c) $\mu^*(A) = \mu(A)$ für alle $A \in \mathcal{S}$. Der Beweis ist dann vollendet, weil nach dem vorherigen Satz \mathcal{A}_{μ^*} eine σ -Algebra ist und μ^* ein Maß auf \mathcal{A}_{μ^*} ist. Wegen (b) gilt $\sigma(\mathcal{S}) \subseteq \mathcal{A}_{\mu^*}$, weil die kleinste σ -Algebra die \mathcal{S} enthält, auch Teilmenge von allen σ -Algebren ist, die \mathcal{S} enthalten. Damit ist auch die Einschränkung von μ^* auf $\sigma(\mathcal{S})$ ein Maß (wir setzen dann $\bar{\mu} := \mu^*|_{\sigma(\mathcal{S})}$) und wegen (c) ist $\bar{\mu}$ eine Fortsetzung von μ .

(a) Wir checken die definierenden Eigenschaften eines äußeren Maßes:

- $\mu^*(\emptyset) = 0$ ist klar weil $\emptyset \in \mathcal{S}$ und $\mu(\emptyset) = 0$.
- Monotonie folgt direkt aus der Definition.

(iii) Nun zur Subadditivität. Seien dazu $A_1, A_2, \dots \in \mathcal{P}(\Omega)$ und $A = \bigcup_{k=1}^{\infty} A_k$. Wir können annehmen, dass $\mu^*(A_k) < \infty$ für alle $k \in \mathbb{N}$ gilt (sonst gilt die Ungleichung sowieso). Sei nun $\varepsilon > 0$ beliebig. Für jedes $k \in \mathbb{N}$ existiert qua Definition (Infimum=größte untere Schranke) eine Folge von Mengen $A_{k,1}, A_{k,2}, \dots \in \mathcal{S}$ mit

$$A_k \subseteq \bigcup_{j=1}^{\infty} A_{k,j} \quad \text{und} \quad \sum_{j=1}^{\infty} \mu(A_{k,j}) \leq \mu^*(A_k) + \frac{\varepsilon}{2^k}.$$

Wem das nicht klar ist, der schaue bitte in den Analysis 1 Mitschrieb! Weil

$$A \stackrel{\text{Def.}}{=} \bigcup_{k=1}^{\infty} A_k \subseteq \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{\infty} A_{k,j}$$

gilt, folgt (das Infimum einer Menge ist kleiner gleich jedem Element der Menge)

$$\mu^*(A) \stackrel{\text{Def.}}{\leq} \mu^* \left(\bigcup_{k=1}^{\infty} \bigcup_{j=1}^{\infty} A_{k,j} \right) \leq \sum_{k=1}^{\infty} \left(\mu^*(A_k) + \frac{\varepsilon}{2^k} \right) = \sum_{k=1}^{\infty} \mu^*(A_k) + \underbrace{\sum_{k=1}^{\infty} \frac{\varepsilon}{2^k}}_{\varepsilon}.$$

Weil ε beliebig gewählt wurde, gilt damit die Subadditivität von μ^* .

- (b) Wir zeigen $\mathcal{S} \subseteq \mathcal{A}_{\mu^*}$. Dazu müssen wir $\mu^*(Z) = \mu^*(Z \cap S) + \mu^*(Z \cap S^C)$ für alle $Z \subseteq \Omega$ und $S \in \mathcal{S}$ nachrechnen.

„ \leq “: $\mu^*(Z) = \mu^*(Z \cap S \cup Z \cap S^C) \leq \mu^*(Z \cap S) + \mu^*(Z \cap S^C)$ aufgrund der gezeigten Subadditivität von μ^* .

„ \geq “: Sei $A_1, A_2, \dots \subseteq \mathcal{S}$ mit $Z \subseteq \bigcup_{k=1}^{\infty} A_k$. Weil \mathcal{S} ein Semiring ist, existieren $C_{k,1}, \dots, C_{k,m_k} \in \mathcal{S}$ mit

$$A_k \cap S^C = A_k \setminus S = \bigcup_{j=1}^{m_k} C_{k,j}.$$

Es gelten

$$Z \cap S \subseteq \left(\bigcup_{k=1}^{\infty} A_k \right) \cap S = \bigcup_{k=1}^{\infty} (A_k \cap S)$$

sowie analog

$$Z \cap S^C \subseteq \left(\bigcup_{k=1}^{\infty} A_k \right) \cap S^C = \bigcup_{k=1}^{\infty} (A_k \cap S^C) = \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{m_k} C_{k,j}.$$

Mit den Definitionen folgt

$$\begin{aligned} \mu^*(Z \cap S) + \mu^*(Z \cap S^C) &\leq \sum_{k=1}^{\infty} \left(\mu(A_k \cap S) + \sum_{j=1}^{m_k} \mu(C_{k,j}) \right) \\ &\stackrel{\mu \text{ } \sigma\text{-add.}}{=} \sum_{k=1}^{\infty} \mu \left((A_k \cap S) \cup \bigcup_{j=1}^{m_k} C_{k,j} \right) \\ &= \sum_{k=1}^{\infty} \mu \left((A_k \cap S) \cup (A_k \cap S^C) \right) \\ &= \sum_{k=1}^{\infty} \mu(A_k). \end{aligned}$$

Daraus folgt $\mu^*(Z \cap S) + \mu^*(Z \cap S^C) \leq \mu^*(Z)$, weil

$$\mu^*(Z) = \inf \left\{ \sum_{k=1}^{\infty} \mu(A_k) : A_1, A_2, \dots \in \mathcal{S} \text{ mit } Z \subseteq \bigcup_{k=1}^{\infty} A_k \right\}.$$

Somit ist „ \geq “ gezeigt. Also ist jedes $S \in \mathcal{A}_{\mu^*}$ und damit gilt $\mathcal{S} \subseteq \mathcal{A}_{\mu^*}$.

- (c) Fehlt noch $\mu^*(A) = \mu(A)$ für alle $A \in \mathcal{S}$. Im Prinzip ist das Lemma 1.3.3.

„ \leq “:

$$\mu^*(A) = \inf \left\{ \sum_{k=1}^{\infty} \mu(A_k) : A_1, A_2, \dots \in \mathcal{S} \text{ mit } A \subseteq \bigcup_{k=1}^{\infty} A_k \right\} \leq \mu(A)$$

für alle $A \in \mathcal{S}$.

„ \geq “: Ist $A \subseteq \bigcup_{n=1}^{\infty} A_n$ für $A_1, A_2, \dots \in \mathcal{S}$, so gilt

$$\mu(A) \stackrel{1.3.3}{\leq} \sum_{n=1}^{\infty} \mu(A_n).$$

Folglich gilt

$$\mu(A) \leq \inf \left\{ \sum_{k=1}^{\infty} \mu(A_k) : A_1, A_2, \dots \in \mathcal{S} \text{ mit } A \subseteq \bigcup_{k=1}^{\infty} A_k \right\} = \mu^*(A).$$

□

Satz 1.3.8.  [Existenz und Eindeutigkeit von Maßen]Ist (Ω, \mathcal{A}) ein messbarer Raum, \mathcal{E} ein \cap -stabiler Semiring mit $\sigma(\mathcal{E}) = \mathcal{A}$. Sei $\mu: \mathcal{E} \rightarrow [0, \infty]$ mit

- $\mu(\emptyset) = 0$
- μ ist σ -additiv
- es gibt Folge eine $E_1, E_2, \dots \in \mathcal{E}$ mit $E_n \uparrow \Omega$ und $\mu(E_n) < \infty$ für alle $n \in \mathbb{N}$.

Dann existiert genau ein Maß $\bar{\mu}$ auf $\mathcal{A} = \sigma(\mathcal{E})$, so dass $\bar{\mu}(A) = \mu(A)$ für alle $A \in \mathcal{E}$.*Beweis.* Existenz: Folgt direkt aus 1.3.7.

Eindeutigkeit: Folgt direkt aus Satz 1.2.13. □

1.4 Das Beispiel - Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$ aus Verteilungsfunktionen

Definition 1.4.1.  $F: \mathbb{R} \rightarrow \mathbb{R}$ heißt **Verteilungsfunktion**, falls

- (i) $0 \leq F(t) \leq 1$ für alle $t \in \mathbb{R}$,
- (ii) F ist nicht fallend,
- (iii) F ist rechtsstetig, d. h. $\lim_{s \downarrow t} F(s) = F(t)$,
- (iv) $\lim_{t \rightarrow \infty} F(t) = 1$ und $\lim_{t \rightarrow -\infty} F(t) = 0$.

Satz 1.4.2.  Für jede Verteilungsfunktion F gibt es **genau** ein Wahrscheinlichkeitsmaß \mathbb{P}_F auf $\mathcal{B}(\mathbb{R})$ mit $\mathbb{P}_F((-\infty, t]) = F(t)$, $t \in \mathbb{R}$. Man sagt dann, „ \mathbb{P}_F ist gemäß F verteilt“ oder „ \mathbb{P}_F hat Verteilung F “ und schreibt $\mathbb{P}_F \sim F$.

Vorlesung 6

Beweis. Eindeutigkeit: Das haben wir uns schon in Beispiel 1.2.15 überlegt.*Existenz:* Um den Fortsetzungssatz von Carathéodory zu nutzen, müssen wir zunächst einen Semiring wählen, der die Borel- σ -Algebra erzeugt. Wir wissen bereits, dass alle möglichen Arten von Intervallen $\mathcal{B}(\mathbb{R})$ erzeugt, die meisten sind aber keine Semiringe. Wir nehmen

$$\mathcal{S} = \{(a, b] : a, b \in \mathbb{R}\},$$

und stellen sofort fest (Eigenschaften checken), dass \mathcal{S} ein Semiring mit $\sigma(\mathcal{S}) = \mathcal{B}(\mathbb{R})$ ist. Als Mengenfunktion auf \mathcal{S} definieren wir

$$\mu((a, b]) := F(b) - F(a).$$

Weil F nicht-fallend ist und $0 \leq F \leq 1$ gilt, bildet μ nach $[0, 1]$ ab. Checken wir als nächstes die Voraussetzungen vom Fortsetzungssatz:

- (i) $\mu(\emptyset) = \mu((a, a]) = F(a) - F(a) = 0$
- (ii) Für die σ -Additivität seien $(a_n, b_n] \in \mathcal{S}$ paarweise disjunkt mit

$$\bigcup_{k=1}^{\infty} (a_k, b_k] \in \mathcal{S}, \quad \text{also} \quad \bigcup_{k=1}^{\infty} (a_k, b_k] =: (a, b),$$

für geeignete $a, b \in \mathbb{R}$. Als Anschauungsbeispiel haltet ihr am besten das konkrete Beispiel $(0, 1] = \bigcup_{k=1}^{\infty} (\frac{1}{k+1}, \frac{1}{k}]$ im Kopf. Um den Beweis besser zu verstehen, schauen wir uns erstmal den endlichen Fall an, d. h.

$$(a, b] = \bigcup_{k=1}^N (a_k, b_k],$$

für ein $N \in \mathbb{N}$ mit geordneten Intervallen $a = a_1 < b_1 = a_2 < \dots < b_N = b$. Dann bekommen wir die σ -Additivität sofort:

$$\mu((a, b]) \stackrel{\text{Def.}}{=} F(b) - F(a) \stackrel{\text{Teleskop}}{=} \sum_{k=1}^N (F(b_k) - F(a_k)) = \sum_{k=1}^N \mu((a_k, b_k]),$$

wobei wir $F(a_k) = F(b_{k-1})$ genutzt haben. Nun aber zurück zum allgemeinen Fall: Wir zeigen

$$F(b) - F(a) = \sum_{k=1}^{\infty} (F(b_k) - F(a_k)), \quad (1.6)$$

denn das ist gerade die σ -Additivität

$$\mu\left(\bigcup_{k=1}^{\infty} (a_k, b_k]\right) = \sum_{k=1}^{\infty} \mu((a_k, b_k])$$

für $(a, b] = \bigcup_{k=1}^{\infty} (a_k, b_k]$. Für die Gleichheit (1.6) zeigen wir beide Ungleichungen:

„ \geq “: Weil F nicht-fallend ist, folgt

$$F(b) - F(a) \geq \sum_{k=1}^N (F(b_k) - F(a_k))$$

für alle $N \in \mathbb{N}$. Wegen der Monotonie von Folgentrennwerten gilt

$$F(b) - F(a) \geq \lim_{N \rightarrow \infty} \sum_{k=1}^N (F(b_k) - F(a_k)) = \sum_{k=1}^{\infty} (F(b_k) - F(a_k)).$$

„ \leq “: Sei $\varepsilon > 0$ und seien $b_n < \tilde{b}_n$, so dass

$$0 \leq F(\tilde{b}_n) - F(b_n) < \frac{\varepsilon}{2^n} \quad (1.7)$$

für alle $n \in \mathbb{N}$. Die \tilde{b}_n existieren weil F rechtsstetig ist (schreibt mal die Definition der Stetigkeit mit $\frac{\varepsilon}{2^n}$ statt ε hin). Weil

$$(a, b] = \bigcup_{k=1}^{\infty} (a_k, b_k] \stackrel{b_k \leq \tilde{b}_k}{\subseteq} \bigcup_{k=1}^{\infty} (a_k, \tilde{b}_k)$$

gilt, gilt auch

$$[a + \varepsilon, b] \subseteq \bigcup_{k=1}^{\infty} (a_k, \tilde{b}_k).$$

Nach Heine-Borel ist $[a + \varepsilon, b]$ kompakt. Aufgrund der Definition der Kompaktheit reichen endlich viele (a_k, \tilde{b}_k) , um $[a + \varepsilon, b]$ zu überdecken. Also gibt es ein $N \in \mathbb{N}$ mit

$$[a + \varepsilon, b] \subseteq \bigcup_{k=1}^N (a_k, \tilde{b}_k).$$

Daraus folgt dann

$$\begin{aligned}
 F(b) - F(a + \varepsilon) &\stackrel{F \text{ monoton}}{\leq} \sum_{k=1}^N (F(\tilde{b}_k) - F(a_k)) \\
 &\stackrel{(1.7)}{\leq} \sum_{k=1}^{\infty} (F(b_k) + \frac{\varepsilon}{2^k} - F(a_k)) = \sum_{k=1}^{\infty} (F(b_k) - F(a_k)) + \varepsilon,
 \end{aligned}$$

wobei wir im letzten Schritt die geometrische Reihe $\sum_{k=1}^{\infty} \frac{1}{2^k} = 1$ genutzt haben. Wegen der Rechtsstetigkeit von F folgt damit

$$\begin{aligned}
 F(b) - F(a) &= \lim_{\varepsilon \downarrow 0} (F(b) - F(a + \varepsilon)) \\
 &\leq \lim_{\varepsilon \downarrow 0} \left(\sum_{k=1}^{\infty} (F(b_k) - F(a_k)) + \varepsilon \right) = \sum_{k=1}^{\infty} (F(b_k) - F(a_k)).
 \end{aligned}$$

Der Fortsetzungssatz impliziert nun die Existenz eines Maßes \mathbb{P}_F auf $\mathcal{B}(\mathbb{R})$ mit

$$\mathbb{P}_F((a, b]) = \mu((a, b]) = F(b) - F(a), \quad a < b.$$

Das Maß ist nicht automatisch ein Wahrscheinlichkeitsmaß, das folgt aber direkt aus der Stetigkeit von Maßen und der Charakterisierung von \mathbb{P}_F auf den Intervallen:

$$\begin{aligned}
 \mathbb{P}_F(\mathbb{R}) &= \mathbb{P}_F\left(\bigcup_{k=1}^{\infty} (-k, k]\right) \\
 &\stackrel{\text{Stet. Maße}}{=} \lim_{n \rightarrow \infty} \mathbb{P}_F((-n, n]) \\
 &\stackrel{\text{Def.}}{=} \lim_{n \rightarrow \infty} (F(n) - F(-n)) \\
 &= \lim_{n \rightarrow \infty} F(n) - \lim_{n \rightarrow \infty} F(-n) = 1.
 \end{aligned}$$

Achtung, das Argument werden wir jetzt immer wieder nutzen!

Ganz ähnlich zeigen wir den Zusammenhang von F und \mathbb{P}_F auf unendlichen Intervallen, wie im Satz behauptet wird:

$$\mathbb{P}_F((-\infty, t]) = \mathbb{P}_F\left(\bigcup_{k=\lceil |t| \rceil}^{\infty} (-k, t]\right) = \lim_{k \rightarrow \infty} \mathbb{P}_F((-k, t]) = F(t) - \lim_{k \rightarrow \infty} F(-k) = F(t),$$

wobei $\lceil |t| \rceil$ die obere Gaußklammer von $|t|$ ist, also $|t|$ aufgerundet.

□

Bemerkung 1.4.3. ▶ Es gibt ganz analog eine Definition für Verteilungsfunktionen auf dem \mathbb{R}^d , sogenannte „multivariate Verteilungsfunktionen“ – das machen wir später.

Bevor wir zu Beispielen von Wahrscheinlichkeitsmaßen auf $\mathcal{B}(\mathbb{R})$ kommen, hier noch das wichtigste Beispiel eines Maßes auf der Borel- σ -Algebra - das Lebesgue Maß auf $\mathcal{B}(\mathbb{R}^d)$.

Satz 1.4.4. ▶ Es gibt ein eindeutiges Maß λ auf $\mathcal{B}(\mathbb{R}^d)$ mit $\lambda(Q) = \text{Volumen}(Q)$ für alle Quader $Q \subseteq \mathbb{R}^d$. λ heißt Lebesgue-Maß auf \mathbb{R}^d , λ ist ein unendliches Maß.

Beweis. Übung, ziemlich analog zum vorherigen Beweis. Betrachte

$$\mathcal{S} := \{(a_1, b_1] \times \dots \times (a_n, b_n] : a_1, \dots, a_d, b_1, \dots, b_d \in \mathbb{R}\},$$

\mathcal{S} ist ein Semiring. $\mu(Q) := \text{Volumen}(Q) = \prod_{k=1}^d (b_k - a_k)$ ist eine σ -additive Mengenfunktion auf \mathcal{S} (die σ -Additivität ist der einzige komplizierte Schritt). $E_n := (-n, n] \times \dots \times (-n, n]$ ist die

benötigte Folge in \mathcal{S} mit endlichem Volumen, die gegen \mathbb{R}^d wächst. λ sei nun das eindeutige Maß auf $\mathcal{B}(\mathbb{R})$ das μ fortsetzt. Es fehlt noch, dass λ ein unendliches Maß ist. Aber auch das geht mit den Argumenten des vorherigen Beweises:

$$\lambda(\mathbb{R}^d) \stackrel{\text{Stet. Maße}}{=} \lim_{n \rightarrow \infty} \lambda((n, -n] \times \dots \times (n, -n]) = \lim_{n \rightarrow \infty} (2n)^d = \infty.$$

□

Meistens betrachten wir das Lebesguemaß auf \mathbb{R} , das im Prinzip die „Länge“ einer Menge misst, zumindest gilt das für Intervalle (oder disjunkte Vereinigungen von Intervallen).

Bemerkung 1.4.5. ▶ Auf den Übungsblättern diskutieren wir das Lebesgue Maß auf Teilmengen vom \mathbb{R}^d , insbesondere auf Intervalle oder Quadern. Beispielsweise sind dann die messbaren Mengen

$$\mathcal{B}([0, 1]) := \{B \subseteq [0, 1] : B \in \mathcal{B}(\mathbb{R})\} = \sigma(\{[a, b] : 0 \leq a < b \leq 1\})$$

die Borel-messbaren Teilmengen von $[0, 1]$ und $\lambda_{[0,1]}$ das eindeutige Maß auf $\mathcal{B}([0, 1])$ mit $\lambda_{[0,1]}(B) = \lambda(B)$ für Borel-messbare Teilmengen von $[0, 1]$. Das Lebesgue Maß auf $[0, 1]$ ist das eindeutige Maß auf den Borel-messbaren Teilmengen von $[0, 1]$, so dass das Maß von Intervallen die Länge ist.

Jetzt kommen wir zu konkreten Beispielen von Verteilungsfunktionen, die uns erneut in der Stochastik begegnen werden. Im Folgenden werden wir regelmässig **Indikatorfunktionen** benutzen:

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

die auch in Analysis 2 (vielleicht in anderer Schreibweise) im Rahmen der Integrationstheorie vermutlich schon über den Weg gelaufen sind.

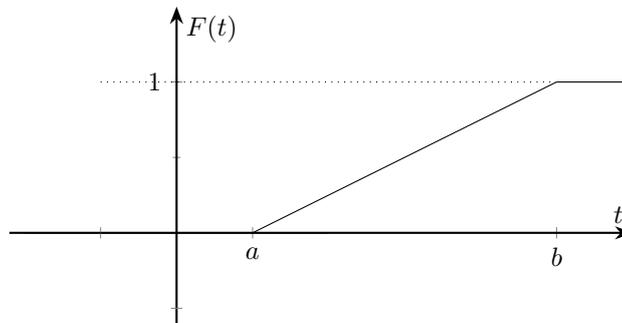
Beispiel 1.4.6. ▶ Für $a < b$ sei

$$F(t) = \frac{t-a}{b-a} \mathbf{1}_{[a,b]}(t) + \mathbf{1}_{(b,\infty)}(t), \quad t \in \mathbb{R},$$

oder anders geschrieben als

$$F(t) = \begin{cases} 0 & : t < a \\ \frac{t-a}{b-a} & : t \in [a, b] \\ 1 & : t > b \end{cases}.$$

Natürlich erfüllt F die Eigenschaften einer Verteilungsfunktion, das zugehörige Maß \mathbb{P}_F nennt man **Gleichverteilung** auf $[a, b]$ und man schreibt $\mathbb{P}_F \sim \mathcal{U}([a, b])$.



Man nennt das Maß auch $\mathcal{U}([a, b])$, \mathcal{U} steht dabei für uniform.

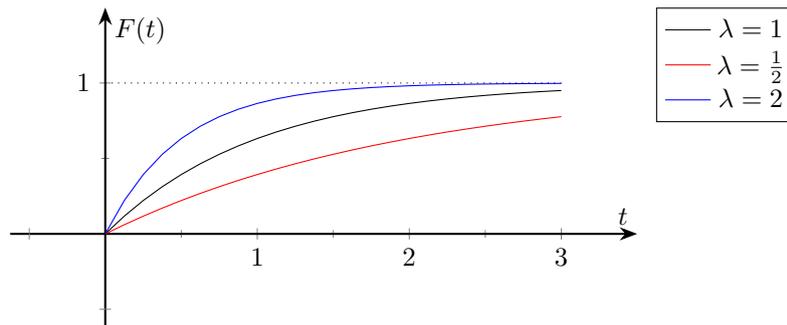
Beispiel 1.4.7. ▶ Für $\lambda > 0$ sei

$$F(t) = (1 - e^{-\lambda t})\mathbf{1}_{[0, \infty)}(t), \quad t \in \mathbb{R},$$

oder anders geschrieben als

$$F(t) = \begin{cases} 0 & : t \leq 0 \\ 1 - e^{-\lambda t} & : t > 0 \end{cases}.$$

Aufgrund der Eigenschaften der Exponentialfunktion erfüllt F die Eigenschaften der Exponentialfunktion, das zugehörige Maß \mathbb{P}_F nennt man **Exponentialverteilung mit Parameter** $\lambda > 0$ und schreibt $\mathcal{P}_F \sim \text{Exp}(\lambda)$.



Man nennt das Maß auch $\text{Exp}(\lambda)$. In der Graphik ist $\text{Exp}(\lambda)$ für drei verschiedene λ geplottet.

Definition 1.4.8. ▶ Ist $f: \mathbb{R} \rightarrow [0, \infty)$ integrierbar mit $\int_{\mathbb{R}} f(x)dx = 1$, dann heißt f **Dichtefunktion** der Verteilungsfunktion

$$F(t) = \int_{-\infty}^t f(x)dx, \quad t \in \mathbb{R}. \quad (1.8)$$

Ist umgekehrt F von der Form (1.8), so heißt f **Dichte** von F . Verteilungsfunktionen mit Dichten nennt man auch **absolutstetig**, Maße mit absolutstetiger Verteilungsfunktion nennt man **absolutstetige Maße**. In der großen Übung wurde diskutiert, warum solch ein F die vier Eigenschaften einer Verteilungsfunktion erfüllt.

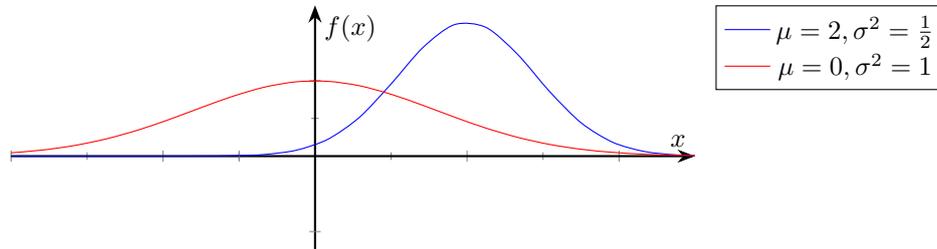
Zwei Beispiele haben wir schon gesehen: $\mathcal{U}([a, b])$ und $\text{Exp}(\lambda)$ haben beide absolutstetige Verteilungsfunktionen. Die zugehörigen Dichten berechnet ihr in den Übungsaufgaben. Aber wie findet man die Dichten von absolutstetigen Verteilungsfunktionen? Ableiten! Das ist, zumindest für stetige Dichten, der Hauptsatz der Integral- und Differentialrechnung. Probiert das bei den zwei Beispielen mal aus. Ableiten ohne nachzudenken erlaubt es die Dichte f zu erraten, wenn man dann durch Integrieren $F(t) = \int_{-\infty}^t f(x)dx$ nachrechnen kann, so ist f eine Dichte von F . Warum es praktisch ist eine absolutstetige Verteilungsfunktion zu haben, wird zum Beispiel in Diskussion 1.4.13 klarer. Man kann direkt wichtige Eigenschaften des Maßes \mathbb{P}_F aus der Dichte f ablesen.

Beispiel 1.4.9. ▶ Die schönste Anwendung von Polarkoordinaten und Fubini (siehe Analysis 2) ist die Berechnung des Integrals $\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$. Damit ist $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ eine Dichtefunktion. Man nennt die zugehörige Verteilungsfunktion

$$F(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} dx, \quad t \in \mathbb{R}.$$

Verteilungsfunktion der **(standard) Normalverteilung**. Das Maß \mathbb{P}_F nennt man dann auch (standard) normalverteilt und man schreibt $\mathbb{P}_F \sim \mathcal{N}(0, 1)$. In der großen Übung wird diskutiert,

dass für $\mu \in \mathbb{R}$ und $\sigma^2 \geq 0$ auch $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ eine Dichtefunktion ist. Die zugehörige Verteilung nennt man auch normalverteilt und schreibt $\mathcal{N}(\mu, \sigma^2)$.



Die Bedeutung von μ und σ^2 diskutierten wir später. Warnung: Warum schreiben wir nur eine Formel für die Dichte f , jedoch nicht für die Verteilungsfunktion F hin? Es gibt einfach keine Formel für das Integral $\int_{-\infty}^t e^{-x^2/2} dx$! Aufgrund der Form der Kurve spricht man auch von der Glockenkurve und weil diese von Gauß entdeckt wurde, von der Gausschen Glockenkurve.

Das Gegenstück zu absolutstetigen Verteilungen sind sogenannte diskrete Verteilungen:

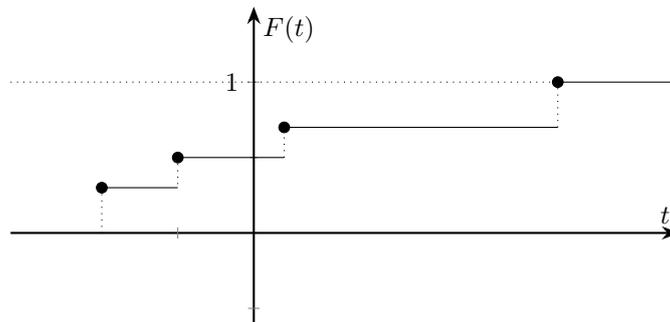
Beispiel 1.4.10.  Für $a_1, \dots, a_N \in \mathbb{R}$, $N \in \mathbb{N}$ oder $N = +\infty$, mit $p_1, \dots, p_N \geq 0$ und $\sum_{k=1}^N p_k = 1$ ist

$$F(t) := \sum_{k=1}^N p_k \mathbf{1}_{[a_k, \infty)}(t) = \sum_{a_k \leq t} p_k, \quad t \in \mathbb{R},$$

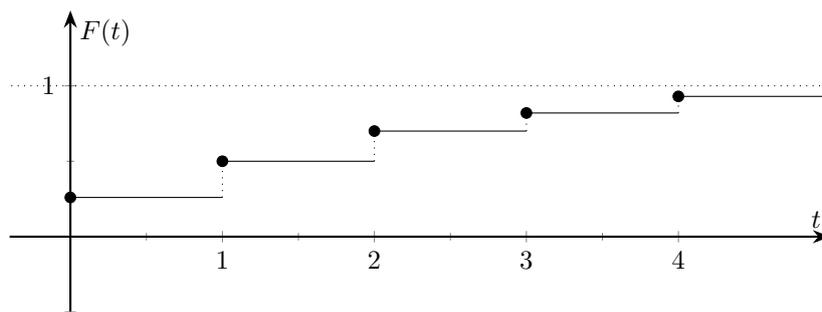
eine Verteilungsfunktion. Die zugehörigen Maße \mathbb{P}_F werden (**endliche**) **diskrete Verteilungen** genannt. In den Übungen zeigt ihr, dass die Maße im diskreten Fall ganz einfach angegeben werden können, es sind Mischungen aus Dirac-Maßen an den Stellen a_1, \dots, a_N :

$$\mathbb{P}_F = \sum_{k=1}^N p_k \delta_{a_k}.$$

Wie zeigt man das? Einfach die Menge $(-\infty, t]$ in das Maß einsetzen, das gibt das gewünschte F .



 Ganz konkret heißt \mathbb{P}_F für $a_k = k$ und $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}$, **Poissonverteilung mit Parameter** $\lambda > 0$ auf $\mathcal{B}(\mathbb{R})$. Beachte: Weil wir die Poissonverteilung bereits auf $\mathcal{P}(\mathbb{N})$ definiert haben gibt es eine gewisse Doppeldeutigkeit, mit der Diskussion der nächsten Vorlesung wird aber klar, dass beide Maße das gleiche beschreiben, nämlich die Verteilung einer Einheit Masse auf \mathbb{N} mit den Wahrscheinlichkeiten p_k für die natürliche Zahl k . Die Poissonverteilung mit Parameter λ wird auch als **Poi**(λ) genannt.



Vorlesung 7

Manche werden sich fragen, wo denn jetzt die Stochastik geblieben ist. Wir haben schließlich gerade Begriffe der Stochastik benutzt, z. B. den Begriff der Uniformverteilung, auch die Gaußsche Glockenfunktion ist bereits aufgetaucht, über zufällige Experimente haben wir aber schon länger nicht gesprochen. Als konkrete Motivation zur Nutzung der abstrakten Theorie zur Modellierung zufälliger Experimente, schauen wir uns das uniforme Ziehen aus $[0, 1]$ an.

Diskussion 1.4.11. ▶ [Stochastische Modellierung, Nr. 2]

Das Modellieren von endlich vielen Möglichkeiten ist relativ einfach, siehe Diskussion 1.1.8. Man kommt recht natürlich auf die Eigenschaften der σ -Algebra und des Maßes. Zur Erinnerung war das gleichverteilte Ziehen aus einer endlichen Menge modelliert durch den endlichen Zustandsraum Ω (=Möglichkeiten zum Ziehen), $\mathcal{A} = \mathcal{P}(\Omega)$ und der diskreten Gleichverteilung $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$. Das Modellieren von Experimenten mit unendlich vielen Möglichkeiten ist dagegen schwieriger. Wie modelliert man zum Beispiel das Ziehen aus dem Intervall $[0, 1]$, sodass kein Bereich von $[0, 1]$ bevorteilt wird? Wenn wir beobachten wollen, ob eine feste Zahl gezogen wurde oder nicht, müssen die einelementigen Mengen $\{t\}$ in der σ -Algebra sein. Wenn kein Element bevorzugt werden soll, also $\mathbb{P}(\{t\})$ für alle t gleich sein soll, führt die Unendlichkeit automatisch zu $\mathbb{P}(\{t\}) = 0$ für alle $t \in [0, 1]$. Warum das? Wenn man irgendeine Folge (a_n) unterschiedlicher Zahlen in $[0, 1]$ wählt, z. B. $a_n = \frac{1}{n}$, und $\mathbb{P}(\{t\}) =: c$ für alle t setzt, so gilt wegen der σ -Additivität von Maßen

$$1 \geq \mathbb{P}\left(\bigcup_{k=1}^{\infty} \{a_k\}\right) = \sum_{k=1}^{\infty} \mathbb{P}(\{a_k\}) = \sum_{k=1}^{\infty} c,$$

also $c = 0$. Hier sehen wir deutlich den Unterschied zur Gleichverteilung auf endlichen Mengen, die einfache Definition durch Einpunktfolgen führt zu nichts! Im Gegensatz zum endlichen Fall legen wir für gleichverteilten Zufall in $[0, 1]$ jetzt fest, dass die Wahrscheinlichkeit von Teilintervallen von $[0, 1]$ nur von der Länge abhängen soll. Das führt zur Forderung $\mathbb{P}((a, b)) = b - a = F(b) - F(a)$ für $a < b$ aus $[0, 1]$, wobei F die Verteilungsfunktion aus Beispiel 1.4.6 ist. Da wir als mathematisches Modell des zufälligen Ziehens eine σ -Algebra und ein Maß haben wollen, wählen wir nun die kleinste σ -Algebra die all diese Intervalle enthält (die Borel- σ -Algebra) und darauf ein Maß, das den Intervallen die geforderten Wahrscheinlichkeiten gibt. Aufgrund des Fortsetzungssatzes gibt es so ein Maß, das ist gerade $\mathcal{U}([0, 1])$.

Hoffentlich ist jetzt einsichtig, warum die Modellierung von komplizierten reellen zufälligen Experimenten mit der Borel- σ -Algebra Sinn macht. Eine Frage bleibt aber noch: Warum nehmen wir nicht einfach die ganze Potenzmenge auf \mathbb{R} als Modell, so wie beim zufälligen Ziehen in endlichen Mengen?

Bemerkung 1.4.12. ▶

- (i) $\mathcal{B}(\mathbb{R})$ funktioniert wunderbar! Insbesondere weil wir sehr handliche Erzeuger haben (z. B. verschiedene Arten von Intervallen) und deshalb aufgrund der bewiesenen Theoreme (fast) nur mit Intervallen arbeiten müssen.
- (ii) $\mathcal{P}(\mathbb{R})$ ist zu groß, z. B. das Lebesgue-Maß oder die Normalverteilung kann zwar auf $\mathcal{B}(\mathbb{R})$, aber nicht auf $\mathcal{P}(\mathbb{R})$ definiert werden (\rightsquigarrow Vitali-Menge). Es gilt tatsächlich $\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{P}(\mathbb{R})$, ganz einfache Beispiele für nicht Borel-messbare Mengen gibt es aber nicht.

Die nächste Runde der Modellierung zufälliger Experimente findet erst in ein paar Wochen statt. Bis dahin könnt ihr die Ideen sacken lassen und euch wieder an der abstrakten Theorie erfreuen.

Das Umschalten im Kopf von Verteilungsfunktionen auf Maße ist anfangs extrem schwierig. Wir wissen zwar abstrakt, dass es für jede Verteilungsfunktion genau ein Maß auf $\mathcal{B}(\mathbb{R})$ gibt und andersrum für jedes Maß eine eindeutige Verteilungsfunktion, aber was bedeutet das konkret? Das versteht man am besten, wenn man Eigenschaften von F in Eigenschaften von \mathbb{P}_F übersetzt:

Diskussion 1.4.13. ▶ Wir starten mit einer nicht sehr rigorosen aber dennoch hilfreichen Interpretation:

„ F beschreibt, wie durch \mathbb{P}_F eine Einheit Zufall auf \mathbb{R} verteilt wird.“

Dazu sei $F(b) - F(a)$ der Anteil des gesamten Zufalls ($F(b) - F(a)$ ist immer zwischen 0 und 1), der in $(a, b]$ gelandet ist. Man spricht auch statt „Anteil“ von der „Masse“ Zufall in $(a, b]$.

Wir schauen uns jetzt an, was drei Eigenschaften von F (stetig, konstant, stark wachsend) für die Verteilung der Masse bedeuten.

Stetigkeit vs. Sprünge: Zunächst berechnen wir die Masse einer Einpunktmenge $\{t\}$ aus den bekannten Eigenschaften von Maßen und Verteilungsfunktionen. Wie immer versuchen wir die gesuchte Menge durch Mengen der Form $(a, b]$ auszudrücken, weil wir für diese Mengen eine Verbindung zwischen F und \mathbb{P}_F haben:

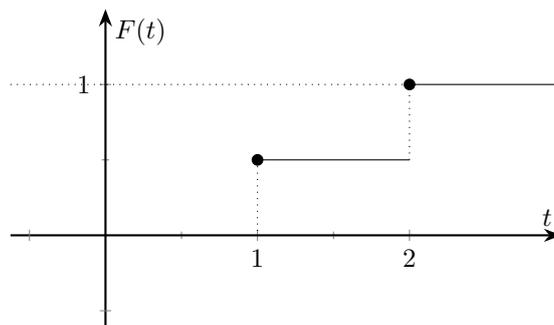
$$\begin{aligned} \mathbb{P}_F(\{t\}) &= \mathbb{P}_F\left(\bigcap_{n=1}^{\infty} \left(t - \frac{1}{n}, t\right]\right) \\ &\stackrel{\text{Stet. Maße}}{=} \lim_{n \rightarrow \infty} \mathbb{P}_F\left(\left(t - \frac{1}{n}, t\right]\right) \\ &\stackrel{\text{Def. } \mathbb{P}_F}{=} \lim_{n \rightarrow \infty} \left(F(t) - F\left(t - \frac{1}{n}\right)\right) \\ &= F(t) - \lim_{n \rightarrow \infty} F\left(t - \frac{1}{n}\right) = F(t) - F(t-), \end{aligned}$$

wobei $F(t-) := \lim_{s \uparrow t} F(s)$ der Linksgrenzwert aus der Analysis ist. Konsequenz: Ist F stetig in t , so hat die Einpunktmenge $\{t\}$ keine Masse. Hierzu beachte man, dass F an jeder Stelle rechtsstetig ist, die Stetigkeit somit äquivalent zu $F(t) = F(t-)$ ist. Insbesondere haben alle einpunktigen Mengen keine Masse, sofern F eine stetige Funktion ist (z. B. bei $\mathcal{U}([a, b])$, $\text{Exp}(\lambda)$, $\mathcal{N}(\mu, \sigma^2)$). Klingt komisch, oder? Ist es aber nicht. Hier sehen wir, warum Maße erst auf überabzählbaren Mengen wirklich spannend werden:

$$\mathbb{P}_F((a, b]) = \mathbb{P}_F\left(\bigcup_{t \in (a, b]} \{t\}\right) \neq \sum_{t \in (a, b]} \mathbb{P}_F(\{t\}),$$

weil σ -Additivität nur für Vereinigungen abzählbar vieler Mengen gilt. Was sollte die überabzählbare Summe auf der rechten Seite auch bedeuten?

F konstant: Überlegen wir nun, was es für \mathbb{P}_F bedeutet, wenn F auf einem Intervall konstant ist. Schauen wir dazu zunächst ein Beispiel an. Betrachten wir folgende einfache Verteilungsfunktion



aus der Klasse der diskreten Verteilungen. Nach der Diskussion zur Stetigkeit wissen wir, dass das zugehörige Maß \mathbb{P}_F folgendes erfüllt: $\mathbb{P}_F(\{1\}) = \mathbb{P}_F(\{2\}) = \frac{1}{2}$. Wegen der σ -Additivität folgt natürlich (es gibt insgesamt nur eine Einheit Zufall zu verteilen), dass $\mathbb{P}_F(A) = 0$ für alle Borelmengen A mit $1, 2 \notin A$. Das Maß \mathbb{P}_F hat also keine Masse außerhalb der Menge $\{1, 2\}$. Schauen wir uns F an, so sehen wir also, dass \mathbb{P}_F keine Masse in den konstanten Bereichen hat. Für Intervalle $(a, b]$ folgt das allgemein natürlich aus $\mathbb{P}_F((a, b]) = F(b) - F(a)$ was gerade 0 ist, wenn F zwischen a und b konstant ist:

„ \mathbb{P}_F hat keine Masse dort, wo F konstant ist.“

Wenn wir die Beobachtung auf $\text{Poi}(\lambda)$ aus Beispiel 1.4.10 anwenden, so sehen wir, dass das zugehörige Maß \mathbb{P}_F nur Masse auf \mathbb{N} hat. Damit kann man ein $\text{Poi}(\lambda)$ -verteiltes Maß auf $\mathcal{B}(\mathbb{R})$ mit der Definition aus Beispiel 1.1.12 identifizieren, wir verteilen eine Einheit Zufall jeweils auf \mathbb{N} (einmal wird die Einheit Zufall direkt auf \mathbb{N} verteilt, einmal auf \mathbb{N} als Teilmenge von \mathbb{R}).

F stark wachsend: Wir wissen nun wieviel Masse an Sprungstellen liegt und auch, dass keine Masse in konstanten Bereichen liegt. Fragt sich also, wo die Masse sonst noch zu finden ist:

„ \mathbb{P}_F hat dort viel Masse dort, wo F am stärksten wächst.“

Formell folgt das natürlich aus $\mathbb{P}_F((a, b]) = F(b) - F(a)$ weil dann auf ein kleines Intervall $(a, b]$ viel Masse verteilt wird, wenn $F(b)$ deutlich größer als $F(a)$ ist. Ist a nah an b , so bedeutet das natürlich, dass F dort stark wächst. Schauen wir uns wieder ein passendes Beispiel an, die Exponentialverteilung $\text{Exp}(\lambda)$ für verschiedene $\lambda > 0$. Am Bildchen in Beispiel 1.4.7 ist zu erkennen, dass viel Masse nah bei der 0 liegt wenn λ groß ist, die Verteilungsfunktion bei 0 also steil ist. Natürlich sehen wir das auch formell aus der Verteilungsfunktion weil für alle $\epsilon > 0$

$$\mathbb{P}_F((0, \epsilon]) = F(\epsilon) - F(0) = (1 - e^{-\lambda\epsilon}) - (1 - e^{-\lambda 0}) = 1 - e^{-\lambda\epsilon},$$

was monoton wachsend in λ ist.

Der Fall mit Dichten: Die obige Diskussion können wir für Verteilungsfunktionen mit Dichten noch konkretisieren. Sei dazu F eine Verteilungsfunktion mit Dichte f , also $F(t) = \int_{-\infty}^t f(x)dx$. Weil F stetig ist, haben alle einpunktigen Mengen keine Masse. Aber wie können wir an f direkt sehen, wo die Masse verteilt ist? Ist f stetig, so folgt aus dem Hauptsatz der Analysis $F'(t) = f(t)$ für alle $t \in \mathbb{R}$. Folglich impliziert ein an der Stelle t großes f ein in t stark wachsendes F und damit viel Masse um t . Andersrum impliziert ein an der Stelle t kleines f ein in t wenig wachsendes F und damit wenig Masse um t . Im Extremfall impliziert natürlich $f = 0$ in $(a, b]$ auch F konstant in $(a, b]$ und damit wird keine Masse auf $(a, b]$ verteilt. Wir merken uns grob

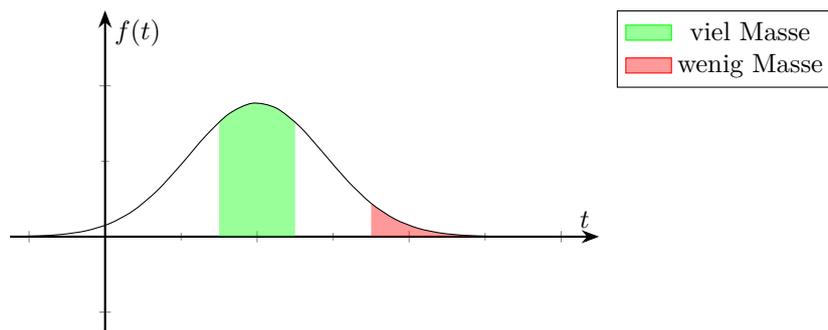
„Hat F eine Dichte, so ist viel Masse dort, wo f groß ist.“

Die nützlichste Interpretation ist durch den Flächeninhalt zwischen Graphen von f und der x -Achse gegeben. Wegen

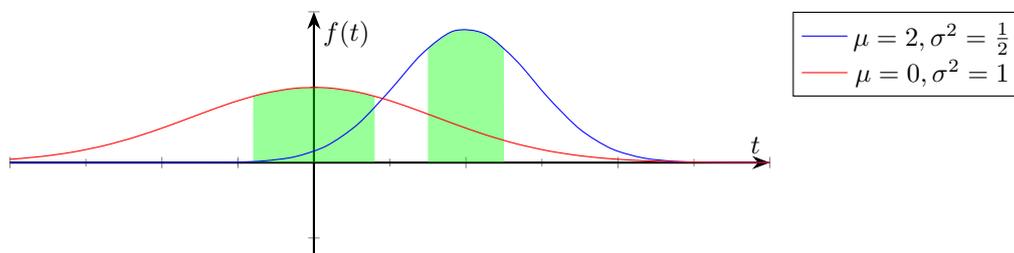
$$\mathbb{P}_F((a, b]) = F(b) - F(a) = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = \int_a^b f(x)dx,$$

ist die Masse in $(a, b]$ gerade die Fläche unter f zwischen a und b . Dazu ist zu beachten, dass nach Annahme die Gesamtfläche zwischen Graphen und x -Achse 1 ist.

In folgendem Beispiel ist die Dichte von $\mathcal{N}(2, 1)$ geplottet:



Wir sehen also, dass viel Masse des Maßes $\mathcal{N}(2, 1)$ um die 2 herum verteilt ist und sehr wenig Masse weit weg von der 2 verteilt ist. Der grüne Bereich ist gerade so gewählt, dass dieser Flächeninhalt $\frac{1}{3}$ ist. Ein Drittel der Masse von $\mathcal{N}(2, 1)$ liegt also im Schnittbereich des grünen Bereichs mit der x -Achse, sehr nah an der 2. Man sagt, die Verteilung ist um 2 konzentriert. Wenn wir zwei verschiedene Normalverteilungen vergleichen, sieht es wie im folgenden Beispiel aus:



Der Inhalt der grünen Flächen ist wieder $\frac{1}{3}$, die zugehörigen normalverteilten Maße auf $\mathcal{B}(\mathbb{R})$ haben deshalb Masse $\frac{1}{3}$ im jeweiligen Schnittbereich mit der x -Achse. Wir sehen schon an dem Bild, dass niedrigeres σ dafür sorgt, dass die Verteilung mehr Masse nah an μ hat. Darauf gehen wir in ein paar Wochen noch viel ausführlicher ein.

Kapitel 2

Abbildungen zwischen messbaren Räumen

Vorlesung 8

Bevor wir messbare Abbildungen definieren, erinnern wir kurz an bereits bekannte Konzepte in der Mathematik. Wir betrachten immer Objekte und Abbildungen zwischen Objekten, die auf eine gewisse Art „natürlich“ (strukturerhaltend) sind:

Mengen	Abbildungen
Gruppen	Homomorphismen
Vektorräume	Lineare Abbildungen
Metrische Räume	stetige Abbildungen

Passend dazu diskutieren wir jetzt die strukturerhaltenden Abbildungen zwischen messbaren Räumen, sogenannte messbare Abbildungen.

2.1 Messbare Abbildungen

Definition 2.1.1. ▶ Seien (Ω, \mathcal{A}) , (Ω', \mathcal{A}') messbare Räume und $f: \Omega \rightarrow \Omega'$. f heißt **messbar**, falls Urbilder messbarer Mengen messbar sind; in Formeln

$$A' \in \mathcal{A}' \Rightarrow f^{-1}(A') \in \mathcal{A}.$$

Es gibt verschiedene Notationen für messbare Abbildungen, man nutzt synonym

- $f: \Omega \rightarrow \Omega'$ ist $(\mathcal{A}, \mathcal{A}')$ -messbar,
- $f: (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ ist messbar,
- $f: \Omega \rightarrow \Omega'$ ist messbar bezüglich \mathcal{A} und \mathcal{A}' .

Genau wie Stetigkeit zwischen metrischen Räumen von den gewählten Metriken abhängt, hängt auch die Messbarkeit von den gewählten σ -Algebren ab. Wenn klar ist, welche σ -Algebren gewählt sind, redet man trotzdem einfach nur von messbaren Abbildungen.

Bemerkung 2.1.2. ▶ Die Definition der Messbarkeit ist analog zur Stetigkeit zwischen metrischen Räumen, dabei werden messbare Mengen durch offene Mengen ersetzt.

Definition 2.1.3. ▶ Ist $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, dann nennt man eine messbare Abbildung auch **Zufallsvariable** und schreibt X statt f .

Wie bei der Konstruktion von Maßen haben wir das Problem, dass wir alle messbaren Mengen testen müssen. Das ist gerade bei der Borel- σ -Algebra unmöglich, wir kennen die Mengen nicht alle. Zum Glück ist es wie im Kapitel zuvor, es reicht einen Erzeuger zu betrachten:

Proposition 2.1.4. ▶ Ist \mathcal{E}' ein Erzeuger von \mathcal{A}' und $f: \Omega \rightarrow \Omega'$. Dann ist f messbar bzgl. \mathcal{A} und \mathcal{A}' genau dann, wenn

$$A' \in \mathcal{E}' \Rightarrow f^{-1}(A') \in \mathcal{A}.$$

Beweis.

„ \Rightarrow “: \checkmark weil $\mathcal{E}' \subseteq \mathcal{A}'$

„ \Leftarrow “: Mal wieder der Trick der guten Mengen. Sei dazu

$$\mathcal{F}' := \{A' \in \mathcal{A}' : f^{-1}(A') \in \mathcal{A}\},$$

wir zeigen $\mathcal{F}' = \mathcal{A}'$. Nach Annahme gilt $\mathcal{E}' \subseteq \mathcal{F}'$. Wenn \mathcal{F}' eine σ -Algebra ist, dann sind wir fertig, weil dann

$$\mathcal{A}' = \sigma(\mathcal{E}') \subseteq \sigma(\mathcal{F}') = \mathcal{F}' \subseteq \mathcal{A}'$$

und folglich $\mathcal{A}' = \mathcal{F}'$ gilt. Doch wenn man die Definition von \mathcal{F}' anschaut, ist das gerade die Messbarkeit.

Wir überprüfen die definierenden Eigenschaften einer σ -Algebra und können dazu auf elementare Eigenschaften des Urbildes von Abbildungen in Analysis 1 zurückgreifen:

(i) $\emptyset \in \mathcal{F}'$, weil $f^{-1}(\emptyset) = \emptyset \in \mathcal{A}$

(ii) Ist $A' \in \mathcal{F}'$, so gilt

$$f^{-1}((A')^C) = (f^{-1}(A'))^C \in \mathcal{A}$$

weil $A \in \mathcal{F}'$ ist und \mathcal{A} als σ -Algebra abgeschlossen bezüglich Komplementbildung ist.

(iii) Sind $A'_1, A'_2, \dots \in \mathcal{F}'$, so gilt

$$f^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(A'_n) \in \mathcal{A}$$

weil die Mengen in \mathcal{F}' sind und \mathcal{A} als σ -Algebra abgeschlossen bezüglich Vereinigungen ist.

□

Definition 2.1.5. ▶ Ist $f: (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ messbar, so heißt f **Borel-messbar**.

Beispiel 2.1.6. ▶

- Jede stetige Abbildung $f: \mathbb{R} \rightarrow \mathbb{R}$ ist auch Borel-messbar. Warum? Wir nutzen Proposition 2.1.4, angewandt auf $\sigma(\{O \subseteq \mathbb{R} : O \text{ offen}\}) = \mathcal{B}(\mathbb{R})$ mit der Erinnerung, dass Urbilder offener Mengen unter stetigen Abbildungen offen (insbesondere Borel-messbar) sind.
- Indikatorfunktionen

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

sind $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ -messbar genau dann, wenn A messbar ist. Das zu prüfen ist relativ simpel, weil wir alle möglichen Urbilder direkt hinschreiben können:

$$\mathbf{1}_A^{-1}(B) = \{\omega \in \Omega : \mathbf{1}_A(\omega) \in B\} = \begin{cases} A & : 1 \in B, 0 \notin B \\ A^C & : 1 \notin B, 0 \in B \\ \mathbb{R} & : 1, 0 \in B \\ \emptyset & : 1, 0 \notin B \end{cases}.$$

Wie für stetige Abbildungen zeigt man, dass die Verknüpfung messbarer Abbildungen wieder messbar ist. Auch das ist eine kleine Übungsaufgabe.

Bemerkung 2.1.7.  Wir erinnern daran, dass

$$\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, t] : t \in \mathbb{R}\}) = \sigma(\{(-\infty, t) : t \in \mathbb{R}\}) = \sigma(\{(a, b) : a < b\}).$$

Wegen Proposition 2.1.4 ist deshalb $f: \Omega \rightarrow \mathbb{R}$ $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ -messbar genau dann, wenn

$$f^{-1}((-\infty, t]) = \{\omega \in \Omega : f(\omega) \leq t\} =: \{f \leq t\} \in \mathcal{A}$$

für alle $t \in \mathbb{R}$. Die Kurzschreibweise $\{f \leq t\}$ ist etwas ungewohnt, wird ab jetzt aber oft genutzt. Analog ist auch f messbar genau dann, wenn

$$f^{-1}((-\infty, t)) = \{\omega \in \Omega : f(\omega) < t\} =: \{f < t\} \in \mathcal{A}$$

für alle $t \in \mathbb{R}$ oder

$$f^{-1}((a, b)) = \{\omega \in \Omega : f(\omega) \in (a, b)\} =: \{f \in (a, b)\} \in \mathcal{A}$$

für alle reellen Zahlen $a < b$. Analog kann man auch halb-offene Intervalle, abgeschlossene Mengen, kompakte Mengen, offene Mengen und so weiter nutzen, jeder Erzeuger von $\mathcal{B}(\mathbb{R})$ gibt eine Möglichkeit um Messbarkeit zu prüfen.

Definition 2.1.8.  Sei $f: \Omega \rightarrow \Omega'$ für einen messbaren Raum (Ω', \mathcal{A}') ist. Dann ist die Menge aller Urbilder

$$\mathcal{A} := \{f^{-1}(A') : A' \in \mathcal{A}'\}$$

eine σ -Algebra und \mathcal{A} natürlich ist die kleinste σ -Algebra auf Ω , für die f $(\mathcal{A}, \mathcal{A}')$ -messbar. Wir nennen die σ -Algebra \mathcal{A} auch $\sigma(f)$.

Beispiel 2.1.9. 

- $\sigma(\mathbf{1}_A) = \{\emptyset, \Omega, A, A^C\}$
- Sei $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = c$, dann ist $\sigma(f) = \{\emptyset, \mathbb{R}\}$.

Definition 2.1.10.  Seien $(\Omega'_i, \mathcal{A}'_i)$ messbare Räume, $f_i: \Omega \rightarrow \Omega'_i$ für $i \in I$. Dann ist

$$\sigma(f_i, i \in I) := \sigma\left(\bigcup_{i \in I} \sigma(f_i)\right) = \sigma(\{f_i^{-1}(A'_i) : A'_i \in \mathcal{A}'_i, i \in I\})$$

die kleinste σ -Algebra auf Ω , bezüglich derer alle f_i messbar sind.

2.2 Bildmaße oder „push-forward“ eines Maßes

Wir nutzten die Messbarkeit einer $(\mathcal{A}, \mathcal{A}')$ -messbaren Abbildung $f: \Omega \rightarrow \Omega'$, um ein Maß μ auf \mathcal{A} auf ein Maß μ_f auf \mathcal{A}' rüberzuschieben (deshalb „push-forward“). In dem Stochastikteil werden wir noch sehen, dass der push-forward extrem wichtig ist.

Satz 2.2.1.  Sei $f: \Omega \rightarrow \Omega'$ $(\mathcal{A}, \mathcal{A}')$ -messbar und μ ein Maß auf \mathcal{A} . Dann ist

$$\mu_f(B) := \mu(f^{-1}(B)), \quad B \in \mathcal{A}'$$

ein Maß auf \mathcal{A}' . Dieses Maß heißt „Bildmaß“ oder „push-forward“ von f .

Beweis. μ_f ist wohldefiniert weil f messbar ist und daher $f^{-1}(B) \in \mathcal{A}$ gilt. Auf \mathcal{A} ist μ definiert, also macht die Definition von μ_f Sinn. Die Positivität von μ_f folgt natürlich direkt aus der Positivität von μ . Checken wir noch die zwei definierenden Eigenschaften eines Maßes:

- (i) $\mu_f(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$
- (ii) Seien $B_1, B_2, \dots \in \mathcal{A}'$ paarweise disjunkt, dann folgt aus der Definition und den Maßeigenschaften von μ

$$\begin{aligned} \mu_f\left(\bigcup_{k=1}^{\infty} B_k\right) &\stackrel{\text{Def.}}{=} \mu\left(f^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right)\right) \\ &\stackrel{\text{Urbild}}{=} \mu\left(\bigcup_{k=1}^{\infty} f^{-1}(B_k)\right) \\ &\stackrel{\mu \text{ Ma\ss}}{=} \sum_{k=1}^{\infty} \mu\left(f^{-1}(B_k)\right) \stackrel{\text{Def.}}{=} \sum_{k=1}^{\infty} \mu_f(B_k). \end{aligned}$$

Damit ist μ_f auch σ -additiv. □

Beispiel 2.2.2. ▶ Sei $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x + a$. f ist Borel-messbar weil f stetig ist. Sei $\mu := \lambda$ das Lebesgue-Maß auf $\mathcal{B}(\mathbb{R})$, was ist dann der push-forward μ_f ? μ_f ist laut Satz 2.2.1 ein Maß, aber welches?

Behauptung: $\mu_f = \lambda$.

Warum? Berechnen wir dazu μ_f auf einem \cap -stabilen Erzeuger von $\mathcal{B}(\mathbb{R})$:

$$\mu_f((c, d]) \stackrel{\text{Def.}}{=} \mu(f^{-1}((c, d])) = \lambda((c-a, d-a]) = (d-a) - (c-a) = d-c = \lambda((c, d]).$$

Weil $\mathcal{E} = \{(c, d] : c < d\}$ \cap -stabil ist mit $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$, gilt aufgrund von Folgerung 1.2.13 auch $\lambda = \mu_f$ (wir wählen dabei $E_n = (-n, n]$).

Weil a beliebig war, gilt also

$$\lambda(B) = \lambda(B + a)$$

für alle $a \in \mathbb{R}$, wobei $B + a := \{b + a : b \in B\}$ die um a verschobene Menge ist. Man sagt, das Lebesgue-Maß ist **translationsinvariant**, Verschieben von Mengen ändert ihr Maß (die „Größe“) nicht. Diese Eigenschaft gilt natürlich nicht für alle Maße. Mehr noch, bis auf triviale Modifikationen (Konstanten addieren) ist das Lebesgue-Maß das einzige translationsinvariante Maß auf $\mathcal{B}(\mathbb{R})$.

2.3 Messbare numerische Funktionen

Wir nutzen wie in Kapitel 1 die erweiterte Zahlengerade $\overline{\mathbb{R}} = [-\infty, +\infty]$. Dabei nutzen wir die definierten „Rechenregeln“ aus Kapitel 1 und auch die Konvergenzen am Rand:

$$a_n \rightarrow +\infty, n \rightarrow \infty, \quad \text{und} \quad a_n \rightarrow -\infty, n \rightarrow \infty,$$

wie in Analysis 1 definiert. Oft schreiben wir ∞ statt $+\infty$.

Definition 2.3.1. ▶ Auf $\overline{\mathbb{R}}$ definieren wir die erweiterte Borel- σ -Algebra:

$$\mathcal{B}(\overline{\mathbb{R}}) := \{B \subseteq \overline{\mathbb{R}} : B \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\}.$$

Kurz überlegen zeigt uns, dass $\mathcal{B}(\overline{\mathbb{R}})$ folgende Mengen enthält: alle $B \in \mathcal{B}(\mathbb{R})$, sowie $B \cup \{+\infty\}$, $B \cup \{-\infty\}$ und $B \cup \{-\infty, +\infty\}$.

Definition 2.3.2. ▶ Für einen messbaren Raum (Ω, \mathcal{A}) heißt $f: \Omega \rightarrow \overline{\mathbb{R}}$ **messbare numerische Funktion**, falls f $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -messbar ist.

In der Stochastik 1 spielen numerische Funktionen noch keine besonders wichtige Rolle. Ihr solltet euch nicht erschrecken lassen, bei (fast) allen Argumenten spielt es keine Rolle, ob eine Funktion reell oder numerisch ist. Numerische Funktionen sind einfach nur eine etwas größere Klasse von Funktionen, die reelle Funktionen enthalten. Gewöhnt euch einfach direkt daran, dass unsere messbaren Funktionen auch die Werte $+\infty$ oder $-\infty$ annehmen dürfen.

Bemerkung 2.3.3. ▶

- (i) Jede $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ -messbare Funktion $f: \Omega \rightarrow \mathbb{R}$ ist auch eine messbare numerische Funktion, denn $f^{-1}(A \cup B) = f^{-1}(B) \in \mathcal{A}$ für $A \in \{\{+\infty\}, \{-\infty\}, \{+\infty, -\infty\}\}$.
- (ii) Aussagen für messbare reelle Funktionen gelten ganz analog für messbare numerische Funktionen. So gilt etwa: $f: \Omega \rightarrow \overline{\mathbb{R}}$ ist $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -messbar genau dann, wenn $\{f \leq t\} \in \mathcal{A}$ für alle $t \in \overline{\mathbb{R}}$. Das folgt auch aus Proposition 2.1.4 weil $\mathcal{E} = \{[-\infty, t]: t \in \overline{\mathbb{R}}\}$ die σ -Algebra $\mathcal{B}(\overline{\mathbb{R}})$ erzeugt (überlegt mal, warum das stimmt).

Definition 2.3.4. ▶ Für $a, b \in \overline{\mathbb{R}}$ definieren wir

$$a \wedge b := \min\{a, b\} \quad \text{und} \quad a \vee b := \max\{a, b\}$$

sowie

$$a^+ := \max\{0, a\} \quad \text{und} \quad a^- := -\min\{0, a\}.$$

Für numerische Funktionen werden entsprechend punktweise $f \wedge g, f \vee g, f^+, f^-$ definiert. f^+ heißt **Positivteil** von f und f^- **Negativteil** von f .

Beachte: Postivteil und Negativteil sind beide positiv aufgrund des zusätzlichen Minus in der Definition des Negativteils.

Es gelten direkt aus der Definition folgende wichtige Identitäten

$$f = f^+ - f^- \quad \text{und} \quad |f| = f^+ + f^-,$$

die uns zeigen, weshalb es oft reicht f^+ und f^- zu untersuchen.

Lemma 2.3.5. ▶ Sind $f, g: \Omega \rightarrow \overline{\mathbb{R}}$ $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -messbar, so sind die Mengen

$$\{f < g\}, \quad \{f \leq g\}, \quad \{f = g\} \quad \text{und} \quad \{f \neq g\}$$

messbar, also in \mathcal{A} .

Beweis. Der Trick ist es, die Mengen als abzählbare Vereinigungen, Komplemente, Schnitte, etc. von messbaren Mengen zu schreiben. Weil f und g messbar sind, führen wir also auf Urbilder offener Mengen von f und g zurück. Als erstes schreiben wir

$$\{f < g\} \stackrel{\text{Trick!}}{=} \bigcup_{t \in \mathbb{Q}} \{f < t < g\} = \bigcup_{t \in \mathbb{Q}} \underbrace{\{f < t\}}_{\in \mathcal{A}} \cap \underbrace{\{t < g\}}_{\in \mathcal{A}}.$$

Der wesentliche Trick war natürlich die erste Gleichheit. Genauso zeigt man auch $\{f > g\} \in \mathcal{A}$. Weil $\{f = g\} = (\{f < g\} \cup \{f > g\})^C$ und $\{f \neq g\} = \{f = g\}^C$ gelten, sind auch die letzten beiden Mengen in \mathcal{A} . Die zweite Menge schreiben wir als $\{f \leq g\} = \{f < g\} \cup \{f = g\}$, die rechte Seite ist in \mathcal{A} . □

Lemma 2.3.6. ▶ Sind $f, g: \Omega \rightarrow \overline{\mathbb{R}}$ $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -messbar, so sind auch $f + g, \alpha f$ für $\alpha \in \mathbb{R}, f \cdot g, f \wedge g, f \vee g$, und $|f|$ messbar.

Beweis. Tricks aus dem letzten Beweis ausprobieren, und in Übungen/Übungsaufgaben üben! □

Auch sehr wichtig ist, dass punktweise Grenzwerte von Folgen messbarer numerischer Funktionen wieder messbar sind:

Proposition 2.3.7. ▶ Es sei $f_1, f_2, \dots : \Omega \rightarrow \overline{\mathbb{R}}$ eine Folge $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -messbarer numerischer Funktionen.

(i) Dann sind auch die punktweise definierten Funktionen

$$\begin{aligned} - g_1(\omega) &:= \inf_{n \in \mathbb{N}} f_n(\omega), & \omega \in \Omega, \\ - g_2(\omega) &:= \sup_{n \in \mathbb{N}} f_n(\omega), & \omega \in \Omega, \\ - g_3(\omega) &:= \limsup_{n \rightarrow \infty} f_n(\omega), & \omega \in \Omega, \\ - g_4(\omega) &:= \liminf_{n \rightarrow \infty} f_n(\omega), & \omega \in \Omega, \end{aligned}$$

messbare numerische Funktionen. Beachte: Weil wir über numerische Funktionen reden, sind alle Ausdrücke wohldefiniert, die Werte $+\infty$ und $-\infty$ dürfen auftauchen.

(ii) Existieren die Grenzwerte in $\overline{\mathbb{R}}$ für alle $\omega \in \Omega$, so ist auch die punktweise definierte Funktion

$$g(\omega) := \lim_{n \rightarrow \infty} f_n(\omega), \quad \omega \in \Omega,$$

messbar.

Beweis. Der Beweis (und Beispiele) wird in der großen Übung diskutiert, hier nur einer der Fälle. Wie immer reicht es, für alle $t \in \mathbb{R}$, $\{g_i \leq t\} \in \mathcal{A}$ zu zeigen. Die Menge wird wieder geschrieben als abzählbare Vereinigungen, Komplemente, Schnitte, etc. von messbaren Mengen:

$$\{g_1 < t\} = \left\{ \omega \in \Omega : \inf_{n \in \mathbb{N}} f_n(\omega) < t \right\} = \underbrace{\bigcup_{n \in \mathbb{N}} \underbrace{\{\omega \in \Omega : f_n(\omega) < t\}}_{\in \mathcal{A}}}_{\in \mathcal{A}}.$$

□

An dieser Stelle ist noch nicht so klar, warum Messbarkeit nützlich ist. Die gerade gezeigten Aussagen sind der Grund, weshalb die im Anschluss zu entwickelnde Lebesgue Integrationstheorie so erfolgreich ist: Alle möglichen Manipulationen mit messbaren Funktionen bleiben messbar.

Kapitel 3

Integrationstheorie

Vorlesung 9

Im Folgenden entwickeln wir die Integrationstheorie im Sinne von Henry Lebesgues. An und für sich hat das nichts mit Stochastik zu tun und gehört eher in den Bereich der Analysis. Im Sinne der abstrakten Modellierung eines zufälligen Experimentes durch einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ wollen wir aber Lebesgue Integrale nutzen, um Begriffe wie Erwartungswert und Varianz als Integrale über $(\Omega, \mathcal{A}, \mathbb{P})$ definieren. Damit kann man dann in der Stochastik die Massenverteilung zufälliger Experimente genauer untersuchen.

3.1 Das (allgemeine) Lebesgue Integral

In diesem Abschnitt werden wir Integrale der Form $\int_{\Omega} f d\mu$ für beliebige Maßräume $(\Omega, \mathcal{A}, \mu)$ und messbare numerische Funktionen $f : \Omega \rightarrow \bar{\mathbb{R}}$ definieren. Das Maß kann endlich sein (z. B. ein Wahrscheinlichkeitsmaß) oder unendlich sein (z. B. das Lebesguemaß auf $\mathcal{B}(\mathbb{R})$).

Bevor wir mit der Konstruktion starten, diskutieren wir ganz kurz den Zusammenhang zum Riemann Integral, das ihr vermutlich aus der Schule oder den Analysis Vorlesungen kennt. Dort habt ihr für reelle Funktionen Integrale $\int_a^b f(x) dx$ definiert, indem Treppenfunktionen (stückweise konstant auf einer Zerlegung von $[a, b]$ in kleine Intervalle) über und unter den Graphen von f gelegt wurden. Wenn die Treppenfunktionen von oben und unten immer feiner an f angenähert werden und dabei die Integrale (Ober- und Untersummen) im Grenzwert gleich sind, so heißt f Riemann integrierbar und das Integral von f ist als dieser Grenzwert definiert (wer alles vergessen hat, kann mal schnell die Bildchen bei Wikipedia zum Riemann Integral anschauen). Die Interpretation des Integrals als Flächeninhalt wird dadurch visuell klar. Anschließend habt ihr das uneigentliche Riemann Integral $\int_{\mathbb{R}} f(x) dx$ als Grenzwert des (eigentlichen) Riemann Integrals $\lim_{n \rightarrow \infty} \int_{-n}^n f(x) dx$ definiert, falls der Grenzwert existiert. Wenn ihr also eigentliche Riemann Integrale durch Stammfunktionen, partielle Integration oder Substitution berechnen könnt, so könnt ihr auch uneigentliche Riemann Integrale berechnen.

Wenn wir jetzt für $f : \Omega \rightarrow \mathbb{R}$ statt für $f : \mathbb{R} \rightarrow \mathbb{R}$ genauso vorgehen wollen, haben wir ein Problem: Wie zerlegen wir Ω in kleine Intervalle? Das geht nicht einfach so, Ω ist schließlich eine völlig beliebige Menge! Was ist aber in beiden Fällen gleich? Der Bildbereich! Der Trick beim Lebesgue Integral ist deshalb, nicht das Urbild in Intervalle zu zerlegen, sondern den Bildbereich in Intervalle zu zerlegen! Warum dafür gerade messbare Funktionen geeignet sind, wird in Satz 3.1.6 deutlich werden. Die Idee von Lebesgue den Bildbereich zu zerteilen, wird es uns daher später erlauben, Erwartungswerte, Varianzen, etc. für beliebige zufällige Experimente zu definieren.

Um leichter zu folgen, kann es nützlich sein, den Spezialfall $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ im Kopf zu halten, denn da können wir besser zeichnen. Wir schreiben in dem Fall statt $\int_{\mathbb{R}} f d\lambda$ auch $\int_{\mathbb{R}} f(x) dx$, um klar zu stellen, dass das Lebesgue Integral für viele „nette“ Integranden $f : \mathbb{R} \rightarrow \mathbb{R}$ das gleiche ist, wie das (uneigentliche) Riemann Integral. Merkt euch für später schon mal, dass beispielsweise für

- nicht-negative stückweise stetige Integranden,

- stückweise stetige Integranden, die außerhalb eines Intervalls 0 sind,

das neue Lebesgue Integral und das schon bekannte uneigentliche Riemann Integral gleich sind. Später wird das nützlich sein, weil ihr dann die Rechenregeln der Analysis 1 (oder Schule) nutzen könnt, um $\int_{\mathbb{R}} f \, d\lambda$ als Grenzwert von $\int_{-n}^n f(x) \, dx$ auszurechnen. Alternativ könnten wir die Rechenregeln aus der Analysis nochmal für das Lebesgue Integral nachrechnen, aber das wäre vielleicht etwas langweilig.

Genug der Vorrede, kommen wir nun zum Lebesgue Integral:

Definition 3.1.1.  Eine messbare Abbildung $f: \Omega \rightarrow \overline{\mathbb{R}}$ heißt **einfach** (alternativ **elementar**, manchmal auch **Treppenfunktion**), falls f nur endlich viele Werte annimmt. Wir definieren auch noch

$$\mathcal{E} = \{f \mid f \text{ einfache Funktion}\} \quad \text{und} \quad \mathcal{E}^+ = \{f \mid f \text{ einfache Funktion, } f \geq 0\}.$$

Eine Darstellung der Form

$$f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k} \tag{3.1}$$

nennen wir disjunkte Darstellung, wenn $\alpha_1, \dots, \alpha_n \in \overline{\mathbb{R}}$ und $A_1, \dots, A_n \in \mathcal{A}$ paarweise disjunkt sind. Nimmt eine einfache Funktion die Werte $\alpha_1, \dots, \alpha_n$ an, so gilt (3.1) zum Beispiel mit den messbaren Mengen

$$A_k = \{f = \alpha_k\} = \{\omega: f(\omega) = \alpha_k\} = f^{-1}([\alpha_k, \alpha_k]) \in \mathcal{A}.$$

Bemerkung. Wenn wir von einfachen Funktionen sprechen, meinen wir also immer, dass entweder f endlich viele Werte annimmt, oder f die obige Darstellung als Summe von Indikatorfunktionen hat. Meistens nutzen wir aber die disjunkten Darstellungen weil die für Integrale benötigt werden.

Disjunkte Darstellungen messbarer Funktionen sind nicht eindeutig, z. B. gilt

$$\mathbf{1}_{[-2,-1]} + 2 \cdot \mathbf{1}_{[1,2]} = \mathbf{1}_{[-2,-3/2]} + \mathbf{1}_{(-3/2,-1]} + 2 \cdot \mathbf{1}_{[1,2]}.$$

Wir definieren im Folgenden das Integral nicht-negativer einfacher Funktionen, dann durch Approximation das Integral nicht-negativer messbarer Funktionen und schließlich durch die Zerlegen $f = f^+ - f^-$ das Integral beliebiger messbarer Funktionen.

Integrale nicht-negativer einfacher Funktionen

Definition 3.1.2.  Für $f \in \mathcal{E}^+$ definiert man

$$\int_{\Omega} f \, d\mu := \sum_{k=1}^n \alpha_k \mu(A_k) \in [0, +\infty],$$

wenn f die disjunkte Darstellung $f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$ hat. $\int_{\Omega} f \, d\mu$ heißt Integral von f bezüglich μ , f heißt Integrand. Weil $\alpha_k = +\infty$ sowie $\mu(A_k) = +\infty$ möglich sind, muss festgelegt werden, wie $+$ und \cdot mit ∞ geht. Siehe dazu die Definition in Abschnitt 1.1.

Beispiel. Weil das Integral den „Flächeninhalt“ zwischen Graphen und Achse beschreiben soll,

sind die Rechenregeln $+$ und \cdot mit ∞ durchaus sinnvoll definiert worden:

$$\begin{aligned}\int_{\mathbb{R}} 0 \cdot \mathbf{1}_{\mathbb{R}} \, d\lambda &= 0 \cdot \lambda(\mathbb{R}) = 0 \cdot (+\infty) = 0, \\ \int_{\mathbb{R}} \mathbf{1}_{\mathbb{R}} \, d\lambda &= 1 \cdot \lambda(\mathbb{R}) = 1 \cdot (+\infty) = +\infty, \\ \int_{\mathbb{R}} (+\infty) \cdot \mathbf{1}_{[a,b]} \, d\lambda &= +\infty \cdot \lambda([a,b]) = +\infty \cdot (b-a) = +\infty, \\ \int_{\mathbb{R}} 3 \cdot \mathbf{1}_{[0,1]} \, d\lambda &= 3 \cdot \lambda([0,1]) = 3.\end{aligned}$$

Was sollten die Integrale auch sonst sein?

Rechnen wir noch nach, dass das Integral einer nicht-negativen einfachen Funktion nicht von der disjunkten Darstellung abhängt. Weil es verschiedene disjunkte Darstellungen für die gleiche Funktion gibt, würde die Definition sonst keinen Sinn machen.

Lemma 3.1.3.  Es gelte

$$\sum_{k=1}^n \alpha_k \mathbf{1}_{A_k} = \sum_{l=1}^m \beta_l \mathbf{1}_{B_l}$$

mit $\alpha_k, \beta_l \geq 0$ und paarweise disjunkten $A_1, \dots, A_n, B_1, \dots, B_m$, so gilt

$$\sum_{k=1}^n \alpha_k \mu(A_k) = \sum_{l=1}^m \beta_l \mu(B_l).$$

Beweis. Ohne Einschränkung der Allgemeinheit seien alle $\alpha_k, \beta_l \neq 0$. Wegen $\bigcup_{k=1}^n A_k = \{f > 0\} = \bigcup_{l=1}^m B_l$ und der σ -Additivität von μ gilt dann

$$\begin{aligned}\sum_{k=1}^n \alpha_k \mu(A_k) &\stackrel{\sigma\text{-add.}}{=} \sum_{k=1}^n \alpha_k \sum_{l=1}^m \mu(A_k \cap B_l) = \sum_{k=1}^n \sum_{l=1}^m \alpha_k \mu(A_k \cap B_l) \\ &\stackrel{(\star)}{=} \sum_{l=1}^m \sum_{k=1}^n \beta_l \mu(B_l \cap A_k) = \sum_{l=1}^m \beta_l \mu(B_l).\end{aligned}$$

(\star) gilt, weil entweder $\mu(A_k \cap B_l) = 0$ oder $\mu(A_k \cap B_l) > 0$ gilt. Im ersten Fall gilt $\alpha_k \mu(A_k \cap B_l) = 0 = \beta_l \mu(A_k \cap B_l)$ trivialerweise, im zweiten Fall impliziert $\mu(A_k \cap B_l) > 0$ schon $A_k \cap B_l \neq \emptyset$ und damit $\alpha_k = \beta_l$ weil die beiden Darstellungen disjunkt sind. \square

Lemma 3.1.4.  Für $f, g \in \mathcal{E}^+$, $\alpha \geq 0$ und $A \in \mathcal{A}$ gelten

- (i) $\mathbf{1}_A \in \mathcal{E}^+$ und $\int_{\Omega} \mathbf{1}_A \, d\mu = \mu(A)$.
- (ii) $\alpha f \in \mathcal{E}^+$ und $\int_{\Omega} \alpha f \, d\mu = \alpha \int_{\Omega} f \, d\mu$,
- (iii) $f + g \in \mathcal{E}^+$ und $\int_{\Omega} (f + g) \, d\mu = \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu$,
- (iv) $f \leq g \Rightarrow \int_{\Omega} f \, d\mu \leq \int_{\Omega} g \, d\mu$.

Beweis.

- (i) \checkmark

- (ii) αf nimmt auch nur endlich viele Werte an (ist also eine einfache Funktion) und hat die disjunkte Darstellung

$$\alpha f = \alpha \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k} = \sum_{k=1}^n (\alpha \alpha_k) \mathbf{1}_{A_k}.$$

Das Integral berechnet sich aus der Definition:

$$\int_{\Omega} (\alpha f) \, d\mu \stackrel{\text{Def.}}{=} \sum_{k=1}^n (\alpha \alpha_k) \mu(A_k) = \alpha \sum_{k=1}^n \alpha_k \mu(A_k) \stackrel{\text{Def.}}{=} \alpha \int_{\Omega} f \, d\mu$$

- (iii) Wir nehmen an, dass f und g die disjunkten Darstellungen

$$f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k} \quad \text{und} \quad g = \sum_{l=1}^m \beta_l \mathbf{1}_{B_l}$$

haben. Ohne Einschränkung gelte $\bigcup_{k=1}^n A_k = \Omega = \bigcup_{l=1}^m B_l$. Wäre das nicht der Fall, so würden wir $A_{n+1} := (\bigcup_{k=1}^n A_k)^C$ und $\alpha_{n+1} = 0$ wählen (analog für g). Damit ist dann wegen $\mathbf{1} = \mathbf{1}_{\Omega} = \sum_{k=1}^n \mathbf{1}_{A_k} = \sum_{l=1}^m \mathbf{1}_{B_l}$ und $\mathbf{1}_{A_k} \cdot \mathbf{1}_{B_l} = \mathbf{1}_{A_k \cap B_l}$ auch

$$\begin{aligned} f + g &= \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k} + \sum_{l=1}^m \beta_l \mathbf{1}_{B_l} \\ &= \sum_{k=1}^n \alpha_k \sum_{l=1}^m \mathbf{1}_{A_k \cap B_l} + \sum_{l=1}^m \beta_l \sum_{k=1}^n \mathbf{1}_{A_k \cap B_l} \\ &= \sum_{k=1}^n \sum_{l=1}^m (\alpha_k + \beta_l) \mathbf{1}_{A_k \cap B_l}. \end{aligned}$$

Damit haben wir eine disjunkte Darstellung für die einfache Funktion $f + g$ und es gilt

$$\begin{aligned} \int_{\Omega} (f + g) \, d\mu &\stackrel{\text{Def.}}{=} \sum_{k=1}^n \sum_{l=1}^m (\alpha_k + \beta_l) \mu(A_k \cap B_l) \\ &\stackrel{\sigma\text{-add.}}{=} \sum_{k=1}^n \alpha_k \mu\left(\bigcup_{l=1}^m (A_k \cap B_l)\right) + \sum_{l=1}^m \beta_l \mu\left(\bigcup_{k=1}^n (A_k \cap B_l)\right) \\ &= \sum_{k=1}^n \alpha_k \mu(A_k) + \sum_{l=1}^m \beta_l \mu(B_l) \\ &\stackrel{\text{Def.}}{=} \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu, \end{aligned}$$

und damit die Behauptung.

- (iv) Monotonie \checkmark , folgt direkt aus der Definition. □

Integral nicht-negativer messbarer numerischer Funktionen

Weiter geht's für nicht-negative Integranden. Wir definieren zunächst einen sofort wohldefinierten Ausdruck und zeigen danach, dass dies auch der Grenzwert beliebiger wachsender Folgen von unten ist.

Definition 3.1.5. ▶ Für $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -messbares $f : \Omega \rightarrow \overline{\mathbb{R}}$ mit $f \geq 0$ definieren wir

$$\int_{\Omega} f \, d\mu = \sup \left\{ \int_{\Omega} g \, d\mu : 0 \leq g \leq f, g \in \mathcal{E}^+ \right\}.$$

$\int_{\Omega} f \, d\mu$ heißt wieder Integral von f bezüglich μ , f heißt Integrand. Wie bei nicht-negativen einfachen Funktionen ist $\int_{\Omega} f \, d\mu = +\infty$ ausdrücklich erlaubt!

Jetzt wollen wir diese komplizierte Definition (wie soll man damit irgendwas zeigen?) durch eine handlichere äquivalente Darstellung ersetzen.

Satz 3.1.6. ▶ [Darum sind messbare Funktionen so wichtig!!!]

Für jede nicht-negative messbare numerische Funktion existiert eine wachsende Folge von Treppenfunktionen $(f_n)_{n \in \mathbb{N}} \subseteq \mathcal{E}^+$ mit $f_n \uparrow f$ punktweise für $n \rightarrow \infty$.

Beweis. Wir definieren

$$f_n = \underbrace{\sum_{k=0}^{n \cdot 2^n - 1} \frac{k}{2^n} \mathbf{1}_{f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right)} + n \mathbf{1}_{f^{-1}([n, +\infty])}}_{\text{messbar}}.$$

Fürs bessere Verständnis zeichne man die Folge f_n für das Beispiel $f : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ mit $f(x) = +\infty \cdot \mathbf{1}_{(-\infty, 0]} + \frac{1}{x} \cdot \mathbf{1}_{(0, +\infty)}$ hin! Weil f messbar ist, sind die A_k messbare Mengen. Also sind die f_n einfache Funktionen. Aufgrund der Definition gelten sofort die geforderten Eigenschaften:

- $0 \leq f_n \leq f$ für alle $n \in \mathbb{N}$.
- Die Folge (f_n) ist punktweise wachsend.
- Die Folge (f_n) konvergiert punktweise gegen f .

□

Lemma 3.1.7. ▶ [Montone Konvergenz Theorem (MCT) für einfache Funktionen]

Sei $(f_n) \subseteq \mathcal{E}^+$ mit $f_n \uparrow f$, $n \rightarrow \infty$, für eine nicht-negative messbare numerische Funktion f . Dann gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu,$$

wobei in der Gleichheit $+\infty = +\infty$ möglich ist.

Für monoton wachsende Folgen einfacher Funktionen darf der Limes also in das Integral getauscht werden.

Beweis. Die Folge $(\int_{\Omega} f_n \, d\mu)_{n \in \mathbb{N}}$ wächst (Monotonie des Integrals für einfache Integranden) und konvergiert also in $[0, +\infty]$.

„ \leq “: Folgt direkt aus der Definition

$$\int_{\Omega} f \, d\mu = \sup \left\{ \int_{\Omega} g \, d\mu : g \leq f, g \in \mathcal{E}^+ \right\},$$

weil das Supremum einer Menge eine obere Schranke der Menge ist und $f_n \leq f$ nach Voraussetzung.

„ \geq “: Wir behaupten: Ist $g \in \mathcal{E}^+$ mit $g \leq f$, so gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \geq \int_{\Omega} g \, d\mu. \quad (3.2)$$

Weil das Supremum einer Menge M die *kleinste* obere Schranke ist, sind wir dann fertig weil aufgrund von (3.2) auch $\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu$ eine obere Schranke ist.

Warum gilt die Behauptung? Sei $\varepsilon \in (0, 1)$ beliebig und sei

$$g = \sum_{k=1}^r \gamma_k \mathbf{1}_{C_k} \in \mathcal{E}^+ \quad \text{mit} \quad g \leq f.$$

Wegen $f_n \uparrow f$ gilt $A_n \uparrow \Omega$, $n \rightarrow \infty$, für

$$A_n := \{f_n \geq (1 - \varepsilon)g\} = \{\omega : f_n(\omega) \geq (1 - \varepsilon)g(\omega)\}.$$

Weil aufgrund der Definition der Mengen A_n und $f \geq 0$

$$f_n(\omega) \geq f_n(\omega) \mathbf{1}_{A_n}(\omega) \geq (1 - \varepsilon)g(\omega) \mathbf{1}_{A_n}(\omega)$$

für alle $\omega \in \Omega$ gilt (man teste die zwei Möglichkeiten $\omega \in A_n$ und $\omega \notin A_n$), folgt

$$\begin{aligned} \int_{\Omega} f_n \, d\mu &\stackrel{\text{Mon.}}{\geq} \int_{\Omega} f_n \mathbf{1}_{A_n} \, d\mu \\ &\stackrel{\text{Mon.}}{\geq} \int_{\Omega} (1 - \varepsilon)g \mathbf{1}_{A_n} \, d\mu \\ &\stackrel{\text{Lin.}}{=} (1 - \varepsilon) \int_{\Omega} \left(\sum_{k=1}^r \gamma_k \mathbf{1}_{C_k} \right) \mathbf{1}_{A_n} \, d\mu \\ &= (1 - \varepsilon) \int_{\Omega} \sum_{k=1}^r \gamma_k \mathbf{1}_{A_n \cap C_k} \, d\mu \\ &\stackrel{\text{Def.}}{=} (1 - \varepsilon) \sum_{k=1}^r \gamma_k \mu(A_n \cap C_k). \end{aligned}$$

Wegen Stetigkeit von Maßen gilt

$$\lim_{n \rightarrow \infty} \mu(A_n \cap C_k) = \mu\left(\bigcup_{n=1}^{\infty} (A_n \cap C_k)\right) = \mu\left(\underbrace{\left(\bigcup_{n=1}^{\infty} A_n\right)}_{=\Omega} \cap C_k\right) = \mu(C_k),$$

also gilt zusammen

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \geq (1 - \varepsilon) \sum_{k=1}^r \gamma_k \mu(C_k) = (1 - \varepsilon) \int_{\Omega} g \, d\mu.$$

Weil ε beliebig gewählt war folgt die Hilfsbehauptung und damit ist der Beweis fertig. □

Vorlesung 10

Warum war das Lemma so wichtig? Die Definition des Integrals als Supremum ist sehr unhandlich. Es hat natürlich den Vorteil, dass das Integral sofort sinnvoll definiert ist, dafür können wir mit der Definition nichts anstellen. Schauen wir uns als Beispiel die Beweise der folgenden elementaren Rechenregeln an. Per Approximation durch einfache Funktionen sind die Argumente sehr einfach, per Definition als Supremum wären die Argumente ziemlich fies.

Lemma 3.1.8. ▶ Für $f, g: \Omega \rightarrow [0, +\infty]$ messbar und $\alpha \geq 0$ gelten

- (i) $\int_{\Omega} \alpha f \, d\mu = \alpha \int_{\Omega} f \, d\mu,$
(ii) $\int_{\Omega} (f + g) \, d\mu = \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu,$
(iii) $f \leq g \Rightarrow \int_{\Omega} f \, d\mu \leq \int_{\Omega} g \, d\mu.$

Beweis. Wir zeigen nur (ii), (i) geht analog und (iii) folgt direkt aus der Definition als Supremum. Beachtet dabei folgende Eigenschaften vom Supremum: $M \subseteq N$ impliziert natürlich $\sup M \leq \sup N$.

Seien $(f_n), (g_n) \subseteq \mathcal{E}^+$ mit $f_n \uparrow f, g_n \uparrow g, n \rightarrow \infty$. Weil dann auch $f_n + g_n \in \mathcal{E}^+$ und $f_n + g_n \uparrow f + g$ gelten, folgt mit Lemma 3.1.7 und der Linearität des Integrals für einfache Funktionen

$$\int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu = \lim_{n \rightarrow \infty} \left(\int_{\Omega} f_n \, d\mu + \int_{\Omega} g_n \, d\mu \right) = \lim_{n \rightarrow \infty} \int_{\Omega} (f_n + g_n) \, d\mu = \int_{\Omega} (f + g) \, d\mu.$$

□

Integral messbarer numerischer Funktionen

Im letzten Schritt wollen wir noch die Annahme der Nichtnegativität weglassen. Sei dazu $f: \Omega \rightarrow \overline{\mathbb{R}}$ ($\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}})$)-messbar. Um f auf nicht-negative Funktionen zurückzuführen, erinnern wir an die Zerlegung von f in Postiv- und Negativteil

$$f = f^+ - f^- \quad \text{und} \quad |f| = f^+ + f^-$$

aus dem Kapitel über messbare Abbildungen.

Definition 3.1.9. ▶ Sei $f: \Omega \rightarrow \overline{\mathbb{R}}$ messbar und

$$\int_{\Omega} f^+ \, d\mu < \infty \quad \text{oder} \quad \int_{\Omega} f^- \, d\mu < \infty.$$

Dann definieren wir

$$\int_{\Omega} f \, d\mu = \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu \in [-\infty, +\infty]$$

und sagen, das Integral $\int f \, d\mu$ ist wohldefiniert. Ist $\int f \, d\mu \in \mathbb{R}$, d.h. die Integrale über Positiv- und Negativteil sind beide endlich, so heißt f μ -integrierbar und wir sagen, das Integral existiert. Existiert bedeutet also wohldefiniert und endlich. Zur Notation: Man schreibt statt $\int_{\Omega} f \, d\mu$ auch

$$\int_{\Omega} f(\omega) \, d\mu(\omega) \quad \text{oder} \quad \int_{\Omega} f(\omega) \, \mu(d\omega),$$

oft lässt man aus Faulheit auch Ω unter dem Integral weg.

Ohne die Einschränkung, dass eines der Integrale endlich ist, könnten wir das Integral nicht sinnvoll definieren. Das liegt daran, dass $+\infty - (+\infty)$ nicht sinnvoll definierbar ist.

Definition 3.1.10. ▶ Ist $f: \Omega \rightarrow \overline{\mathbb{R}}$ messbar und $A \in \mathcal{A}$, so definiert man

$$\int_A f \, d\mu := \int_{\Omega} f \mathbf{1}_A \, d\mu,$$

wenn die rechte Seite wohldefiniert ist. Alternativ schreibt man auch hier

$$\int_A f(\omega) \, d\mu(\omega) \quad \text{oder} \quad \int_A f(\omega) \, \mu(d\omega)$$

und wenn $\Omega = \mathbb{R}$ ist, auch $\int_{\mathbb{R}} f(x) d\mu(x)$. Für den Spezialfall des Lebesgue-Maßes auf $\mathcal{B}(\mathbb{R})$ schreiben wir wieder

$$\int_a^b f(x) dx \quad \text{statt} \quad \int_{[a,b]} f d\lambda,$$

damit die Analogie zur Analysis nicht verloren geht. Weil das Integral nur für messbare Funktionen definiert ist, ist es ganz essentiell, dass auch $f\mathbf{1}_A$ eine messbare Funktion ist. Das liegt an der großen Flexibilität von messbaren Funktionen: $\mathbf{1}_A$ ist messbar, weil A messbar ist und das Produkt messbarer Funktionen ist messbar.

Lemma 3.1.11.  Für $f, g: \Omega \rightarrow \overline{\mathbb{R}}$ μ -integrierbar und $\alpha \in \mathbb{R}$ gelten

(i) αf ist μ -integrierbar und

$$\int_{\Omega} \alpha f d\mu = \alpha \int_{\Omega} f d\mu.$$

(ii) Wenn $f + g$ sinnvoll definiert ist (d. h. kein $+\infty + (-\infty)$), so ist $f + g$ μ -integrierbar und

$$\int_{\Omega} (f + g) d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu.$$

(iii)

$$f \leq g \quad \Rightarrow \quad \int_{\Omega} f d\mu \leq \int_{\Omega} g d\mu$$

(iv) Δ -Ungleichung:

$$\left| \int_{\Omega} f d\mu \right| \leq \int_{\Omega} |f| d\mu$$

Beweis.

(i) Für $\alpha \geq 0$ gelten

$$(\alpha f)^+ = \alpha f^+ \quad \text{und} \quad (\alpha f)^- = \alpha f^-.$$

Damit ist αf μ -integrierbar, weil mit Lemma 3.1.8(i)

$$\int_{\Omega} \alpha f^+ d\mu = \alpha \int_{\Omega} f^+ d\mu < \infty \quad \text{und} \quad \int_{\Omega} \alpha f^- d\mu = \alpha \int_{\Omega} f^- d\mu < \infty$$

gelten. Es gilt dann per Definition des Integrals als Differenz der Integrale über Positiv- und Negativteil

$$\int_{\Omega} \alpha f d\mu \stackrel{\text{Def.}}{=} \int_{\Omega} \alpha f^+ d\mu - \int_{\Omega} \alpha f^- d\mu \stackrel{3.1.8}{=} \alpha \int_{\Omega} f^+ d\mu - \alpha \int_{\Omega} f^- d\mu = \alpha \int_{\Omega} f d\mu.$$

Der Fall $\alpha < 0$ geht genauso, wir nutzen hierbei $(\alpha f)^+ = -\alpha f^-$ und $(\alpha f)^- = -\alpha f^+$ und gehen dann genauso vor.

(ii) Die Summe ist bei Integralen immer der delikate Teil. Es gelten zunächst punktweise (Fallunterscheidungen)

$$0 \leq (f + g)^+ \leq f^+ + g^+ \quad \text{und} \quad 0 \leq (f + g)^- \leq f^- + g^-.$$

Damit gelten

$$\int_{\Omega} (f + g)^+ d\mu \stackrel{3.1.8}{\leq} \int_{\Omega} (f^+ + g^+) d\mu \stackrel{3.1.8}{=} \int_{\Omega} f^+ d\mu + \int_{\Omega} g^+ d\mu < \infty$$

und

$$\int_{\Omega} (f + g)^{-} d\mu \stackrel{3.1.8}{\leq} \int_{\Omega} (f^{-} + g^{-}) d\mu \stackrel{3.1.8}{=} \int_{\Omega} f^{-} d\mu + \int_{\Omega} g^{-} d\mu < \infty.$$

Also ist gemäß Definition $f + g$ μ -integrierbar. Die Berechnung des Integrals von $f + g$ ist clever. Wir kennen die Linearität bisher nur für nicht-negative Funktionen. Führen wir die Behauptung also auf den Fall zurück, indem wir wie folgt $f + g$ auf zwei Arten in Positiv- und Negativteil zerlegen:

$$(f + g)_{\geq 0}^{+} - (f + g)_{\geq 0}^{-} = f + g = (f_{\geq 0}^{+} - f_{\geq 0}^{-}) + (g_{\geq 0}^{+} - g_{\geq 0}^{-}).$$

Umformen ergibt

$$(f + g)^{+} + f^{-} + g^{-} = (f + g)^{-} + f^{+} + g^{+}.$$

Weil jetzt nur noch Summen nicht-negativer Funktionen auftauchen, können wir die bereits bekannte Linearität des Integrals aus Lemma 3.1.8 nutzen:

$$\int_{\Omega} (f + g)^{+} d\mu + \int_{\Omega} f^{-} d\mu + \int_{\Omega} g^{-} d\mu = \int_{\Omega} (f + g)^{-} d\mu + \int_{\Omega} f^{+} d\mu + \int_{\Omega} g^{+} d\mu.$$

Erneutes Auflösen ergibt

$$\int_{\Omega} (f + g)^{+} d\mu - \int_{\Omega} (f + g)^{-} d\mu = \int_{\Omega} f^{+} d\mu - \int_{\Omega} f^{-} d\mu + \int_{\Omega} g^{+} d\mu - \int_{\Omega} g^{-} d\mu$$

und Ausnutzen der Definition des Integrals als Differenz der Positiv- und Negativteile

$$\int_{\Omega} (f + g) d\mu = \int_{\Omega} f d\mu + \int_{\Omega} g d\mu.$$

- (iii) Natürlich gilt $f \leq g \Leftrightarrow g - f \geq 0$. Weil die Nullfunktion sowie $g - f$ nicht-negativ sind, gilt wegen der Linearität und der Definition des Integrals für einfache Funktionen (die Nullfunktion)

$$0 = \int_{\Omega} 0 d\mu \leq \int_{\Omega} (g - f) d\mu \stackrel{(i),(ii)}{=} \int_{\Omega} g d\mu - \int_{\Omega} f d\mu.$$

Umformen gibt die Behauptung.

- (iv) Die Dreiecksungleichung für Integrale folgt aus der Dreiecksungleichung in \mathbb{R} :

$$\begin{aligned} \left| \int_{\Omega} f d\mu \right| &\stackrel{\text{Def.}}{=} \left| \int_{\Omega} f^{+} d\mu - \int_{\Omega} f^{-} d\mu \right| \\ &\triangleq \left| \int_{\Omega} f^{+} d\mu \right| + \left| \int_{\Omega} f^{-} d\mu \right| \\ &\stackrel{\geq 0}{=} \int_{\Omega} f^{+} d\mu + \int_{\Omega} f^{-} d\mu \\ &\stackrel{\text{Linear}}{=} \int_{\Omega} (f^{+} + f^{-}) d\mu = \int_{\Omega} |f| d\mu. \end{aligned}$$

□

Eine Konsequenz der gerade genutzten Linearität kombiniert mit $|f| = f^{+} + f^{-}$ ist folgendes Äquivalenz:

$$f \text{ ist } \mu\text{-integrierbar} \iff \int_{\Omega} f d\mu \text{ existiert} \iff \int_{\Omega} |f| d\mu < \infty. \quad (3.3)$$

Die rechte Seite sieht sehr viel nützlicher aus als die eigentliche Definition von μ -Integrierbarkeit, ist aber nur eine recht triviale Modifikation.

Beispiel 3.1.12. 

(i) Für den wichtigen Spezialfall $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ schauen wir uns mal ein paar Beispiele an:

- (a) Ist f stückweise stetig, so ist f Riemann integrierbar auf Intervallen $[a, b]$ und gleich dem Lebesgue Integral $\int_{[a,b]} f \, d\lambda$. Das erklärt auch wieder, warum wir auch für das Lebesgue Integral die dx -Notation nutzen und $\int_{[a,b]} f \, d\lambda = \int_a^b f(x) \, dx$ schreiben. Ihr könnt also für nette Integranden $\int_{[a,b]} f \, d\lambda$ mit den Rechenregeln aus der Schule und Analysis berechnen (Stammfunktionen, partielle Integration, Substitution). Stückweise stetig ist eigentlich nicht die richtige Annahme, die richtige Formulierung ist folgende (Lebesgue'sches Kriterium für Riemann-Integrierbarkeit): Eine Funktion ist Riemann integrierbar auf $[a, b]$ genau dann, wenn die Unstetigkeitsstellen eine Nullmenge sind. Solch eine Funktion ist auch Lebesgue integrierbar und die Integrale sind gleich. Weil wir uns in dieser Vorlesung nicht für das Riemann Integral interessieren, führen wir das nicht weiter aus.
- (b) Ist $f \geq 0$ und stückweise stetig, so stimmt $\int_{\mathbb{R}} f \, d\lambda$ mit dem uneigentlichen Riemann Integral überein. Das sehen wir später mit dem monotonen Konvergenz Theorem und (a). Ihr dürft das Integral in dem Fall als

$$\int_{\mathbb{R}} f \, d\lambda = \lim_{n \rightarrow \infty} \int_{-n}^n f(x) \, dx \quad (3.4)$$

schreiben, und die rechte Seite mit den Tricks der Analysis berechnen. Deshalb schreiben wir auch wie in Analysis $\int_{\mathbb{R}} f(x) \, dx$ statt $\int_{\mathbb{R}} f \, d\lambda$. Warum das gilt, schauen wir uns nach Satz 3.2.1 nochmal an.

- (c) Warnung: Den Rechentrick aus (3.4) dürft ihr nicht immer nutzen, wir haben schließlich angenommen, dass f nicht-negativ ist. Hier sind zwei Gegenbeispiele:

$$f(x) = \frac{\sin(x)}{x}, \quad x \in \mathbb{R},$$

$$f = \sum_{k=1}^{\infty} \mathbf{1}_{[k-1, k)} (-1)^k \frac{1}{k}.$$

Bei beiden Beispielen sind Positiv- und Negativteil nicht integrierbar (das kann man mit (b) nachrechnen), $\int_{\mathbb{R}} f \, d\lambda$ ist also nicht einmal wohldefiniert, der Grenzwert $\lim_{n \rightarrow \infty} \int_{-n}^n f(x) \, dx$ existiert jedoch. Für das Riemann Integral (auf einem Intervall) gilt zwar, „Riemann integrierbar impliziert Lebesgue integrierbar und die Integrale sind gleich“, für das uneigentliche Riemann Integral gilt das weder der Beispiele jedoch nicht!

- (d) In (a) haben wir gesagt, dass Riemann integrierbare Funktionen auch Lebesgue integrierbar sind. Hier ist ein Beispiel, das zeigt, dass die Umkehrung nicht immer gilt: $f = \mathbf{1}_{\mathbb{Q} \cap [0, 1]}$, die Funktion die nur den Wert 1 für rationale Zahlen in $[0, 1]$ annimmt. Weil f eine einfache Funktion ist, $f = 1 \cdot \mathbf{1}_A$ für die Borel-messbare Menge $A = \mathbb{Q} \cap [0, 1]$, ist sie Lebesgue integrierbar mit Integral $\int_{[0, 1]} \mathbf{1}_{\mathbb{Q}} \, d\lambda = 1 \cdot \lambda(\mathbb{Q} \cap [0, 1]) = 0$. Die Funktion ist aber nicht Riemann integrierbar: Jede Treppenfunktion über kleine Intervalle, die über f liegt, ist auf $[0, 1]$ größer als 1. Liegt eine Treppenfunktion unter f , so ist sie kleiner als 0 auf $[0, 1]$. Damit können die Ober- und Untersummen nicht gegen den gleichen Wert konvergieren.

(ii) Schauen wir uns jetzt das Beispiel $(\Omega, \mathcal{A}, \mu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$, $\mu(A) = \#A$, an. Ganz am Anfang der Vorlesung hatten wir dieses Maß Zählmaß genannt. An diesem Beispiel lernen wir: Summen sind auch nur Integrale! Warum? Für $f : \mathbb{N} \rightarrow [0, \infty)$ gilt

$$\int_{\mathbb{N}} f \, d\mu = \sum_{k=0}^{\infty} f(k).$$

Das folgt direkt aus Lemma 3.1.7 weil f als $f = \sum_{k=0}^{\infty} f(k)\mathbf{1}_{\{k\}}$ geschrieben werden kann (setzt mal m in die rechte Seite ein, es ist immer nur ein Summand ungleich 0!) und damit $f_n \uparrow f$ gilt, mit der Folge $f_n := \sum_{k=1}^n f(k)\mathbf{1}_{\{k\}}$ einfacher Funktionen. Zusammen gibt das aufgrund der Definition des allgemeinen Lebesgue Integrals für einfache Integranden

$$\int_{\mathbb{N}} f \, d\mu \stackrel{(3.1.7)}{=} \lim_{n \rightarrow \infty} \int_{\mathbb{N}} f_n \, d\mu \stackrel{\text{Def.}}{=} \lim_{n \rightarrow \infty} \sum_{k=0}^n f(k) \mu(\{k\}) = \lim_{n \rightarrow \infty} \sum_{k=0}^n f(k) = \sum_{k=0}^{\infty} f(k).$$

Vielleicht habt ihr in der Analysis 2 schon über Nullmengen gesprochen. Dann habt ihr schon gelernt, dass Nullmengen bei Integralen keine Rolle spielen. Wenn nicht, lernt ihr Nullmengen und ihre Bedeutung für Integrale jetzt kennen:

Definition 3.1.13.  $N \in \mathcal{A}$ heißt μ -Nullmenge, falls $\mu(N) = 0$.

Aufgrund der Subadditivität von Maßen (folgt aus der σ -Additivität) folgt sofort, dass abzählbare Vereinigungen von Nullmengen wieder Nullmengen sind (kleine Übung).

Definition 3.1.14. 

- (i) Gilt eine Eigenschaft für alle $\omega \in \Omega$ außer einer μ -Nullmenge (d. h. für $N := \{\omega \in \Omega : \text{Eigenschaft gilt nicht}\}$ gilt $\mu(N) = 0$), so gilt die Eigenschaft μ -fast überall. Man schreibt kurz auch μ -f.ü.
- (ii) Ist μ ein Wahrscheinlichkeitsmaß, so sagt man anstelle von „ μ -fast überall“ auch „ μ -fast sicher“ oder kurz μ -f.s.

Wenn klar ist über welches Maß μ gesprochen wird, sagt man auch nur „fast überall“ oder „fast sicher“.

Vorlesung 11

Weil aufgrund der Definition des Integrals für Indikatorfunktionen $f = \mathbf{1}_N$ über Nullmengen $\int_{\Omega} f \, d\mu = 1 \cdot \mu(N) = 0$, ist es nicht überraschend, dass Nullmengen beim Integrieren keine Rolle spielen. Das wird in (ii) des folgenden sehr wichtigen Satzes deutlich:

Satz 3.1.15.  Sind $f, g: \Omega \rightarrow \overline{\mathbb{R}}$ μ -integrierbar, so gelten:

- (i) f ist μ -fast überall endlich.
- (ii) $f = g$ μ -fast überall impliziert $\int_{\Omega} f \, d\mu = \int_{\Omega} g \, d\mu$.
- (iii) $f \geq 0$ und $\int_{\Omega} f \, d\mu = 0$ impliziert $f = 0$ μ -fast überall.

Beweis.

- (i) Sei $A := \{|f| = \infty\} = f^{-1}(\{-\infty, +\infty\}) \in \mathcal{A}$. Weil $n\mathbf{1}_A \leq |f|$ für alle $n \in \mathbb{N}$ gilt, folgt

$$n\mu(A) \stackrel{\text{Def.}}{=} \int_{\Omega} n\mathbf{1}_A \, d\mu \stackrel{\text{Mon.}}{\leq} \int_{\Omega} |f| \, d\mu = \int_{\Omega} (f^+ + f^-) \, d\mu \stackrel{f}{\underset{\mu\text{-int.}}{<}} \infty$$

für alle $n \in \mathbb{N}$. Weil $\mu(A) > 0$ einen Widerspruch gibt (dann wäre $n\mu(A)$ unbeschränkt aber $\int |f| \, d\mu$ ist eine obere Schranke), folgt die Behauptung.

- (ii) Für $N := \{f \neq g\} \in \mathcal{A}$ gilt aufgrund der Voraussetzung $\mu(N) = 0$. Weil aus $f = g$ fast überall auch $f^+ = g^+$ und $f^- = g^-$ fast überall folgt, impliziert die Definition des Integrals als $\int_{\Omega} f \, d\mu = \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu$ bzw. $\int_{\Omega} g \, d\mu = \int_{\Omega} g^+ \, d\mu - \int_{\Omega} g^- \, d\mu$, dass die Aussage nur für $f, g \geq 0$ gezeigt werden muss (wende Aussage dann auf Positiv- und Negativteil an). Seien also $f, g \geq 0$ und

$$N = \{f \neq g\} = \{\omega : f(\omega) \neq g(\omega)\}$$

die Nullmenge, auf der f und g nicht übereinstimmen. Dann gilt aufgrund der Monotonie und der Definition des Integrals

$$0 \leq \int_{\Omega} f \mathbf{1}_N \, d\mu \leq \int_{\Omega} (+\infty) \mathbf{1}_N \, d\mu = +\infty \cdot \mu(N) = +\infty \cdot 0 = 0.$$

Für die letzte Gleichung haben wir die Konvention $+\infty \cdot 0 = 0$ genutzt. Genauso gilt $\int_{\Omega} g \mathbf{1}_N \, d\mu = 0$. Wenn wir jetzt $1 = \mathbf{1}_N + \mathbf{1}_{N^c}$ schreiben, ergibt sich mit der Linearität des Integrals

$$\int_{\Omega} f \, d\mu = \int_{\Omega} f \mathbf{1}_N \, d\mu + \int_{\Omega} f \mathbf{1}_{N^c} \, d\mu = \int_{\Omega} f \mathbf{1}_{N^c} \, d\mu = \int_{\Omega} g \mathbf{1}_{N^c} \, d\mu = \int_{\Omega} g \, d\mu.$$

(iii) Seien $A_n = \{f > \frac{1}{n}\} = \{\omega : f(\omega) > \frac{1}{n}\}$ für $n \in \mathbb{N}$. Damit ist mit der Monotonie des Integrals

$$0 \stackrel{\text{Ann.}}{=} \int_{\Omega} f \, d\mu \stackrel{\text{Mon.}}{\geq} \int_{\Omega} f \mathbf{1}_{A_n} \, d\mu \stackrel{\text{Mon.}}{\geq} \int_{\Omega} \frac{1}{n} \mathbf{1}_{A_n} \, d\mu \stackrel{\text{Def.}}{=} \frac{1}{n} \mu(A_n) \geq 0,$$

weil $\frac{1}{n} \mathbf{1}_{A_n} < f \mathbf{1}_{A_n} \leq f$. Also ist $\mu(A_n) = 0$ für alle $n \in \mathbb{N}$. Damit gilt

$$0 \leq \mu(\{\omega : f(\omega) > 0\}) = \mu\left(\bigcup_{k=1}^{\infty} A_k\right) \stackrel{\text{subadd.}}{\leq} \sum_{k=1}^{\infty} \mu(A_k) = 0.$$

Es gilt also $f = 0$ μ -fast überall.

□

Für spätere Verwendungen noch ein Satz zur Transformation von Integralen. In Analysis 2 habt ihr den schon in konkreter Form im \mathbb{R}^d kennengelernt.

Satz 3.1.16.  [abstrakter Transformationsatz]

Seien (Ω, \mathcal{A}) , (Ω', \mathcal{A}') messbare Räume, μ ein Maß auf \mathcal{A} , $f: \Omega \rightarrow \Omega'$ messbar, $g: \Omega' \rightarrow \overline{\mathbb{R}}$ messbar und $g \geq 0$. Dann gilt

$$\int_{\Omega'} g \, d\mu_f = \int_{\Omega} g \circ f \, d\mu,$$

wobei $+\infty = +\infty$ möglich ist. Dabei ist μ_f der push-forward (Bildmaß) von μ .

$$\begin{array}{ccc} (\Omega, \mathcal{A}, \mu) & \xrightarrow{f} & (\Omega', \mathcal{A}', \mu_f) \\ & \searrow^{g \circ f} & \downarrow g \\ \int_{\Omega} g \circ f \, d\mu & & \int_{\Omega'} g \, d\mu_f \\ & & (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})) \end{array}$$

Beweis. Einmal durch die „Gebetsmühle“ der Integrationstheorie (d. h. zeige die Aussage für einfache Integranden und nehme dann den Grenzwert):

(A) Sei erstmal

$$g = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k} \geq 0$$

eine nicht-negative einfache Funktion. Wir nehmen die Darstellung mit $A_k = g^{-1}(\{\alpha_k\})$. Dann gilt

$$g \circ f = \sum_{k=1}^n \alpha_k \mathbf{1}_{f^{-1}(A_k)} \geq 0,$$

$g \circ f$ ist also auch eine einfache Funktion. Weil nach Definition des push-forwards $\mu_f(A_k) = \mu(f^{-1}(A_k))$ gilt, bekommen wir

$$\int_{\Omega'} g \circ f \, d\mu \stackrel{\text{Def.}}{=} \sum_{k=1}^n \alpha_k \mu(f^{-1}(A_k)) = \sum_{k=1}^n \alpha_k \mu_f(A_k) \stackrel{\text{Def.}}{=} \int_{\Omega'} g \, d\mu_f.$$

Damit gilt die Behauptung für einfache Funktionen.

(B) Weil g messbar ist, existiert eine Folge $(g_n) \subseteq \mathcal{E}^+$ mit $g_n \uparrow g$, $n \rightarrow \infty$. Also gilt

$$\int_{\Omega'} g \, d\mu_f \stackrel{3.1.7}{=} \lim_{n \rightarrow \infty} \int_{\Omega'} g_n \, d\mu_f \stackrel{(A)}{=} \lim_{n \rightarrow \infty} \int_{\Omega} g_n \circ f \, d\mu \stackrel{3.1.7}{=} \int_{\Omega} g \circ f \, d\mu.$$

Im letzten Schritt haben wir genutzt, dass auch $g_n \circ f$ einfach ist (siehe (A)) und $g_n \circ f \uparrow g \circ f$, $n \rightarrow \infty$, gilt.

□

Wir hätten den letzten Satz auch direkt ohne die Nichtnegativität formulieren können. Aus didaktischen Gründen zerlegen wir die Aussage in den Satz und das folgende Korollar. Der nicht-negative Fall ist einfacher zu formulieren weil Integrale über nicht-negative Funktionen immer definiert sind. Bei allgemeinen Integranden muss man immer aufpassen, dass nicht $+\infty + (-\infty)$ auftaucht, das haben wir Wohldefiniertheit genannt. Die Formulierung des Satzes muss also die Wohldefiniertheit beinhalten. Um sich das Leben leichter zu machen, nimmt man meistens sogar die Integrierbarkeit (Wohldefiniert und endlich) an, das reicht in den Anwendungen ohnehin meistens aus.

Korollar 3.1.17. ▶ Unter den Voraussetzungen von Satz 3.1.16 gelte jetzt nur noch $g: \Omega' \rightarrow \overline{\mathbb{R}}$. Dann ist g μ_f -integrierbar genau dann, wenn $g \circ f$ μ -integrierbar ist. Ist eine dieser Aussagen erfüllt, so gilt ebenfalls die Transformationsformel

$$\int_{\Omega'} g \, d\mu_f = \int_{\Omega} g \circ f \, d\mu.$$

Beweis. Wegen $g^+ \circ f = (g \circ f)^+$ und $g^- \circ f = (g \circ f)^-$ folgt aus dem Transformationssatz für nicht-negative Funktionen

$$\begin{aligned} g \text{ } \mu_f\text{-integrierbar} &\stackrel{\text{Def.}}{\Leftrightarrow} \int_{\Omega'} g^+ \, d\mu_f < \infty \quad \text{und} \quad \int_{\Omega'} g^- \, d\mu_f < \infty \\ &\Leftrightarrow \int_{\Omega} g^+ \circ f \, d\mu < \infty \quad \text{und} \quad \int_{\Omega} g^- \circ f \, d\mu < \infty \\ &\stackrel{\text{Def.}}{\Leftrightarrow} g \circ f \text{ } \mu\text{-integrierbar.} \end{aligned}$$

Nun zur Berechnung der Integrale:

$$\begin{aligned} \int_{\Omega'} g \, d\mu_f &\stackrel{\text{Def.}}{=} \int_{\Omega'} g^+ \, d\mu_f - \int_{\Omega'} g^- \, d\mu_f \\ &\stackrel{3.1.16}{=} \int_{\Omega} g^+ \circ f \, d\mu - \int_{\Omega} g^- \circ f \, d\mu \\ &= \int_{\Omega} (g \circ f)^+ \, d\mu - \int_{\Omega} (g \circ f)^- \, d\mu \\ &\stackrel{\text{Def.}}{=} \int_{\Omega} g \circ f \, d\mu. \end{aligned}$$

□

3.2 Konvergenzsätze

Im folgenden sei $(\Omega, \mathcal{A}, \mu)$ ein fester Maßraum. Gezeigt haben wir schon

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu,$$

wenn $(f_n)_{n \in \mathbb{N}}$ eine Folge nicht-negativer einfacher Funktionen ist, die wachsend gegen f konvergieren. Wir wollen nun die gleiche Aussage für beliebige nicht-negative wachsende Folgen zeigen.

Satz 3.2.1. ▶ [Monotone Konvergenz Theorem (MCT)]

Seien $f, f_1, f_2, \dots: \Omega \rightarrow \overline{\mathbb{R}}$ messbar und es gelte $0 \leq f_1 \leq f_2 \leq \dots \leq f$ sowie $f = \lim_{n \rightarrow \infty} f_n$ μ -f.ü.

Dann gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu,$$

wobei $+\infty = +\infty$ möglich ist.

Für monoton wachsende Funktionenfolgen darf der Limes also in das Integral getauscht werden.

Beweis.

- (i) Wir nehmen an, dass $f_1(\omega) \leq f_2(\omega) \leq \dots \leq f(\omega)$ und $f(\omega) = \lim_{n \rightarrow \infty} f_n(\omega)$ nicht nur fast überall, sondern für alle $\omega \in \Omega$ (also punktweise) gelten.

„ \leq “: Wegen der Monotonie des Integrals gilt

$$\int_{\Omega} f_n \, d\mu \leq \int_{\Omega} f \, d\mu$$

für alle $n \in \mathbb{N}$. Weil die Folge der Integrale aufgrund der Monotonie wachsend ist, existiert nach Analysis 1 der Grenzwert ($+\infty$ ist möglich). Wieder nach Analysis 1, Vergleichssatz für Folgen, gilt damit

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} f \, d\mu.$$

„ \geq “: Weil alle f_n messbar sind, existieren Folgen $(g_{n,k}) \subseteq \mathcal{E}^+$ mit $g_{n,k} \uparrow f_n$, $k \rightarrow \infty$. Sei $h_n = g_{1,n} \vee \dots \vee g_{n,n} = \max\{g_{1,n}, \dots, g_{n,n}\}$. Die h_n sind einfache Funktionen, für die zwei Eigenschaften gelten:

- (a) $h_n \leq f_n$
 (b) $h_n \uparrow f$

Zu (a): Es gilt $g_{i,n} \leq f_i \leq f_n$ für alle $i \leq n$, also ist auch das punktweise Maximum kleiner als f_n . Zu (b): Weil $h_n \geq g_{i,n}$ für alle $i = 1, \dots, n$ gilt, ist auch

$$\lim_{n \rightarrow \infty} h_n \geq \lim_{n \rightarrow \infty} g_{i,n} = f_i$$

für alle festen $i \in \mathbb{N}$. Weil aber $\lim_{i \rightarrow \infty} f_i = f$ gilt, ist $\lim_{n \rightarrow \infty} h_n \geq f$, erneut nach dem Vergleichssatz für Folgen aus Analysis 1. Weil auch noch $h_n \leq f_n \leq f$ gilt, folgt mit der letzten Aussage $\lim_{n \rightarrow \infty} h_n = f$ punktweise. Die Folge (h_n) ist also eine wachsende Folge von einfachen Funktionen, so dass (f_n) zwischen (h_n) und f liegt und (h_n) punktweise gegen f konvergiert. Damit bekommen wir aus dem Monotone Konvergenz Theorem für einfache Funktionen durch Einschachtelung die Behauptung:

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \stackrel{(a)}{\geq} \lim_{n \rightarrow \infty} \int_{\Omega} h_n \, d\mu \stackrel{3.1.7}{=} \int_{\Omega} \lim_{n \rightarrow \infty} h_n \, d\mu \stackrel{(b)}{=} \lim_{n \rightarrow \infty} \int_{\Omega} f \, d\mu.$$

- (ii) Sei N die Nullmenge, auf der die Annahme aus (i) nicht gilt, also $N = \{\omega \in \Omega : f_n(\omega) \not\rightarrow f(\omega)\}$. Es gilt $f_n \mathbf{1}_{N^c} \rightarrow f \mathbf{1}_{N^c}$, $n \rightarrow \infty$, punktweise, weil für alle $\omega \in N$ die Folge auf konstant 0 gesetzt wird. Wegen (i) gilt dann

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \stackrel{3.1.15}{=} \lim_{n \rightarrow \infty} \int_{\Omega} f_n \mathbf{1}_{N^c} \, d\mu \stackrel{(i)}{=} \int_{\Omega} f \mathbf{1}_{N^c} \, d\mu \stackrel{3.1.15}{=} \int_{\Omega} f \, d\mu,$$

weil Integrale zweier Funktionen gleich sind, wenn sie nur auf Nullmengen unterschiedlich sind. □

Kommen wir zu einer Anwendung, die für die Stochastik sehr wichtig ist. Gerade in der Finanzmathematik wird folgender „Maßwechsel“ essentiell sein!

Anwendung 3.2.2. ▶ Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $f: \Omega \rightarrow \overline{\mathbb{R}}$ messbar und nicht-negativ und μ ein Maß auf \mathcal{A} . Dann ist

$$\nu(A) := \int_A f \, d\mu = \int_{\Omega} f \mathbf{1}_A \, d\mu$$

ein Maß auf \mathcal{A} .

Beweis.

$\nu \geq 0$ ✓ wegen Integral über nicht-negative Funktion.

$\nu(\emptyset) = 0$ ✓ weil Integral über die Nullfunktion 0 ist.

σ -Additivität: Seien $A_1, A_2, \dots \in \mathcal{A}$ paarweise disjunkt. Dann gilt

$$\begin{aligned} \nu\left(\bigcup_{k=1}^{\infty} A_k\right) &\stackrel{\text{Def.}}{=} \int_{\Omega} f \mathbf{1}_{\bigcup_{k=1}^{\infty} A_k} \, d\mu \\ &= \int_{\Omega} f \cdot \left(\sum_{k=1}^{\infty} \mathbf{1}_{A_k}\right) \, d\mu \\ &\stackrel{\text{Def. Reihe}}{=} \int_{\Omega} \lim_{n \rightarrow \infty} \sum_{k=1}^n f \mathbf{1}_{A_k} \, d\mu \\ &\stackrel{3.2.1}{=} \lim_{n \rightarrow \infty} \int_{\Omega} \sum_{k=1}^n f \mathbf{1}_{A_k} \, d\mu \\ &\stackrel{\text{Lin.}}{=} \lim_{n \rightarrow \infty} \sum_{k=1}^n \int_{\Omega} f \mathbf{1}_{A_k} \, d\mu \\ &\stackrel{\text{Def.}}{=} \lim_{n \rightarrow \infty} \sum_{k=1}^n \nu(A_k) \stackrel{\text{Def. Reihe}}{=} \sum_{k=1}^{\infty} \nu(A_k). \end{aligned}$$

Weil hier die Folge $\left(\sum_{k=1}^n f \mathbf{1}_{A_k}\right)_{n \in \mathbb{N}} \not\subseteq \mathcal{E}^+$, reicht die einfache Version der monotonen Konvergenz aus 3.1.7 nicht aus, wir brauchen monotone Konvergenz wirklich für allgemeine messbare Funktionen. □

Anschließend an den Maßwechsel noch eine Bemerkung, die für diese Stochastik 1 Vorlesung nicht wichtig ist. Da das Thema zum Beispiel in der Finanzmathematik extrem relevant ist, schreiben wir die Notation zum Gewöhnen schon einmal auf:

Bemerkung 3.2.3. ▶

- (i) Man schreibt mit
- ν
- aus vorheriger Anwendung auch

$$\frac{d\nu}{d\mu} = f$$

und nennt f die **Radon-Nikodým-Ableitung** oder **-Dichte von ν bezüglich μ** . Man sagt auch, dass ν absolutstetig bezüglich μ ist. ν ist ein Wahrscheinlichkeitsmaß (d. h. $\nu(\Omega) = 1$) genau dann, wenn $\int_{\Omega} f d\mu = 1$. Das kennen wir ja schon!

- (ii) Wir kennen den Begriff der Absolutstetigkeit schon für Verteilungsfunktionen (siehe Definition 1.4.8). In der Tat passt das genau zu dem neuen Begriff der Absolutstetigkeit für Maße: Ist
- \mathbb{P}_F
- ein Wahrscheinlichkeitsmaß auf
- $\mathcal{B}(\mathbb{R})$
- mit Verteilungsfunktion
- F
- und Dichte
- f
- , so gilt aufgrund der Definition von
- \mathbb{P}_F
- (Maße sind auf einem
- \cap
- stabilen Erzeuger eindeutig festgelegt!), dass

$$\frac{d\mathbb{P}_F}{d\lambda} = f, \quad \lambda = \text{Lebesguemaß.}$$

Ist also F absolutstetig mit Dichte f , so ist das Maß \mathbb{P}_F absolutstetig bezüglich dem Lebesguemaß mit Dichte f . Die zwei Begriffe der Absolutstetigkeit für Verteilungsfunktionen und Maße passen damit zusammen.

Ganz analog funktioniert das für diskrete Verteilungen. Ist F diskret mit Sprungstellen $(a_k)_{k=1, \dots, N}$ und Sprunghöhen $(p_k)_{k=1, \dots, N}$, so gilt

$$\frac{d\mathbb{P}_F}{d\mu} = \sum_{k=1}^N p_k \mathbf{1}_{\{a_k\}}, \quad \mu = \sum_{k=1}^N \delta_{a_k}.$$

Das ist der Grund dafür, weshalb bei diskreten Verteilungen die Wahrscheinlichkeiten $(p_k)_{k=1, \dots, N}$ manchmal auch Zähldichte genannt werden. Die Formel ist natürlich furchtbar, sie hat aber eine einfache Bedeutung: Das Maß μ wird durch die Dichte „umgewichtet“. Die Masse liegt auf den gleichen Werten a_1, \dots, a_N , \mathbb{P}_F hat aber eine andere Verteilung der Masse auf a_1, \dots, a_N : Auf den Werten a_k liegt nun nicht Masse 1, sondern Masse p_k .

Vorlesung 12

Satz 3.2.4.  **[Lemma von Fatou]**

Seien $f_1, f_2, \dots: \Omega \rightarrow \overline{\mathbb{R}}$ messbar und nicht-negativ, die Folge (f_n) muss dabei nicht konvergieren. Dann gilt

$$\int_{\Omega} \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n d\mu,$$

$+\infty \leq +\infty$ ist dabei möglich.

Wenn (f_n) sogar μ -f.ü. konvergiert und $(\int_{\Omega} f_n d\mu)$ konvergiert, gilt damit

$$\int_{\Omega} \lim_{n \rightarrow \infty} f_n d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu$$

weil für konvergente Folgen $\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} a_n$ gilt. Die Ungleichung „ \leq “ gilt in den Konvergenzsätzen also mit weniger Annahmen als Satz 3.2.1 (und anschliessend in Satz 3.2.5).

Beweis. Definieren wir $g_n := \inf_{k \geq n} f_k$, so ist g_n messbar für alle $n \in \mathbb{N}$, wachsend in n und erfüllt $g_n \leq f_n$, $n \in \mathbb{N}$, sowie punktweise $\lim_{k \rightarrow \infty} g_k = \liminf_{n \rightarrow \infty} f_n$ (das ist eine der äquivalenten Definitionen des Limes Inferiors). Satz 3.2.1 und Monotonie von Integralen und Folggrenzwerten gibt dann die Aussage:

$$\int_{\Omega} \liminf_{n \rightarrow \infty} f_n d\mu = \int_{\Omega} \lim_{n \rightarrow \infty} g_n d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} g_n d\mu = \liminf_{n \rightarrow \infty} \int_{\Omega} g_n d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n d\mu.$$

Das dritte Gleichheitszeichen gilt weil g_n (und damit das Integral) wachsend in n ist. \square

Satz 3.2.5.  [Dominierte Konvergenz Theorem (DCT)]

Seien $f, f_1, f_2, \dots: \Omega \rightarrow \overline{\mathbb{R}}$ messbar. Es sollen gelten

- (a) $\lim_{n \rightarrow \infty} f_n = f$ μ -fast überall,
- (b) $|f_n| \leq g$ μ -fast überall für alle $n \in \mathbb{N}$, für eine beliebige μ -integrierbare nicht-negative messbare numerische Funktion g .

Dann sind f, f_1, f_2, \dots μ -integrierbar und

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu.$$

Die Funktion g spielt keine große Rolle (sie muss nur existieren) und wird integrierbare Majorante für die Folge (f_n) genannt.

Beweis. Wie beim Beweis der monotonen Konvergenz nehmen wir zunächst an, dass die Konvergenz sogar für alle $\omega \in \Omega$ gilt. In einem zweiten Schritt kann man dann wie bei monotoner Konvergenz mit der Hilfsfolge $(f_n \mathbf{1}_{N^c})$ für die Nullmenge $N = \{\omega : f_n(\omega) \not\rightarrow f(\omega)\}$ den Fall der μ -f.ü. Konvergenz zeigen.

Der Beweis beruht auf einer elementaren Erkenntnis: Wenn $|f_n| \leq g$ gilt, so gilt $f_n \leq g$ und $-f_n \leq g$ oder umgeformt auch $0 \leq f_n + g$ und $0 \leq g - f_n$. In anderen Worten: Wir können die f_n so geschickt verschieben, dass wir nicht-negative Funktionen bekommen und Fatou anwenden können.

(i)

$$\begin{aligned} \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu &\stackrel{\text{Lin.}}{=} \int_{\Omega} (f + g) \, d\mu \\ &\stackrel{\text{Ann.}}{=} \int_{\Omega} \left(\lim_{n \rightarrow \infty} f_n + g \right) \, d\mu \\ &= \int_{\Omega} \left(\liminf_{n \rightarrow \infty} f_n + g \right) \, d\mu \\ &\stackrel{3.2.4}{\leq} \liminf_{n \rightarrow \infty} \int_{\Omega} (f_n + g) \, d\mu \\ &\stackrel{\text{Lin.}}{=} \liminf_{n \rightarrow \infty} \left(\int_{\Omega} f_n \, d\mu + \int_{\Omega} g \, d\mu \right) \\ &= \liminf_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu + \int_{\Omega} g \, d\mu. \end{aligned}$$

Wenn wir nun auf beiden Seiten das Integral über g abziehen, dann bekommen wir

$$\int_{\Omega} f \, d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu.$$

(ii) Das selbe Argument wenden wir auf $0 \leq g - f_n$ an. Dieselbe Rechnung gibt

$$\int_{\Omega} -f \, d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} -f_n \, d\mu \stackrel{\text{Lin.}}{=} \liminf_{n \rightarrow \infty} - \int_{\Omega} f_n \, d\mu = - \limsup_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu,$$

also

$$\limsup_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} f \, d\mu.$$

Beide Schritte zusammen ergeben

$$\int_{\Omega} f \, d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \leq \limsup_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu \leq \int_{\Omega} f \, d\mu.$$

Also stimmen \liminf und \limsup überein und geben nach Analysis 1 den Grenzwert

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu.$$

□

Beispiel. In beiden Beispielen sei $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.

- (i) Schauen wir uns mal ein Gegenbeispiel für die Konvergenzsätze an, dafür nehmen wir die Folge $f_n = n \mathbf{1}_{(0, \frac{1}{n})}$. Das sind Indikatorfunktionen, deren Breite kleiner wird, die Höhe allerdings größer wird. Es gilt punktweise $\lim_{n \rightarrow \infty} f_n = f$, wobei f die Nullfunktion ist. Wegen

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} f_n \, d\lambda \stackrel{\text{Höhe mal Breite}}{=} \lim_{n \rightarrow \infty} 1 = 1 \neq 0 = \int_{\mathbb{R}} f \, d\lambda,$$

scheinen MCT und DCT nicht zu funktionieren. Warum? Die Annahme von MCT ist nicht erfüllt weil die Folge f_n nicht punktweise wächst. Die Folge ist auch nicht beschränkt durch „die selbe“ integrierbare Funktion g , daher ist die Annahme von DCT nicht erfüllt.

- (ii) Für reelle Integrale können wir jetzt einmal schnell den Berechnungsweg

$$\int_{\mathbb{R}} f \, d\lambda = \lim_{n \rightarrow \infty} \int_{[-n, n]} f \, d\lambda$$

für $f \geq 0$ begründen. Das folgt direkt aus MCT mit der Wahl $f_n = f \mathbf{1}_{[-n, n]}$ (f wird also außerhalb von $[-n, n]$ auf 0 gesetzt) weil $f_n \uparrow f$ (das gilt nur für $f \geq 0$!). In dx -Notation steht da also, wie in Beispiel 3.1.12 (b) behauptet,

$$\int_{\mathbb{R}} f(x) \, dx = \lim_{n \rightarrow \infty} \int_{-n}^n f(x) \, dx.$$

Warnung: Gegenbeispiele wie $\frac{\sin(x)}{x}$ aus Beispiel 3.1.12 (c) zeigen, dass das Integral über ganz \mathbb{R} nicht unbedingt der Grenzwert der Integrale von $-n$ bis n sein muss! Weil die Folge (f_n) in dem Fall nicht wachsend ist, kann in dem Fall MCT auch nicht angewandt werden.

Eine ganz wichtige Folgerung für die spätere Stochastik ist der Spezialfall, wenn μ ein endliches Maß (insbesondere ein Wahrscheinlichkeitsmaß) ist:

Korollar 3.2.6. ▶ Ist μ ein endliches Maß (z. B. ein Wahrscheinlichkeitsmaß) und $|f_n| \leq C$ für alle $n \in \mathbb{N}$ μ -f.ü., d. h. alle f_n sind *beschränkt* durch das gleiche C , und $\lim_{n \rightarrow \infty} f_n = f$ μ -f.ü., so gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \int_{\Omega} f \, d\mu.$$

Beweis. Das ist dominierte Konvergenz mit der Majorante $g \equiv C$. Als Indikatorfunktion ist die Majorante integrierbar, weil

$$\int_{\Omega} g \, d\mu = \int_{\Omega} C \mathbf{1}_{\Omega} \, d\mu \stackrel{\text{Def.}}{=} C \mu(\Omega) < \infty.$$

□

3.3 Das Beispiel - Integrale über Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$

Weil in dem Spezialfall $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_F)$ die Integrale später in der Stochastik unter dem Namen „Erwartungswerte“ eine enorm wichtige Rolle spielen werden, schauen wir uns den Spezialfall jetzt schon mal in Ruhe an. Das gibt euch genug Zeit, die wichtigsten Rechnungen über das Semester mehrfach zu üben.

Kurze Erinnerung: Ein Wahrscheinlichkeitsmaß \mathbb{P}_F auf $\mathcal{B}(\mathbb{R})$ mit Verteilungsfunktion F beschreibt ein reellwertiges Zufallsexperiment, bei dem die „gezogene Zahl“ mit Wahrscheinlichkeit $\mathbb{P}_F((a, b]) = F(b) - F(a)$ in $(a, b]$ liegt. Hat F eine Dichte f , so gilt

$$F(t) = \int_{-\infty}^t f(x) dx, \quad \text{also } \mathbb{P}_F((a, b]) = \int_a^b f(x) dx.$$

Ist F diskret mit Werten a_1, \dots, a_N und Wahrscheinlichkeiten p_1, \dots, p_N , so gilt

$$F(t) = \sum_{k=1}^N p_k \mathbf{1}_{[a_k, +\infty)} = \sum_{a_k \leq t} p_k, \quad \text{also } \mathbb{P}_F((a, b]) = \sum_{a < a_k \leq b} p_k.$$

Wir summieren also die Wahrscheinlichkeiten der Werte in $(a, b]$.

Ein paar konkreten Integralen geben wir jetzt Namen und überlegen uns anschließend, wie man die Integrale in vielen Beispielen berechnen kann.

Definition 3.3.1.

(i) Für $k \in \mathbb{N}$ heißt

$$\int_{\mathbb{R}} x^k d\mathbb{P}_F(x)$$

k-tes Moment von \mathbb{P}_F oder k-tes Moment der Verteilungsfunktion F , wenn das Integral wohldefiniert ist.

(ii)

$$\int_{\mathbb{R}} e^{\lambda x} d\mathbb{P}_F(x)$$

heißt **exponentielles Moment** von \mathbb{P}_F oder exponentielles Moment der Verteilungsfunktion F .

Allgemein betrachten wir für messbare Abbildungen $g: \mathbb{R} \rightarrow \overline{\mathbb{R}}$ auch die Integrale

$$\int_{\mathbb{R}} g(x) d\mathbb{P}_F(x),$$

jedoch ohne ihnen extra einen Namen zu geben.

Beachte: Alle Integrale über nicht-negative messbare Integranden sind wohldefiniert, das Integral könnte aber $+\infty$ sein. Damit sind exponentielle Momente und gerade Momente immer in $[0, +\infty]$ definiert, existieren nach unserer Konvention aber nur, wenn sie endlich sind.

Satz 3.3.2. [Integrale für absolutstetige Verteilungen]

Sei \mathbb{P}_F ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R})$ und F habe Dichte f . Dann gilt für $g: \mathbb{R} \rightarrow \overline{\mathbb{R}}$ Borel-messbar:

$$\int_{\mathbb{R}} g d\mathbb{P}_F \text{ ist wohldefiniert} \quad \Leftrightarrow \quad \int_{\mathbb{R}} g(x) f(x) dx \text{ ist wohldefiniert} \quad (3.5)$$

und, wenn die Integrale wohldefiniert sind, gilt die Rechenregel

$$\int_{\mathbb{R}} g d\mathbb{P}_F = \int_{\mathbb{R}} g(x) f(x) dx.$$

Der Satz besagt, dass wir die abstrakten Integrale $\int_{\mathbb{R}} g d\mathbb{P}_F$ durch sehr viel weniger abstrakte Integrale behandeln können. Beachtet, dass mit $\int_{\mathbb{R}} g(x)f(x) dx$ das Lebesgue Integral $\int_{\mathbb{R}} gf d\lambda$ gemeint ist. Nach dem Beweis diskutieren wir nochmal ausführlich, wie ihr das mit den Tricks aus der Analysis berechnen könnt.

Beweis.

(i) Für $g \geq 0$ starten wir die Gebetsmühle der Integrationstheorie:

- Sei zunächst $g = \mathbf{1}_A$ für $A \in \mathcal{B}(\mathbb{R})$. Es gilt

$$\int_{\mathbb{R}} g d\mathbb{P}_F \stackrel{\text{Def.}}{=} \mathbf{1} \cdot \mathbb{P}_F(A) \stackrel{(*)}{=} \int_A f(x) dx = \int_{\mathbb{R}} \mathbf{1}_A(x)f(x) dx = \int_{\mathbb{R}} g(x)f(x) dx.$$

Warum gilt (*), also $\mathbb{P}_F(A) = \int_A f(x) dx$? Ist $A = (a, b]$, so gilt $\mathbb{P}_F((a, b]) = \int_a^b f(x) dx$ weil F Dichte f hat. In Anwendung 3.2.2 haben wir gezeigt, dass $\nu(A) = \int_A f(x) dx$ ein Maß auf $\mathcal{B}(\mathbb{R})$ ist. Es gilt also $\mathbb{P}_F = \nu$ auf einem \cap -stabilen Erzeuger von $\mathcal{B}(\mathbb{R})$ und damit aufgrund von Korollar 1.2.12 auch auf $\mathcal{B}(\mathbb{R})$. Also gilt (*) für alle $A \in \mathcal{B}(A)$.

- Ist g eine einfache Funktion, so folgt nun

$$\int_{\mathbb{R}} g d\mathbb{P}_F = \int_{\mathbb{R}} g(x)f(x) dx$$

aufgrund der Linearität des Integrals.

- Ist $g \geq 0$, wählen wir eine Folge $(g_n) \subseteq \mathcal{E}^+$ mit $g_n \uparrow g$, $n \rightarrow \infty$. Die Folge existiert weil g messbar ist. Mit dem Monotone Konvergenz Theorem und dem gerade Gezeigten für einfache Funktionen folgt

$$\int_{\mathbb{R}} g d\mathbb{P}_F \stackrel{3.2.1}{=} \lim_{n \rightarrow \infty} \int_{\mathbb{R}} g_n d\mathbb{P}_F = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} g_n(x)f(x) dx \stackrel{3.2.1}{=} \int_{\mathbb{R}} g(x)f(x) dx.$$

(ii) Für g beliebig zerlegt man g in $g^+ - g^-$ und wende (i) auf die Integrale der Positiv- und Negativteile an. Vergleiche den Anfang des Beweises von Korollar 3.1.17.

□

Weil ihr mit dem Satz in dieser Vorlesung viel rechnen werdet, verbinden wir den Satz noch schnell mit Beispiel 3.1.12. Das Integral auf der rechten Seite der Äquivalenzen ist das Lebesgue Integral $\int_{\mathbb{R}} gf d\lambda$ in der weniger angstverbreitenden dx -Notation. Für nette Integranden (stückweise stetig) ist das Integral aufgrund der Diskussion aus Beispiel 3.1.12 wie folgt zu berechnen ist:

- Ist $g \geq 0$ (z. B. $g(x) = e^x$), so ist alles einfach: Weil f als Dichte immer nicht-negativ ist, ist auch das Produkt gf nicht-negativ. Damit ist das Integral immer wohldefiniert (es könnte aber unendlich sein). Daher könnt ihr den ersten Teil im Satz ignorieren und sofort

$$\int_{\mathbb{R}} g d\mathbb{P}_F = \int_{\mathbb{R}} g(x)f(x) dx \stackrel{\text{MCT}}{=} \lim_{n \rightarrow \infty} \int_{-n}^n g(x)f(x) dx$$

mit den Tricks der Analysis ausrechnen.

- Habt ihr Pech, so gilt $g \geq 0$ nicht (z. B. für $g(x) = x$). Dann überlegt ihr euch, was Positivteil g^+ ist und was Negativteil g^- ist. Weil beide nicht-negativ sind, macht ihr 2x das gerade beschriebene, ihr berechnet also $\int_{\mathbb{R}} g^+(x)f(x) dx$ und $\int_{\mathbb{R}} g^-(x)f(x) dx$. Wenn eines von beiden (oder beide) endlich ist, so ist nach dem Satz $\int_{\mathbb{R}} g d\mathbb{P}_F$ wohldefiniert und ihr habt den Wert

$$\int_{\mathbb{R}} g d\mathbb{P}_F = \int_{\mathbb{R}} g^+(x)f(x) dx - \int_{\mathbb{R}} g^-(x)f(x) dx$$

schon ausgerechnet. Wir haben dabei $(gf)^+ = g^+f$ und $(gf)^- = g^-f$ benutzt, weil f als Dichte nicht-negativ ist. Vorsicht: In diesem zweiten Fall dürft ihr nicht einfach $\int_{\mathbb{R}} g d\mathbb{P}_F = \lim_{n \rightarrow \infty} \int_{-n}^n g(x)f(x) dx$ nutzen! Nach Beispiel 3.1.12 (iii) kann das falsch sein! Wenn ihr neugierig seid, könnt ihr für ein konkretes Beispiel zu Warnung 3.3.7 vorblättern.

Angelehnt an (3.3) können wir auch folgendes formulieren, wenn wir uns nur für die Endlichkeit des Integrals interessieren:

$$g \text{ ist } \mathbb{P}_F\text{-integrierbar} \iff \int_{\mathbb{R}} g \, d\mathbb{P}_F \text{ existiert} \iff \int_{\mathbb{R}} |g(x)|f(x) \, dx < \infty. \quad (3.6)$$

Vorlesung 13

Im diskreten Fall nehmen wir mal diese Formulierung:

Satz 3.3.3.  [Integrale für diskrete Verteilungen]

Sei \mathbb{P}_F ein Maß auf $\mathcal{B}(\mathbb{R})$ und F sei diskret, d. h.

$$F(t) = \sum_{k=1}^N p_k \mathbf{1}_{[a_k, \infty)}(t), \quad t \in \mathbb{R},$$

mit $N \in \mathbb{N} \cup \{+\infty\}$, $a_1, \dots, a_N \in \mathbb{R}$ und $\sum_{k=1}^N p_k = 1$. Dann gilt für $g: \mathbb{R} \rightarrow \overline{\mathbb{R}}$ messbar:

$$g \text{ } \mathbb{P}_F\text{-integrierbar} \iff \sum_{k=1}^N |g(a_k)|p_k < \infty$$

und, wenn g \mathbb{P}_F -integrierbar ist, gilt

$$\int_{\mathbb{R}} g \, d\mathbb{P}_F = \sum_{k=1}^N g(a_k)p_k.$$

Zu beachten ist, dass in vielen diskreten Modellen N endlich ist und g die Werte $+\infty$ und $-\infty$ nicht annimmt, dann ist das Integral $\int_{\mathbb{R}} g \, d\mathbb{P}_F$ natürlich immer definiert und ihr merkt euch einfach die Rechenregel $\int_{\mathbb{R}} g \, d\mathbb{P}_F = \sum_{k=1}^N g(a_k)p_k$.

Beweis.

(i) Sei zunächst $g \geq 0$:

- (a) Am einfachsten ist der Fall $N \in \mathbb{N}$, denn dann sprechen wir nur von endlichen Summen. Weil das Maß \mathbb{P}_F von der Form $\mathbb{P}_F = \sum_{k=1}^N p_k \delta_{a_k}$ ist, gilt $g = g \mathbf{1}_{\{a_1, \dots, a_N\}}$ \mathbb{P}_F -fast überall. Es folgt dann

$$\begin{aligned} \int_{\mathbb{R}} g \, d\mathbb{P}_F &\stackrel{\text{Satz 3.1.15}}{=} \int_{\mathbb{R}} g \mathbf{1}_{\{a_1, \dots, a_N\}} \, d\mathbb{P}_F \\ &= \int_{\mathbb{R}} g \sum_{k=1}^N \mathbf{1}_{\{a_k\}} \, d\mathbb{P}_F \\ &= \int_{\mathbb{R}} \sum_{k=1}^N g(a_k) \mathbf{1}_{\{a_k\}} \, d\mathbb{P}_F \\ &\stackrel{\text{Lin.}}{=} \sum_{k=1}^N \int_{\mathbb{R}} \underbrace{g(a_k) \mathbf{1}_{\{a_k\}}}_{\text{einfach}} \, d\mathbb{P}_F \\ &\stackrel{\text{Def. Int.}}{=} \sum_{k=1}^N g(a_k) \mathbb{P}_F(\{a_k\}) \stackrel{\text{Def. Maß}}{=} \sum_{k=1}^N g(a_k)p_k. \end{aligned}$$

- (b) $N = +\infty$ funktioniert im Prinzip genauso, wir müssen nur einmal monotone Konvergenz für die wachsende Folge von messbaren Funktionen $g_n := \sum_{k=1}^n g(a_k) \mathbf{1}_{\{a_k\}}$

nutzen, um Integral und Summe zu vertauschen:

$$\begin{aligned}
 \int_{\mathbb{R}} g \, d\mathbb{P}_F &= \int_{\mathbb{R}} g \mathbf{1}_{\{a_1, a_2, \dots\}} \, d\mathbb{P}_F \\
 &= \int_{\mathbb{R}} g \sum_{k=1}^{\infty} \mathbf{1}_{\{a_k\}} \, d\mathbb{P}_F \\
 &\stackrel{\text{Def. Reihe}}{=} \int_{\mathbb{R}} \lim_{n \rightarrow \infty} \sum_{k=1}^n g(a_k) \mathbf{1}_{\{a_k\}} \, d\mathbb{P}_F \\
 &\stackrel{3.2.1}{=} \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \sum_{k=1}^n g(a_k) \mathbf{1}_{\{a_k\}} \, d\mathbb{P}_F \\
 &\stackrel{\text{Lin.}}{=} \lim_{n \rightarrow \infty} \sum_{k=1}^n \int_{\mathbb{R}} \underbrace{g(a_k) \mathbf{1}_{\{a_k\}}}_{\text{einfach}} \, d\mathbb{P}_F \\
 &\stackrel{\text{Def. Int.}}{=} \sum_{k=1}^{\infty} g(a_k) \mathbb{P}_F(\{a_k\}) \stackrel{\text{Def. Maß}}{=} \sum_{k=1}^{\infty} g(a_k) p_k.
 \end{aligned}$$

(ii) Sei nun g messbar, aber nicht mehr nicht-negativ. Es gilt also wegen (3.3) und Teil (i)

$$\int_{\mathbb{R}} g \, d\mathbb{P}_F \text{ existiert} \Leftrightarrow \int_{\mathbb{R}} |g| \, d\mathbb{P}_F < \infty \Leftrightarrow \sum_{k=1}^N |g(a_k)| p_k < \infty.$$

Wenn das Integral existiert, gilt wegen (i)

$$\begin{aligned}
 \int_{\mathbb{R}} g \, d\mathbb{P}_F &\stackrel{\text{Def.}}{=} \int_{\mathbb{R}} g^+ \, d\mathbb{P}_F - \int_{\mathbb{R}} g^- \, d\mathbb{P}_F \\
 &\stackrel{(i)}{=} \sum_{k=1}^N g^+(a_k) p_k - \sum_{k=1}^N g^-(a_k) p_k \\
 &= \sum_{k=1}^N (g^+(a_k) - g^-(a_k)) p_k = \sum_{k=1}^N g(a_k) p_k.
 \end{aligned}$$

□

Wir werden in der Stochastik das 1.te Moment Erwartungswert nennen, das soll also so etwas wie der Mittelwert über Versuchsausführungen sein (In Vorlesung 26 wird das mit dem Gesetz der großen Zahlen klar). Um mal zwei Beispiele konkret mit den gerade gezeigten Rechenregeln auszurechnen, schauen wir uns die Erwartungswerte (1.te Momente) vom Würfelexperiment und vom gleichverteilten Ziehen aus $[0, 1]$ mal an:

Beispiel 3.3.4.

- Seien $a_1 = 1, \dots, a_6 = 6$ und $p_1 = \dots = p_6 = \frac{1}{6}$. Dann gilt

$$\int_{\mathbb{R}} x \, d\mathbb{P}_F(x) = \sum_{k=1}^6 k \frac{1}{6} = 3,5.$$

Dies ist so ein Beispiel, bei dem die Summe natürlich endlich ist (N endlich, g endlich), wir uns also gar keine Gedanken um Wohldefiniertheit und Endlichkeit machen müssen. Wir können Satz 3.3.3 also als einfache Rechenregeln benutzen.

- Sei jetzt \mathbb{P}_F absolutstetig mit Dichte $f = \mathbf{1}_{[0,1]}$, dafür nutzen wir Satz 3.3.2. Der Integrand von $\int_{\mathbb{R}} x f(x) \, dx$ ist nicht-negativ, also sind alle Integrale wohldefiniert (könnten aber den Werte $+\infty$ annehmen). Wir können also direkt die Rechenregel aus dem Satz benutzen:

$$\int_{\mathbb{R}} x \, d\mathbb{P}_F(x) = \int_{\mathbb{R}} x f(x) \, dx = \int_{\mathbb{R}} x \mathbf{1}_{[0,1]}(x) \, dx = \int_0^1 x \, dx = \left[\frac{1}{2} x^2 \right]_0^1 = \frac{1}{2}.$$

Falls ihr eine grobe Vorstellung von dem Begriff Erwartungswert habt, so sollten diese zwei Beispiele dazu passen. In ein paar Wochen wird euch das auch klar sein.

Warnung 3.3.5. ▶ Es ist nicht immer der Fall, dass eine Verteilungsfunktion diskret oder absolutstetig ist. In diesen Fällen gibt es keine einfache Formel für $\int_{\mathbb{R}} g \, d\mathbb{P}_F$! Es gibt drei Typen von Verteilungsfunktionen:

- F ist absolutstetig,
- F ist diskret,
- F ist „stetigsingulär“ (stetig, hat aber keine Dichte).

Jede stetige Verteilungsfunktion lässt sich zerlegen in absolutstetigen und stetigsingulären Anteil (Satz von Lebesgue). Das geht hier aber zu weit, Beispiele für stetigsinguläre Verteilungen sind tricky (z. B. die Cantorverteilung).

In vielen Beispielen müssen wir gar nicht rechnen, sondern sehen das Ergebnis direkt. Kurze Erinnerung an die Schule: f heißt punktsymmetrisch, falls $f(x) = -f(-x)$ und achsensymmetrisch, falls $f(x) = f(-x)$ für alle $x \in \mathbb{R}$. Für integrierbare punktsymmetrische Funktionen gilt $\int_{\mathbb{R}} f(x) \, dx = 0$. Das können wir direkt ausnutzen, um viele Momente direkt als 0 zu erkennen:

Lemma 3.3.6. ▶ Ist F absolutstetig mit Dichte f und ist das $(2k+1)$ -te Moment von F wohldefiniert, so gilt

$$f \text{ achsensymmetrisch} \quad \Rightarrow \quad \int_{\mathbb{R}} x^{2k+1} \, d\mathbb{P}_F(x) = 0.$$

Beweis. Ist f achsensymmetrisch, so ist $h(x) := x^{2k+1}f(x)$ punktsymmetrisch weil dann $h(-x) = (-x)^{2k+1}f(-x) = -x^{2k+1}f(x) = -h(x)$. Wegen 3.3.2 ist also

$$\int_{\mathbb{R}} x^{2k+1} \, d\mathbb{P}_F(x) = \int_{\mathbb{R}} x^{2k+1}f(x) \, dx \stackrel{\text{punktsym.}}{=} 0.$$

□

Warnung 3.3.7. ▶ [Cauchyverteilung]

Wir müssen in Lemma 3.3.6 auf jeden Fall annehmen, dass die Momente existieren. Beispielsweise hat die Cauchyverteilung eine achsensymmetrische Dichte, das erste Moment ist aber gar nicht definiert, damit insbesondere nicht 0! Dies ist einer der fiesen $\infty - \infty$ Fälle. Rechnen wir das mal nach. Wir betrachten also die Cauchy-Dichte $f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ und berechnen zunächst den Positivteil, wobei wir die konkrete Form des Positivteils x^+ einsetzen:

$$\begin{aligned} \int_{\mathbb{R}} x^+ \, d\mathbb{P}_F(x) &\stackrel{3.3.2, x^+ \geq 0}{=} \frac{1}{\pi} \int_{\mathbb{R}} x^+ \frac{1}{1+x^2} \, dx \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} x \mathbf{1}_{[0, \infty)}(x) \frac{1}{1+x^2} \, dx \\ &= \frac{1}{\pi} \int_0^{\infty} \frac{1}{1/x+x} \, dx \\ &\geq \frac{1}{\pi} \int_1^{\infty} \frac{1}{1+x} \, dx \\ &\stackrel{3.2.1}{=} \frac{1}{\pi} \lim_{N \rightarrow \infty} \int_1^N \frac{1}{1+x} \, dx \\ &\stackrel{\text{Hauptsatz}}{=} \frac{1}{\pi} \lim_{N \rightarrow \infty} \left[\ln(1+x) \right]_1^N = +\infty. \end{aligned}$$

Genauso gibt die selbe Rechnung, mit der konkreten Form des Negativteils x^- ,

$$\int_{\mathbb{R}} x^- d\mathbb{P}_F(x) \stackrel{3.3.2; x^+ \geq 0}{=} \frac{1}{\pi} \int_{-\infty}^0 -x \frac{1}{1+x^2} dx = \frac{1}{\pi} \int_0^{\infty} x \frac{1}{1+x^2} dx = +\infty.$$

Damit ist $\int_{\mathbb{R}} x d\mathbb{P}_F(x) = \int_{\mathbb{R}} x^+ d\mathbb{P}_F(x) - \int_{\mathbb{R}} x^- d\mathbb{P}_F(x) = +\infty - (+\infty)$ nicht wohldefiniert. Für die Cauchyverteilung ist das erste Moment also nicht wohldefiniert!

Die Abschätzung für die Cauchyverteilung war nicht so einfach. Daher wäre es nützlich, den Integralen direkt anzusehen, ob sie existieren, oder nicht. Wenn man weiß was zu tun ist, dann ist die formelle Rechnung viel leichter. Hier ist eine grobe Heuristik:

Bemerkung 3.3.8. ▶ [Heuristik mit Dichten]

Wie „sieht“ man, ob ein Integral existiert? Man vergleicht mit bekannten Integralen. Erinnern wir uns kurz an die Analysisvorlesung:

$$\begin{aligned} \int_1^{\infty} \frac{1}{x^a} dx &\stackrel{3.2.1}{=} \lim_{N \rightarrow \infty} \int_1^N \frac{1}{x^a} dx \\ &= \lim_{N \rightarrow \infty} \begin{cases} \left[\ln(x) \right]_1^N & : a = 1 \\ \left[\frac{1}{1-a} x^{1-a} \right]_1^N & : a \neq 1 \end{cases} \\ &= \lim_{N \rightarrow \infty} \begin{cases} \ln(N) & : a = 1 \\ \frac{1}{1-a} (N^{1-a} - 1) & : a \neq 1 \end{cases} \\ &= \begin{cases} +\infty & : a \leq 1 \\ \frac{1}{a-1} < \infty & : a > 1 \end{cases}. \end{aligned}$$

Unsere Heuristik für die Integrierbarkeit bei unendlich ist es, grob mit der Funktion $\frac{1}{x}$ zu vergleichen. Fällt ein Integrand f deutlich schneller gegen 0, ist vermutlich $\int_1^{\infty} f(x) dx < \infty$. Fällt der Integrand langsamer als $\frac{1}{x}$ gegen 0, so ist sicherlich $\int_1^{\infty} f(x) dx = +\infty$.

Hier sind ein paar Beispiele:

- (i) Nochmal die Cauchyverteilung: $\frac{1}{\pi} x \frac{1}{1+x^2}$ ist bei ∞ ungefähr wie $\frac{1}{x}$, das ist aber *nicht* integrierbar. Die Heuristik sagt uns also auf einen Blick, dass etwas schief gehen sollte. Um daraus ein sauberes Argument zu machen, müssen wir leider doch die Abschätzung aus 3.3.7 durchgehen.
- (ii) Für welches β hat $\text{Exp}(\lambda)$ ein endliches exponentielle Moment, wann ist also

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

endlich? Weil e^{ax} für $a > 0$ schneller als jedes Polynom wächst, fällt e^{-ax} ($\lambda > 0$) viel schneller als jedes Polynom gegen 0. Also sind alle exponentiellen Momente genau dann endlich, wenn $\beta < \lambda$. In dem Fall können wir alles natürlich sofort ausrechnen, weil der Integrand eine einfache Stammfunktion hat.

- (iii) Für welches β hat $\mathcal{N}(\mu, \sigma^2)$ ein endliches exponentielles Moment, wann ist also

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} e^{\beta x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

definiert? Natürlich geht $e^{\beta x} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ viel schneller als $\frac{1}{x}$ gegen 0, weil x^2 schneller wächst als x , und die Exponentialfunktion alles polynomielle dominiert.

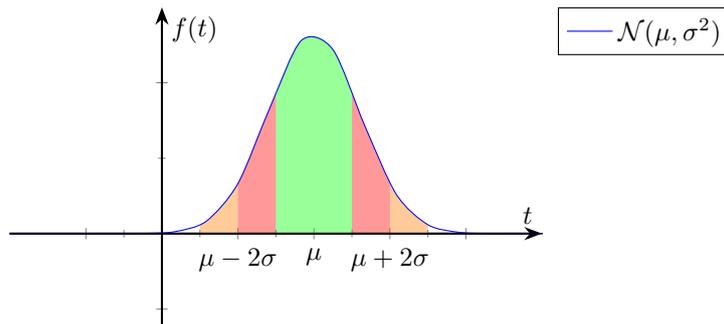
Um einen ersten Eindruck zu bekommen, warum Momente überhaupt nützlich sind, schauen wir uns eine Variante der Markov-Ungleichung an. Wir sehen hier, dass wir mit den Momenten etwas über die Verteilung der Masse aussagen können. Schauen wir uns dazu die Konzentration der Masse der Normalverteilung um μ an:

Beispiel 3.3.9. ▶ [Konzentrationsungleichung]

Wir kennen bereits die Normalverteilung $\mathcal{N}(\mu, \sigma^2)$. Der Parameter μ verschiebt die Stelle des Maximalpunktes der Dichte an die Stelle μ , der Parameter σ^2 ist für die Stauchung zuständig. Für kleines σ^2 ist die Dichte „spitzer“, für großes σ^2 flacher. Qualitativ wissen wir schon, dass viel Masse nah bei μ liegt, dort ist die Dichte schließlich groß (vergleiche die Diskussion 1.4.13). Können wir das auch genauer machen? Wenn jemand fragt, wie weit man von μ weggehen muss, so dass beispielsweise 99,7 Prozent der Masse in $[\mu - a, \mu + a]$ liegt, was antworten wir? Schauen wir das Bildchen an: Reicht der grüne Bereich, um $\mathbb{P}([\mu - a, \mu + a]) \approx 0,997$ zu erreichen? Vermutlich nicht. Grün, rot und orange zusammen könnten vom Bild her aber hinhalten. Solche Fragen durch Ungleichungen möglichst gut zu beantworten, sind das Themengebiet der **Konzentrationsungleichungen**, also Ungleichungen der Art

$$\mathbb{P}([a, b]) \leq \dots \quad \text{oder} \quad \mathbb{P}([a, b]) \geq \dots$$

Das ist natürlich einfach zu beantworten, wenn die Dichte f von \mathbb{P} explizit ist, weil dann $\mathbb{P}_F([a, b]) = F(b) - F(a)$ gilt. Bei der Exponentialverteilung braucht man z. B. nichts abzuschätzen weil F eine einfache Formel hat. Für die Normalverteilung geht das allerdings nicht, F ist als Integral gegeben und das Integral kann nicht ausgerechnet werden.



Das bunte Bildchen visualisiert die sogenannte 1-2-3- σ -Regel für $\mathcal{N}(\mu, \sigma^2)$. Die Regel besagt, dass

- in $[\mu - \sigma, \mu + \sigma]$ ungefähr 68 Prozent (also 0,68) der Masse liegt,
- in $[\mu - 2\sigma, \mu + 2\sigma]$ ungefähr 95 Prozent (also 0,95) der Masse liegt,
- in $[\mu - 3\sigma, \mu + 3\sigma]$ ungefähr 99,7 Prozent (also 0,997) der Masse liegt.

Weil σ auch Standardabweichung genannt wird, heißt das in Worten: „Um zwei Standardabweichungen nach links und rechts von μ liegt 95 Prozent der Masse von $\mathcal{N}(\mu, \sigma^2)$ “.

Wir zeigen jetzt mit der Markov-Ungleichung unsere erste Integralabschätzung und versuchen uns danach an einem Teil der 1-2-3- σ -Regel.

Proposition 3.3.10. ▶ [Markov-Ungleichung für Polynome]

Sei \mathbb{P}_F ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R})$, sodass für eine gerade natürliche Zahl $2k$ das $2k$ -te Moment existiert ist. Dann gilt, für alle $a > 0$.

$$\mathbb{P}_F([-a, a]) \geq 1 - \frac{\int_{\mathbb{R}} x^{2k} d\mathbb{P}_F(x)}{a^{2k}}.$$

Gleichbedeutend (Gegenereignis) gilt

$$\mathbb{P}_F([-a, a]^C) \leq \frac{\int_{\mathbb{R}} x^{2k} d\mathbb{P}_F(x)}{a^{2k}}.$$

Beweis. Der Beweis ist tatsächlich sehr einfach und basiert auf dem kleinen Trick, den wir schon ein paar mal gesehen haben. Wir mogeln einen Indikator über eine Menge in das Integral und schätzen auf dem Indikator die Funktion ab. Das geht natürlich nur, wenn der Indikator über eine messbare Menge so gewählt wird, dass die Menge etwas mit dem Integranden zu tun hat. Wir nehmen dazu den Integranden $g(x) = x^{2k}$ und die Menge $[-a, a]^C$. Für $x \in [-a, a]^C$ gilt natürlich $|x| > a$ und damit, $2k$ ist gerade, $x^{2k} = |x|^{2k} \geq a^{2k}$. Dann müssen wir nur noch die Monotonie vom Integral nutzen:

$$\begin{aligned} \int_{\mathbb{R}} x^{2k} d\mathbb{P}_F(x) &\stackrel{\text{Mon.}}{\geq} \int_{\mathbb{R}} x^{2k} \mathbf{1}_{[-a, a]^C}(x) d\mathbb{P}_F(x) \\ &\geq \int_{\mathbb{R}} \underbrace{a^{2k} \mathbf{1}_{[-a, a]^C}(x)}_{\text{einfach}} d\mathbb{P}_F(x) \\ &\stackrel{\text{Def.}}{=} a^{2k} \mathbb{P}_F([-a, a]^C). \end{aligned}$$

Auflösen gibt

$$\mathbb{P}_F([-a, a]^C) \leq \frac{\int_{\mathbb{R}} x^{2k} d\mathbb{P}_F(x)}{a^{2k}}.$$

Die zweite Ungleichung ist die Gegenwahrscheinlichkeit, weil für Wahrscheinlichkeitsmaße $\mathbb{P}_F(B) = 1 - \mathbb{P}_F(B^C)$ gilt. \square

Kommen wir zurück zur Konzentration der Normalverteilung um μ .

Beispiel 3.3.11.  Aufgabe: Sei \mathbb{P} normalverteilt mit Parametern $\mu = 0$ und $\sigma^2 > 0$. Finde ein $a > 0$ mit $\mathbb{P}([-a, a]) \approx 0.997$. Damit uns die Markov-Ungleichung explizite Zahlen gibt, brauchen wir Formel für gerade Momente, wir nehmen einfach mal das 2.te und das 8.te. Für 2.te und 8.te Momente der Normalverteilungen gelten, das sehen wir später (oder ihr rechnet die Integrale von Hand aus),

$$\int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sigma^2 \quad \text{und} \quad \int_{\mathbb{R}} x^8 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 105\sigma^8.$$

Um die Konzentration der Normalverteilung in $[-a, a]$ abzuschätzen, verwenden wir Proposition 3.3.10 einmal mit $2k = 2$ und einmal mit $2k = 8$:

$$\mathbb{P}([-a, a]) \geq 1 - \frac{\sigma^2}{a^2} \quad \text{sowie} \quad \mathbb{P}([-a, a]) \geq 1 - \frac{105\sigma^8}{a^8}.$$

Wir probieren jetzt mal aus, welche Abschätzung besser ist. Einsetzen, umformen und in Taschenrechner einsetzen, gibt beim zweiten Moment ungefähr

$$1 - \frac{\sigma^2}{a^2} \geq 0.997 \quad \Leftrightarrow \quad a \geq 18,26 \cdot \sigma$$

und beim 8.ten Moment ungefähr

$$1 - \frac{105\sigma^8}{a^8} \geq 0.997 \quad \Leftrightarrow \quad a \geq 3,69 \cdot \sigma.$$

Das ist jetzt zwar nicht ganz an an der richtigen Lösung aus Beispiel 3.3.9, aber für das achte Moment auch nicht ganz weit davon entfernt.

Vorlesung 14

3.4 Integralabschätzungen und L^p -Räume

Sei jetzt wieder $(\Omega, \mathcal{A}, \mu)$ ein beliebiger messbarer Raum und $f: \Omega \rightarrow \overline{\mathbb{R}}$ sei $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -messbar. Wir werden in diesem Kapitel mehrfach nutzen, dass wegen Satz 3.1.15 folgende Äquivalenz gilt:

$$\int_{\Omega} |f| d\mu = 0 \quad \Leftrightarrow \quad |f| = 0 \text{ } \mu\text{-fast überall} \quad \Leftrightarrow \quad f = 0 \text{ } \mu\text{-fast überall.}$$

Satz 3.4.1.  **[Hölder-Ungleichung]**

Seien $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$\int_{\Omega} |fg| \, d\mu \leq \left(\int_{\Omega} |f|^p \, d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} |g|^q \, d\mu \right)^{\frac{1}{q}}.$$

Beweis. Alle auftretenden Integranden sind messbar und nicht-negativ, also sind alle Integrale definiert, $+\infty = +\infty$ ist aber möglich. Wir erinnern an die Young-Ungleichung aus Analysis 2 (das ist gerade die Konkavität des Logarithmus): Für $\alpha, \beta > 0$ gilt

$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}.$$

Ist ein Faktor der rechten Seite der Hölder-Ungleichung 0 oder $+\infty$, so ist nichts zu zeigen. Das ist sofort klar für $+\infty$, aber auch der Fall 0 ist nicht schwer: Wenn nämlich $(\int_{\Omega} |f|^p \, d\mu)^{1/p} = 0$ gilt, so muss $f = 0$ μ -fast überall gelten. Also ist auch $|fg| = 0$ μ -fast überall und damit ist auch die linke Seite 0. Die Ungleichung ergibt dann also $0 \leq 0$ und das ist richtig. Also nehmen wir an, beide Faktoren sind positiv. Wir definieren

$$\sigma = \left(\int_{\Omega} |f|^p \, d\mu \right)^{\frac{1}{p}} > 0 \quad \text{und} \quad \tau = \left(\int_{\Omega} |g|^q \, d\mu \right)^{\frac{1}{q}} > 0$$

sowie die messbaren Abbildungen

$$\alpha = \frac{|f|}{\sigma} \quad \text{und} \quad \beta = \frac{|g|}{\tau}.$$

Mit Young folgt

$$\frac{|f(\omega)g(\omega)|}{\sigma\tau} \leq \frac{|f(\omega)|^p}{\sigma^p p} + \frac{|g(\omega)|^q}{\tau^q q}, \quad \forall \omega \in \Omega.$$

Integrieren beider Seiten gibt wegen der Monotonie des Integrals

$$\frac{1}{\sigma\tau} \int_{\Omega} |fg| \, d\mu \leq \frac{\int_{\Omega} |f|^p \, d\mu}{\sigma^p p} + \frac{\int_{\Omega} |g|^q \, d\mu}{\tau^q q} = \frac{1}{p} + \frac{1}{q} = 1.$$

Durchmultiplizieren gibt die Höldersche Ungleichung. □

Der Fall $p = q = 2$ heißt auch Cauchy-Schwarz:

$$\left(\int_{\Omega} |fg| \, d\mu \right)^2 \leq \int_{\Omega} |f|^2 \, d\mu \int_{\Omega} |g|^2 \, d\mu.$$

Korollar 3.4.2.  Ist \mathbb{P} ein Wahrscheinlichkeitsmaß auf \mathcal{A} , so gilt

$$\left(\int_{\Omega} |f| \, d\mathbb{P} \right)^p \leq \int_{\Omega} |f|^p \, d\mathbb{P}$$

für alle $p > 1$.

Beweis. Wegen $\mathbb{P}(\Omega) = 1$ gilt

$$\int_{\Omega} |f| \, d\mathbb{P} = \int_{\Omega} |f| \cdot 1 \, d\mathbb{P} \leq \left(\int_{\Omega} |f|^p \, d\mathbb{P} \right)^{\frac{1}{p}} \left(\int_{\Omega} \underbrace{|1|^q}_{=1 \cdot 1_{\Omega}} \, d\mathbb{P} \right)^{\frac{1}{q}} = \left(\int_{\Omega} |f|^p \, d\mathbb{P} \right)^{\frac{1}{p}} \cdot 1,$$

also die Behauptung. □

Satz 3.4.3. ▶ [Minkowski-Ungleichung]Sei $p \geq 1$, so gilt

$$\left(\int_{\Omega} |f + g|^p \, d\mu \right)^{1/p} \leq \left(\int_{\Omega} |f|^p \, d\mu \right)^{1/p} + \left(\int_{\Omega} |g|^p \, d\mu \right)^{1/p}.$$

Beide Seiten können den Wert $+\infty$ annehmen.*Beweis.* Wie in Analysis 2, folgt aus Hölder und der Young Ungleichung. Wir zeigen die stärkere Ungleichung

$$\left(\int_{\Omega} (|f| + |g|)^p \, d\mu \right)^{1/p} \leq \left(\int_{\Omega} |f|^p \, d\mu \right)^{1/p} + \left(\int_{\Omega} |g|^p \, d\mu \right)^{1/p}. \quad (3.7)$$

Tatsächlich impliziert (3.7) Minkowski weil die linke Seite von Minkowski kleiner ist als die linke Seite von (3.7). Das folgt direkt aus der Monotonie des Integrals und weil $|f + g| \leq |f| + |g|$ gilt.

- (a) Für $p = 1$ gilt wegen der Linearität des Integrals (3.7) natürlich mit Gleichheit.
- (b) Sei nun $p > 1$. Ist die rechte Seite $+\infty$, so gilt (3.7). Also nehmen wir an, dass beide Integrale der rechten Seite endlich sind. Dann ist aber auch die linke Seite wegen der elementaren Abschätzung

$$(|f| + |g|)^p \leq (2|f| \vee |g|)^p = 2^p (|f|^p \vee |g|^p) \leq 2^p (|f|^p + |g|^p)$$

und der Monotonie des Integrals endlich. Damit nun zum Beweis von (3.7):

$$\begin{aligned} \int_{\Omega} (|f| + |g|)^p \, d\mu &= \int_{\Omega} (|f| + |g|)^{p-1} (|f| + |g|) \, d\mu \\ &\stackrel{\text{ausm., Lin.}}{=} \int_{\Omega} (|f| + |g|)^{p-1} |f| \, d\mu + \int_{\Omega} (|f| + |g|)^{p-1} |g| \, d\mu \\ &\stackrel{2 \times \text{Hölder}}{\leq} \left(\int_{\Omega} |f|^p \, d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} (|f| + |g|)^{(p-1)q} \, d\mu \right)^{\frac{1}{q}} \\ &\quad + \left(\int_{\Omega} |g|^p \, d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} (|f| + |g|)^{(p-1)q} \, d\mu \right)^{\frac{1}{q}} \\ &\stackrel{\text{auskl.}}{=} \left(\left(\int_{\Omega} |f|^p \, d\mu \right)^{\frac{1}{p}} + \left(\int_{\Omega} |g|^p \, d\mu \right)^{\frac{1}{p}} \right) \left(\int_{\Omega} |f| + |g| \, d\mu \right)^{1 - \frac{1}{p}}. \end{aligned}$$

In der letzten Gleichung haben wir genutzt, dass $1 - \frac{1}{p} = \frac{1}{q}$ und $(p-1)q = p$ aufgrund der Voraussetzung an p und q gelten. Rübermultiplizieren des zweiten Faktors gibt dann (3.7). □

Definition 3.4.4. ▶ $f: \Omega \rightarrow \overline{\mathbb{R}}$ messbar heißt **p -fach integrierbar**, falls $\int_{\Omega} |f|^p \, d\mu < \infty$. Statt 2-fach integrierbar sagt man auch **quadratintegrierbar**, statt 1-fach integrierbar sagt man **integrierbar**. Natürlich passt die Notation zu unserer ursprünglichen Definition der Integrierbarkeit:

$$f \mu\text{-int.} \quad \stackrel{\text{alte}}{\Leftrightarrow} \text{Def.} \quad \int_{\Omega} f^+ \, d\mu < \infty, \int_{\Omega} f^- \, d\mu < \infty \quad \stackrel{\text{Übung}}{\Leftrightarrow} \quad \int_{\Omega} |f| \, d\mu < \infty \quad \stackrel{\text{neue}}{\Leftrightarrow} \text{Def.} \quad f \text{ int.}$$

Schauen wir uns ein ganz konkretes Beispiel für die neue Definition an.

Beispiel 3.4.5. ▶

- Für $\Omega = [1, \infty)$, $\mathcal{A} = \mathcal{B}([1, \infty))$, $\mu = \lambda_{|[1, \infty)}$ ist $\int_{\Omega} f \, d\mu = \int_1^{\infty} f(x) \, dx$. Für $f(x) = \frac{1}{x^a}$ ist f p -fach integrierbar genau dann, wenn $p > \frac{1}{a}$.

- Für $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$, $\mu = \lambda_{[0,1]}$ ist $\int_{\Omega} f \, d\mu = \int_0^1 f(x) \, dx$. Für

$$f(x) = \begin{cases} \frac{1}{x^\alpha} & : x \in (0, 1] \\ +\infty & : x = 0 \end{cases}$$

ist f p -fach integrierbar genau dann, wenn $p < \frac{1}{\alpha}$. Für das Integral ist der Funktionswert $+\infty$ unproblematisch, da die Menge $\{0\}$ eine μ -Nullmenge ist.

In der Mathematik wollen wir aus allen Objekten möglichst nützliche Strukturen schaffen, in diesem Fall einen Vektorraum.

Definition 3.4.6. ▶

$$\mathcal{L}^p(\mu) := \left\{ f: \Omega \rightarrow \overline{\mathbb{R}} \text{ messbar} \mid \int_{\Omega} |f|^p \, d\mu < \infty \right\}.$$

Manchmal schreibt man auch $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ oder nur \mathcal{L}^p .

Lemma 3.4.7. ▶ Mit punktweiser Addition und Skalarmultiplikation ist $\mathcal{L}^p(\mu)$ ein reeller Vektorraum, sogar ein Untervektorraum der messbaren Funktionen, d. h.

- (i) $0 \in \mathcal{L}^p(\mu)$, wobei 0 die konstante Nullfunktion ist.
- (ii) $f, g \in \mathcal{L}^p(\mu) \Rightarrow f + g \in \mathcal{L}^p(\mu)$,
- (iii) $\alpha \in \mathbb{R}, f \in \mathcal{L}^p(\mu) \Rightarrow \alpha f \in \mathcal{L}^p(\mu)$,

Beweis. Messbare Funktionen mit punktweiser Addition und skalarer Multiplikation geben einen Vektorraum. Die Eigenschaften (ii)-(i) bedeuten, dass $\mathcal{L}^p(\mu)$ ein Untervektorraum ist. Wir prüfen also nur die dafür benötigten Eigenschaften:

- (i) $\int_{\Omega} |0|^p \, d\mu = 0 < \infty$
- (ii) $\int_{\Omega} |\alpha f|^p \, d\mu \stackrel{\text{Lin.}}{=} |\alpha|^p \int_{\Omega} |f|^p \, d\mu < \infty$ weil $f \in \mathcal{L}^p(\mu)$ angenommen wurde. Also gilt auch $\alpha f \in \mathcal{L}^p(\mu)$.
- (iii) Wegen Minkowski gilt

$$\int_{\Omega} |f + g|^p \, d\mu \leq \left(\underbrace{\left(\int_{\Omega} |f|^p \, d\mu \right)^{\frac{1}{p}}}_{< \infty} + \underbrace{\left(\int_{\Omega} |g|^p \, d\mu \right)^{\frac{1}{p}}}_{< \infty} \right)^p < \infty.$$

Also ist $f + g \in \mathcal{L}^p(\mu)$. □

Aus der Analysis wisst ihr schon, das man aus einem Vektorraum immer gerne einen normierten Raum machen möchte. Dann kann man zum über Folgenkonvergenz und Stetigkeit sprechen. Können wir also aus $\mathcal{L}^p(\mu)$ einen normierten Raum machen? Nein!

Lemma 3.4.8. ▶

$$\|f\|_p := \left(\int_{\Omega} |f|^p \, d\mu \right)^{\frac{1}{p}}$$

ist eine **Halbnorm** auf $\mathcal{L}^p(\mu)$, d. h. es gelten für $f, g \in \mathcal{L}^p(\mu)$ und $\alpha \in \mathbb{R}$

- (i) $0 \leq \|f\|_p < \infty, \|0\|_p = 0$ (Definitheit fehlt)
- (ii) $\|\alpha f\|_p = |\alpha| \|f\|_p$

$$(iii) \|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Beweis. Die ersten zwei Eigenschaften sind klar. Die Dreiecksungleichung in $\mathcal{L}^p(\mu)$ ist gerade die Minkowski Ungleichung! \square

Warnung: $\|\cdot\|_p$ ist *keine* Norm! Jedes f mit $\mu(\{f \neq 0\}) = 0$ erfüllt

$$\|f\|_p = \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} = 0.$$

Wenn f auf einer μ -Nullmenge ungleich 0 ist, so ist aber $f \neq 0$. Also ist die Definitheit nicht erfüllt. Mit einem Trick kann man $\mathcal{L}^p(\mu)$ zu einem normierten Vektorraum ändern, indem man das Problem der Definitheit wegdefiniert. Dazu betrachtet man den Quotientenraum bezüglich der Äquivalenzrelation

$$f \sim g \quad :\Leftrightarrow \quad f = g \text{ } \mu\text{-fast überall,}$$

der aus den Äquivalenzklassen

$$[f] := \{g \in \mathcal{L}^p(\mu) : f \sim g\} = \{g \in \mathcal{L}^p(\mu) : f = g \text{ } \mu\text{-fast überall}\}$$

besteht. Die Operationen und die Norm werden durch beliebige Repräsentanten der Äquivalenzklassen definiert:

$$[f] + [g] := [f + g], \quad \alpha[f] := [\alpha f] \quad \text{und} \quad \|[f]\|_p := \|f\|_p.$$

Der Quotientenraum wird als $L^p(\mu) = \{[f] : f \in \mathcal{L}^p(\mu)\}$ bezeichnet. Gemeinsam mit den Operationen und der Norm auf den Elementen (Äquivalenzklassen) ist $L^p(\mu)$ ein normierter Vektorraum:

Satz 3.4.9.  Für $p \geq 1$ ist $L^p(\mu)$ mit den gerade definierten Operationen ein normierter Vektorraum und mit $\|\cdot\|_p$ ein vollständiger normierter Raum (Banachraum).

Beweis. Wir zeigen, dass $L^p(\mu)$ mit den definierten Operationen ein normierter Vektorraum ist.

- (a) Vektorraum aus Lineare Algebra 1.
- (b) Die Eigenschaften der Halbnorm folgen direkt aus der Definition weil $\|\cdot\|_p$ eine Halbnorm auf $\mathcal{L}^p(\mu)$ ist. Es fehlt also nur noch die Definitheit:

$$\begin{aligned} \|[f]\|_p = 0 &\stackrel{\text{Def.}}{\Leftrightarrow} \|f\|_p = 0 \\ &\stackrel{\text{Def.}}{\Leftrightarrow} \int_{\Omega} |f|^p d\mu = 0 \\ &\Leftrightarrow |f|^p = 0 \text{ } \mu\text{-fast überall} \\ &\Leftrightarrow f = 0 \text{ } \mu\text{-fast überall} \\ &\Leftrightarrow [f] = [0] \end{aligned}$$

Damit ist $\|\cdot\|_p$ eine Norm.

Was wir hier nicht zeigen werden, ist die Vollständigkeit von $L^p(\mu)$. Aber irgendwas sollten wir ja für die Funktionalanalysis übrig lassen! \square

3.5 Produktmaße und Satz von Fubini

Jetzt wird es noch einmal so richtig dreckig, bevor wir die wunderbare Welt der Stochastik erobern können. Sobald wir uns durch die Konstruktion des Produktmaßes und des Satzes von Fubini gequält haben, fällt uns alles andere aber sofort vor die Füße.

Im Folgenden seien $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ Maßräume und

$$\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

das kartesische Produkt aus Analysis 1. Wir wollen auf Ω eine σ -Algebra und darauf ein Maß mit einer schönen Produkteigenschaft definieren (deshalb wird das Maß Produktmaß heißen).

Definition 3.5.1. 

(i) Die σ -Algebra

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\})$$

heißt **Produkt- σ -Algebra** auf $\Omega_1 \times \Omega_2$.

(ii) Ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ heißt **Produktmaß**, falls

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2) \tag{3.8}$$

für alle Mengen $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$.

Natürlich wäre es schön, wenn die Produkt- σ -Algebra einfach nur aus allen Mengen $A_1 \times A_2$ bestehen würde. Leider gibt das keine σ -Algebra (die Komplementbildung geht schief), die Mengen geben nur einen \cap -stabilen Erzeuger von $\mathcal{A}_1 \otimes \mathcal{A}_2$. Die Eigenschaft des Produktmaßes legt das Produktmaß also nur auf einem \cap -stabilen Erzeuger fest. Mit unseren Kenntnissen der Maßtheorie, könnten wir also reflexhaft sagen: Kein Problem, mit Carathéodory können wir ein Produktmaß aus den Werten auf dem Erzeuger konstruieren und wegen Dynkin-Systemen kann es nur ein Maß mit der Produktmaßeigenschaft bekommen. Genau richtig - das funktioniert! Wir quälen uns im nächsten Satz aber gewaltig mehr. Wir konstruieren das Produktmaß nicht mit Carathéodory, sondern schreiben eine Formel hin. Das ist in der Tat viel komplizierter, der Vorteil ist aber, dass wir damit den darauf folgenden Satz von Fubini schon fast bewiesen haben. Würden wir das Produktmaß mit Carathéodory konstruieren, würde der ganze Aufwand in den Beweis von Fubini verschoben.

Satz 3.5.2.  **[Konstruktion Produktmaß]**

Sind μ_1, μ_2 σ -endliche Maße auf $\mathcal{A}_1, \mathcal{A}_2$, so existiert ein eindeutiges Maß $\mu_1 \otimes \mu_2$ auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)$$

für alle Mengen $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$.

Beweis. **Eindeutigkeit:** Für die Eindeutigkeit nutzen wir wieder den Eindeutigkeitssatz 1.2.13, der auf Dynkin-Systemen beruht. Sei

$$\mathcal{S} = \{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\},$$

dann ist \mathcal{S} ein \cap -stabiler Erzeuger von $\mathcal{A}_1 \otimes \mathcal{A}_2$. Um den Eindeutigkeitssatz anzuwenden, müssen wir noch zeigen, dass Produktmaße σ -endlich sein müssen, wir brauchen also eine wachsende Folge in \mathcal{S} , die endliches Maß hat und Ω ausfüllt. Weil μ_1, μ_2 σ -endliche Maße sind, existieren Folgen $(E_n^1)_{n \in \mathbb{N}} \subseteq \mathcal{A}_1, (E_n^2)_{n \in \mathbb{N}} \subseteq \mathcal{A}_2$ mit $E_n^1 \uparrow \Omega_1, E_n^2 \uparrow \Omega_2$ und $\mu_1(E_n^1) < \infty, \mu_2(E_n^2) < \infty$ für alle $n \in \mathbb{N}$. Sei nun $E_n := E_n^1 \times E_n^2$, dann gelten

$$E_n \uparrow \Omega, n \rightarrow \infty, \quad \text{und} \quad \mu_1(E_n^1) \cdot \mu_2(E_n^2) < \infty.$$

Aus dem Eindeutigkeitsatz folgt also, dass zwei Maße μ und $\bar{\mu}$, die die Definition des Produktmaßes erfüllen, gleich sind. Es kann also nur ein Produktmaß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ geben. Ob es so ein Maß gibt, ist natürlich noch nicht klar.

Existenz: Anstatt das Produktmaß mit Carathéodory zu konstruieren, schreiben wir es einfach hin. Hier ist es:

$$\mu(A) := \int_{\Omega_1} \mu_2(A_{\omega_1}) d\mu_1, \quad A \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$

wobei $A_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}$. Wir machen sogar noch mehr, wir schreiben noch ein zweites Produktmaß hin:

$$\bar{\mu}(A) := \int_{\Omega_2} \mu_1(A_{\omega_2}) d\mu_2, \quad A \in \mathcal{A}_1 \otimes \mathcal{A}_2,$$

wobei jetzt $A_{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$. Wir zeigen nun, dass μ ein Produktmaß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ ist. Mit exakt demselben Beweis zeigt man auch, dass $\bar{\mu}$ ein Produktmaß ist, weshalb dann wegen der Eindeutigkeit auch $\mu = \bar{\mu}$ gilt. Zeigen wir also die Eigenschaften eines Maßes sowie die definierende Eigenschaft des Produktmaßes:

- (i) $\mu: \mathcal{A}_1 \otimes \mathcal{A}_2 \rightarrow [0, \infty]$ gilt, weil μ_2 ein Maß ist (deshalb nicht-negativ) und Integrale über nicht-negative Funktionen nicht-negativ sind.
- (ii) $\mu(\emptyset) = 0$ gilt, weil \emptyset_{ω_1} auch die leere Menge ist und Maße der leeren Menge 0 sind.
- (iii) Nun zur σ -Additivität. Seien dazu $A^1, A^2, \dots \in \mathcal{A}_1 \otimes \mathcal{A}_2$ paarweise disjunkt und sei

$$A := \bigcup_{k=1}^{\infty} A^k.$$

Dann gilt $A_{\omega_1} = \bigcup_{k=1}^{\infty} A_{\omega_1}^k$ und mit den Maßeigenschaften sowie monotoner Konvergenz

$$\begin{aligned} \mu(A) &\stackrel{\text{Def.}}{=} \int_{\Omega_1} \mu_2\left(\left(\bigcup_{k=1}^{\infty} A^k\right)_{\omega_1}\right) d\mu_1(\omega_1) \\ &= \int_{\Omega_1} \mu_2\left(\bigcup_{k=1}^{\infty} A_{\omega_1}^k\right) d\mu_1(\omega_1) \\ &\stackrel{\mu_2 \text{ } \sigma\text{-add.}}{=} \int_{\Omega_1} \sum_{k=1}^{\infty} \mu_2(A_{\omega_1}^k) d\mu_1(\omega_1) \\ &\stackrel{3.2.1}{=} \sum_{k=1}^{\infty} \int_{\Omega_1} \mu_2(A_{\omega_1}^k) d\mu_1(\omega_1) \stackrel{\text{Def.}}{=} \sum_{k=1}^{\infty} \mu_1(A^k). \end{aligned}$$

- (iv) μ ist also ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$. Wir müssen noch die definierende Eigenschaft auf dem Erzeuger zeigen. Sei dazu $A = A_1 \times A_2$. Weil aufgrund der Definitionen

$$(A_1 \times A_2)_{\omega_1} = \begin{cases} \emptyset & : \omega_1 \notin A_1 \\ A_2 & : \omega_1 \in A_1 \end{cases}$$

gilt, folgt aufgrund der Definition des Integrals für einfache Integranden

$$\begin{aligned} \mu(A_1 \times A_2) &= \int_{\Omega_1} \mu_2((A_1 \times A_2)_{\omega_1}) d\mu_1(\omega_1) \\ &= \int_{\Omega_1} \mu_2(A_2) \mathbf{1}_{A_1}(\omega_1) d\mu_1(\omega_1) \\ &\stackrel{\text{Linear}}{=} \mu_2(A_2) \int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1) d\mu_1(\omega_1) = \mu_2(A_2) \cdot \mu_1(A_1). \end{aligned}$$

Also ist μ ein Produktmaß.

Eigentlich könnte alles so schön sein, und der Beweis ist hier zu Ende. Leider haben wir geschummelt. Warum ist μ überhaupt sinnvoll definiert? Klingt blöd, ist aber gar nicht so klar. Warum ist der Integrand überhaupt definiert, d. h. warum gilt $A_{\omega_1} \in \mathcal{A}_2$? Unklar. Warum ist $\omega_1 \mapsto \mu_2(A_{\omega_1})$ überhaupt $(\mathcal{A}_1, \mathcal{B}(\overline{\mathbb{R}}))$ -messbar, warum macht das Integral also überhaupt Sinn? Unklar. Wir sollten also beides noch checken.

Wir zeigen jetzt nacheinander

- (a) $A_{\omega_1} \in \mathcal{A}_2$ für alle $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$
- (b) $\omega_1 \mapsto \mu_2(A_{\omega_1})$ ist $(\mathcal{A}_1, \mathcal{B}(\overline{\mathbb{R}}))$ -messbar.

Zu (a): Wir folgen dem Trick der guten Mengen. Seien dazu die guten Mengen

$$\mathcal{F} = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : A_{\omega_1} \in \mathcal{A}_2\}.$$

Wir zeigen: $\mathcal{F} = \mathcal{A}_1 \otimes \mathcal{A}_2$. Dazu zeigen wir zunächst, dass \mathcal{F} eine σ -Algebra ist:

- $\Omega \in \mathcal{F}$ gilt, weil $\Omega_{\omega_1} = \Omega_2 \in \mathcal{F}$ gilt.
- Sei $A \in \mathcal{F}$, dann ist $A^C \in \mathcal{F}$ weil $(A^C)_{\omega_1} = (A_{\omega_1})^C \in \mathcal{A}_2$, da \mathcal{A}_2 als σ -Algebra abgeschlossen unter Komplementbildung ist.
- Genauso mit abzählbaren Vereinigungen: Wegen der Abgeschlossenheit der σ -Algebra \mathcal{A}_2 unter Bildung von abzählbaren Vereinigungen, gilt für $A^1, A^2, \dots \in \mathcal{F}$

$$\left(\bigcup_{k=1}^{\infty} A^k \right)_{\omega_1} = \bigcup_{k=1}^{\infty} \underbrace{A_{\omega_1}^k}_{\in \mathcal{A}_2} \in \mathcal{A}_2.$$

Damit ist $\bigcup_{k=1}^{\infty} A^k \in \mathcal{F}$.

Folglich ist \mathcal{F} eine σ -Algebra. Weil $\mathcal{S} \subseteq \mathcal{F}$ und $\sigma(\mathcal{S}) \stackrel{\text{Def.}}{=} \mathcal{A}_1 \otimes \mathcal{A}_2$ gilt, bekommen wir zusammen:

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{S}) \subseteq \sigma(\mathcal{F}) = \mathcal{F} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$$

Weil links und rechts das gleiche steht, bekommen wir überall Gleichheiten, es gilt also $\mathcal{F} = \mathcal{A}_1 \otimes \mathcal{A}_2$ und die Behauptung folgt. Das war wieder unser „standard“ gute Mengen Trick in der einfacheren Situation mit σ -Algebren.

Nun zu (b): Wir checken erst den Fall $\mu_2(\Omega_2) < \infty$ und schieben die Aussage dann mit der σ -Endlichkeit auf den allgemeinen Fall. Sei also erstmal $\mu_2(\Omega_2) < \infty$. Wir zeigen, wieder mit dem gute Mengen Trick (aber in der Dynkin-System Variante), $\mathcal{A}_1 \otimes \mathcal{A}_2 = \mathcal{F}$, mit der Menge

$$\mathcal{F} := \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : \omega_1 \mapsto \mu_2(A_{\omega_1}) \text{ messbar}\}$$

der guten Mengen. Es gilt $\mathcal{S} \subseteq \mathcal{F}$, da

$$\omega_1 \mapsto \mu_2((A_1 \times A_2)_{\omega_1}) = \underbrace{\mu_2(A_2) \mathbf{1}_{A_1}}_{\text{messbar}}(\omega_1)$$

als Produkt messbarer Abbildungen messbar ist. Um wie in (a) zu argumentieren, zeigen wir nun, dass \mathcal{F} ein Dynkin-System ist:

- $\Omega \in \mathcal{F}$ ist klar, weil $\omega_1 \mapsto \mu_2(\Omega_{\omega_1}) = \mu_2(\Omega_2) < \infty$ konstant und damit messbar ist.
- Sei $A \in \mathcal{F}$, dann gilt

$$\omega_1 \mapsto \mu_2((A^C)_{\omega_1}) = \mu_2(A_{\omega_1}^C) \stackrel{\text{endl. Maß}}{=} \underbrace{\mu_2(\Omega_2)}_{\text{messbar}} - \underbrace{\mu_2(A_{\omega_1})}_{\text{messbar}}.$$

Also gilt $A^C \in \mathcal{F}$ und damit ist \mathcal{F} abgeschlossen bezüglich Komplementbildung.

- Seien nun $A^1, A^2, \dots \in \mathcal{F}$ paarweise disjunkt, dann gilt

$$\begin{aligned} \omega_1 &\mapsto \mu_2\left(\left(\bigcup_{k=1}^{\infty} A^k\right)_{\omega_1}\right) \\ &= \mu_2\left(\bigcup_{k=1}^{\infty} A_{\omega_1}^k\right) \\ &\stackrel{\sigma\text{-add.}}{=} \underbrace{\sum_{k=1}^{\infty} \underbrace{\mu(A_{\omega_1}^k)}_{\text{messbar}}}_{\text{messbar}} = \underbrace{\lim_{m \rightarrow \infty} \sum_{k=1}^m \underbrace{\mu(A_{\omega_1}^k)}_{\text{messbar}}}_{\text{messbar}}, \end{aligned}$$

weil Grenzwerte messbarer Abbildungen wieder messbar sind.

\mathcal{F} ist also ein Dynkin-System. Nun folgt wie immer beim Trick der guten Mengen

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \stackrel{\text{Def.}}{=} \sigma(\mathcal{S}) = d(\mathcal{S}) \subseteq d(\mathcal{F}) = \mathcal{F} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2,$$

wobei wir den Hauptsatz für Dynkin-Systeme (Satz 1.2.11) für \mathcal{S} genutzt haben. Also gilt wieder $\mathcal{F} = \mathcal{A}_1 \otimes \mathcal{A}_2$ und die Behauptung folgt.

Jetzt fehlt nur noch der Fall $\mu_2(\Omega_2) = \infty$. Sei dazu $(E_n^2) \subseteq \mathcal{A}_2$ mit $E_n^2 \uparrow \Omega$ und $\mu_2(E_n^2) < \infty$ für alle $n \in \mathbb{N}$. Wir definieren

$$\mu_2^n(A) = \mu_2(A \cap E_n^2), \quad n \in \mathbb{N},$$

wie wir schon mehrfach gemacht haben (siehe z. B. den Beweis von 1.2.13). Dann sind die μ_2^n endliche Maße auf $(\Omega_2, \mathcal{A}_2)$. Aus dem ersten Schritt folgt, dass $\omega_1 \mapsto \mu_2^n(A_{\omega_1})$ messbar ist. Weil wegen der Stetigkeit von Maßen

$$\lim_{n \rightarrow \infty} \mu_2^n(A_{\omega_1}) = \lim_{n \rightarrow \infty} \mu_2(A_{\omega_1} \cap E_n^2) = \mu_2(A_{\omega_1})$$

gilt, ist auch die Abbildung $\omega_1 \mapsto \mu_2(A_{\omega_1})$ als punktwiser Grenzwert von messbaren Funktionen messbar. Das war's! □

Natürlich kann man per Induktion von $n = 2$ auf $n \in \mathbb{N}$ schließen:

Korollar 3.5.3. ▶ Sind $(\Omega_i, \mathcal{A}_i, \mu_i)_{i=1, \dots, n}$ σ -endliche Maßräume, so existiert genau ein Maß $\mu_1 \otimes \dots \otimes \mu_n$ auf der n -fachen Produkt- σ -Algebra

$$\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n := \sigma(\{A_1 \times \dots \times A_n : A_i \in \mathcal{A}_i\})$$

auf $\Omega_1 \times \dots \times \Omega_n$ mit

$$\mu_1 \otimes \dots \otimes \mu_n(A_1 \times \dots \times A_n) = \mu_1(A_1) \cdot \dots \cdot \mu_n(A_n)$$

für alle Mengen $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$.

Beweis. Induktion. □

Definition 3.5.4. ▶ Sind alle $(\Omega_i, \mathcal{A}_i, \mu_i)$ identisch, so schreibt man $\mu^{\otimes n}$ statt $\mu \otimes \dots \otimes \mu$. $\mu^{\otimes n}$ heißt dann n -faches **Produktmaß** von μ .

Vorlesung 16

Satz 3.5.5. ▶ [Satz von Fubini für $f \geq 0$]

Seien $(\Omega_1, \mathcal{A}_1, \mu_1), (\Omega_2, \mathcal{A}_2, \mu_2)$ σ -endliche Maßräume und $f: \Omega_1 \times \Omega_2 \rightarrow [0, +\infty]$ sei $(\mathcal{A}_1 \otimes \mathcal{A}_2, \mathcal{B}(\mathbb{R}))$ -messbar. Dann gelten:

(i) Die Integranden nach einer Variablen sind messbar, d. h.

$$\begin{aligned}\omega_2 &\mapsto f_{\omega_1}(\omega_2) := f(\omega_1, \omega_2) \text{ ist } (\mathcal{A}_2, \mathcal{B}(\overline{\mathbb{R}}))\text{-messbar f\"ur alle } \omega_1 \in \Omega_1 \text{ fest,} \\ \omega_1 &\mapsto f_{\omega_2}(\omega_1) := f(\omega_1, \omega_2) \text{ ist } (\mathcal{A}_1, \mathcal{B}(\overline{\mathbb{R}}))\text{-messbar f\"ur alle } \omega_2 \in \Omega_2 \text{ fest.}\end{aligned}$$

(ii) Die Integralfunktionen nach einer Variablen sind messbar, d. h.

$$\begin{aligned}\omega_2 &\mapsto \int_{\Omega_1} f_{\omega_2}(\omega_1) \, d\mu_1(\omega_1) \text{ ist } (\mathcal{A}_2, \mathcal{B}(\overline{\mathbb{R}}))\text{-messbar,} \\ \omega_1 &\mapsto \int_{\Omega_2} f_{\omega_1}(\omega_2) \, d\mu_2(\omega_2) \text{ ist } (\mathcal{A}_1, \mathcal{B}(\overline{\mathbb{R}}))\text{-messbar.}\end{aligned}$$

(iii) Es gilt Fubini und der Fubini-Flip:

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \, d\mu_1 \otimes \mu_2 &\stackrel{\text{Fubini}}{=} \int_{\Omega_1} \left(\int_{\Omega_2} f_{\omega_1}(\omega_2) \, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\ &\stackrel{\text{Fubini-Flip}}{=} \int_{\Omega_2} \left(\int_{\Omega_1} f_{\omega_2}(\omega_1) \, d\mu_1(\omega_1) \right) d\mu_2(\omega_2)\end{aligned}$$

Ganz wichtig: (i) und (ii) besagen lediglich, dass alle Integrale in der wesentlichen Aussage (iii) Sinn machen. In (iii) kann auch auf allen Seiten der Gleichheiten $+\infty$ stehen.

Beweis. Weil f messbar ist, existiert eine Folge $(f^n)_{n \in \mathbb{N}}$ einfacher Funktionen, die punktweise (also f\"ur alle (ω_1, ω_2)), gegen f wachsen. Dann gilt auch punktweise $f_{\omega_1}^n \uparrow f_{\omega_1}$, $f_{\omega_2}^n \uparrow f_{\omega_2}$, $n \rightarrow \infty$, weil dort einfach eine Variable festgehalten wird. Da $(f_{\omega_1}^n)_{n \in \mathbb{N}}$, $(f_{\omega_2}^n)_{n \in \mathbb{N}}$ Folgen einfacher Funktionen (in jeweils einer Koordinate) sind, sind f_{ω_1} , f_{ω_2} als punktweise Grenzwerte messbarer Funktionen auch messbar. Also gilt (i). Weil mit monotoner Konvergenz f\"ur alle $\omega_1 \in \Omega_1$

$$\lim_{n \rightarrow \infty} \int_{\Omega_2} f_{\omega_1}^n(\omega_2) \, d\mu_2(\omega_2) \stackrel{3.2.1}{=} \int_{\Omega_2} f_{\omega_1}(\omega_2) \, d\mu_2(\omega_2)$$

gilt, ist

$$\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}(\omega_2) \, d\mu_2(\omega_2)$$

als punktweiser Grenzwert messbarer Funktionen auch messbar. Man beachte dabei, dass

$$\omega_1 \mapsto \int_{\Omega_2} f_{\omega_1}^n(\omega_2) \, d\mu_2(\omega_2)$$

eine endliche Linearkombination von Funktionen der Form $\omega_1 \mapsto \mu_2(A_{\omega_1})$ ist. Diese Abbildungen sind, wie im Beweis von 3.5.3 gezeigt, messbar. Damit gilt also auch (ii).

Nun zur eigentlichen Aussage, (iii). Nicht \"uberraschend, erst f\"ur Indikatoren, dann die Gebetsm\"uhle. Sei also zun\"achst $f = \mathbf{1}_A$ f\"ur ein $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Dann gilt aufgrund der expliziten Konstruktion des Produktma\"a\"es

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f \, d\mu_1 \otimes \mu_2 &\stackrel{\text{Def. Int.}}{=} \mu_1 \otimes \mu_2(A) \\ &\stackrel{\text{Produktma\"a\"}}{=} \int_{\Omega_1} \mu_2(A_{\omega_1}) \, d\mu_1(\omega_1) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} f_{\omega_1}(\omega_2) \, d\mu_2(\omega_2) \right) d\mu_1(\omega_1),\end{aligned}$$

weil $f_{\omega_1} = \mathbf{1}_{A_{\omega_1}}$ gilt. Genau analog ergibt sich

$$\int_{\Omega_1 \times \Omega_2} f \, d\mu_1 \otimes \mu_2 = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \, d\mu_1(\omega_1) \right) d\mu_2(\omega_2),$$

weil wir am Anfang der Konstruktion des Produktmaßes auch schon die Identität gesehen haben:

$$\mu_1 \otimes \mu_2(A) = \int_{\Omega_2} \mu_1(A_{\omega_2}) d\mu_2(\omega_2).$$

Das hatten wir in dem Beweis $\bar{\mu}$ genannt. Damit haben wir beide Gleichheiten in (iii) für Indikatorfunktionen $f = \mathbf{1}_A$ bewiesen. Nun noch durch die Gebetsmühle der Integrationstheorie: Für einfache Funktionen folgt (iii) durch Linearität des Integrals und monotone Konvergenz folgt die Aussage auch für alle messbaren $f \geq 0$. \square

Warnung 3.5.6.  Meistens wird nur der „Fubini-flip“ genutzt, also die zweite Gleichheit. Die $(\mathcal{A}_1 \otimes \mathcal{A}_2, \mathcal{B}(\overline{\mathbb{R}}))$ -Messbarkeit in beiden Variablen muss für f trotzdem gecheckt werden. Sorry. Das ist ein wenig wie in Analysis 2: Für Stetigkeit einer Abbildung $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ reichte es auch nicht, partielle Stetigkeit in den einzelnen Koordinaten zu checken.

Merkt euch trotzdem folgende ganz grobe Behauptung: „Fubini funktioniert immer“. Mathematisch ist das natürlich nicht korrekt, man muss schließlich Voraussetzungen prüfen. Im Gegensatz zu monotoner oder dominierter Konvergenz ist das aber eigentlich nie ein Problem. Integrale drehen ist selten ein Problem, Grenzwerte in Integrale reinziehen ist oft schwierig!

Beispiel 3.5.7. 

- (i) Der Satz von Fubini aus Analysis 2 ist gerade der Spezialfall für $\mu_1 = \mu_2 = \lambda$ und liest sich als

$$\int_{\mathbb{R}^2} f(x_1, x_2) d(x_1, x_2) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, x_2) dx_1 \right) dx_2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x_1, x_2) dx_2 \right) dx_1$$

weil das zweidimensionale Lebesgue-Maß gerade das Produktmaß $\lambda \otimes \lambda$ ist. Wie immer schreiben wir lieber dx (bzw. $d(x_1, x_2)$) statt $d\lambda$, das ist aber nur eine Notationsfrage!

- (ii) Als Übungsaufgabe zeigt ihr, dass wir bei Doppelreihen die Reihenfolge immer ändern dürfen, wenn alle Koeffizienten nicht-negativ sind:

$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} a_{k,n} = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} a_{k,n}.$$

Das ist tatsächlich nur Fubini weil Reihen gerade Integrale mit Zählmaßen sind, vergleiche Beispiel 3.1.12. Vermutlich kennt ihr das Drehen von Reihen schon aus der Analysis 1, jetzt nochmal mit Fubini.

In den meisten Anwendungen reicht Fubini für nicht-negative messbare Funktionen, darum haben wir folgende Verallgemeinerung in der Vorlesung weggelassen. Zur Vollständigkeit wollen wir aber wieder durch Zerlegung in Positiv- und Negativteil eine Variante von Fubini für reell-wertige Integranden formulieren:

Satz 3.5.8. (Fubini für allgemeines integrierbares f)

Unter den selben Voraussetzungen von Satz 3.5.5 sei nun $f: \Omega_1 \times \Omega_2 \rightarrow \overline{\mathbb{R}}$ integrierbar. Dann gelten die Aussagen (i), (ii) und (iii).

Beweis. Wir schreiben $f = f^+ - f^-$, und wenden auf $f^+ \geq 0$ und $f^- \geq 0$ Satz 3.5.5 an. Die Messbarkeitsaussagen folgen direkt aus der Messbarkeit von Differenzen messbarer Abbildungen, Aussage (iii) folgt durch Zerlegung der Integrale. Die erste Gleichheit von (iii) zeigen wir wie

folgt:

$$\begin{aligned}
\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) \, d\mu_1 \otimes \mu_2(\omega_1, \omega_2) &\stackrel{\text{Def.}}{=} \int_{\Omega_1 \times \Omega_2} f^+(\omega_1, \omega_2) \, d\mu_1 \otimes \mu_2(\omega_1, \omega_2) \\
&\quad - \int_{\Omega_1 \times \Omega_2} f^-(\omega_1, \omega_2) \, d\mu_1 \otimes \mu_2(\omega_1, \omega_2) \\
&\stackrel{2 \times \text{Fubini}}{=} \int_{\Omega_1} \left(\int_{\Omega_2} f^+(\omega_1, \omega_2) \, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\
&\quad - \int_{\Omega_1} \left(\int_{\Omega_2} f^-(\omega_1, \omega_2) \, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\
&\stackrel{2 \times \text{Lin.}}{=} \int_{\Omega_1} \left(\int_{\Omega_2} (f^+ - f^-)(\omega_1, \omega_2) \, d\mu_2(\omega_2) \right) d\mu_1(\omega_1) \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \, d\mu_2(\omega_2) \right) d\mu_1(\omega_1).
\end{aligned}$$

Die zweite Gleichheit von (iii) folgt genauso aus Satz 3.5.5.

□

Kapitel 4

Stochastik

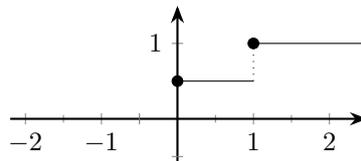
An dieser Stelle beenden wir die allgemeine Maß- und Integrationstheorie und wenden uns endlich vollständig der Stochastik zu. Wir haben bereits zufällige Experimente durch Wahrscheinlichkeitsräume modelliert, die Verteilung „einer Masse Zufall“ auf die reellen Zahlen durch Verteilungsfunktionen diskutiert und die Konzentration des Zufalls unter \mathbb{P}_F abgeschätzt. Jetzt kommt noch ein letzter Modellierungsschritt, wir wenden uns sogenannten Zufallsvariablen zu.

4.1 Zufallsvariablen

Diskussion 4.1.1. ▶ Schauen wir uns nochmal die Modellierung sehr einfacher zufälliger Experimente an. Beispielsweise für den Münzwurf haben wir zwei Modellierungsvarianten diskutiert: Variante 1: $\Omega = \{\text{Kopf}, \text{Zahl}\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ und das Maß

$$\mathbb{P}(\{\text{Kopf}\}) = \mathbb{P}(\{\text{Zahl}\}) = \frac{1}{2}, \quad \mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\emptyset) = 0.$$

Variante 2: $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mathbb{P}_F$, wobei \mathbb{P}_F das Wahrscheinlichkeitsmaß mit diskreter Verteilungsfunktion F ist:



Beide Modelle modellieren ein Experiment, bei dem zwei Elementarereignisse jeweils mit Wahrscheinlichkeit $\frac{1}{2}$ auftreten. Das zweite Modell ist natürlich viel zu kompliziert, hat aber den Vorteil, dass wir reelle Zahlen bekommen. Der Vorteil ist, dass wir so zum Beispiel so etwas wie einen Mittelwert (heißt später Erwartungswert) definieren können. Da wir den Ereignissen die Werte 0 und 1 geben, hättet ihr ohne viel nachzudenken sicherlich als umgangssprachlichen Erwartungswert $\frac{1}{2}$ vorgeschlagen.

Was unterscheidet Variante 1 von Variante 2? In der ersten Variante haben wir nur das zufällige Experiment Würfeln modelliert in der zweiten Variante haben wir zwei Dinge auf einmal modelliert:

- Was passiert? (Welches zufällige Ereignis passiert beim Würfeln)
- Was wird ausgezahlt? (Was wird für Kopf bzw. Zahl ausgezahlt)

Wenn wir bei Variante 1 auch über Auszahlungen sprechen wollen, müssen wir die möglichen Elementarereignisse noch in Auszahlungen übersetzen. Dafür kommen messbare Abbildungen $X : \Omega \rightarrow \mathbb{R}$ ins Spiel.

Ab jetzt: Trenne in der Modellierung „Was passiert?“ (also Ereignisse und Wahrscheinlichkeiten) von „Was wird beobachtet/ausgezahlt?“.

Definition 4.1.2.  Ein **stochastisches Modell** besteht aus

- (i) einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$,
- (ii) einer $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ -messbaren Abbildung $X: \Omega \rightarrow \mathbb{R}$.

Dabei beschreibt (i) das zufällige Experiment, (ii) beschreibt die „Beobachtung/Auszahlung“. X wird auch **Zufallsvariable (ZV)** genannt. Eine konkrete Ausführung $X(\omega)$ nennt man auch **Realisierung** der Zufallsvariablen.

Die Übersetzung des Elementarereignisses ω in die Beobachtung $X(\omega)$ ist dabei deterministisch (nicht zufällig), der Zufall steckt ausschließlich in dem Auftreten von ω . Irgendjemand unbekanntes (die Physik, ein Computer, ein Gott, etc.) entscheidet über die Wahl des zufälligen ω , das wird dann in die Zufallsvariable X eingesetzt und wir beobachten den Wert $X(\omega)$. In dieser Vorlesung sprechen wir nicht weiter über die „Ausführung von Zufall“, wir modellieren Wahrscheinlichkeiten. Für die Ausführung von Zufall (wir sagen die Realisierung der Zufallsvariablen) verweisen wir auf Vorlesungen über Monte Carlo Methoden oder Philosophie.

Definition 4.1.3.  Sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$.

- (i) Die **Verteilung der Zufallsvariablen** X ist definiert als

$$\mathbb{P}_X(B) := \mathbb{P}(X \in B) \stackrel{\text{Notation}}{=} \mathbb{P}(\{\omega \in \Omega: X(\omega) \in B\}) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

Unabhängig von dem zugrundeliegenden Wahrscheinlichkeitsraum ist \mathbb{P}_X also ein Maß auf $\mathcal{B}(\mathbb{R})$ und zwar der push-forward von \mathbb{P} unter X .

- (ii) Die **Verteilungsfunktion der Zufallsvariablen** X ist definiert als

$$F_X(t) := \mathbb{P}_X((-\infty, t]) \stackrel{\text{Def.}}{=} \mathbb{P}(X \leq t), \quad t \in \mathbb{R}.$$

Wir schreiben $X \sim F_X$ und $X \sim \mathbb{P}_X$ und sagen „ X ist verteilt gemäß F_X bzw. X ist verteilt gemäß \mathbb{P}_X “.

Weil so viele Indizes natürlich nerven, werden wir immer $X \sim F$ schreiben, wenn wir meinen, dass X gemäß F verteilt ist. Wir sagen dann auch kurz „ X hat Verteilungsfunktion F “. Beachte: Aufgrund der Definition ist natürlich $\mathbb{P}_X = \mathbb{P}_{F_X}$, das werden wir immer mal wieder nutzen.

Wir haben uns schon in der Maßtheorie langsam an den Begriff $\mu(\{f \leq t\})$ als Abkürzung für $\mu(\{\omega : f(\omega) \leq t\})$ gewöhnen müssen. In der Stochastik gehen wir noch einen Schritt weiter und lassen die Klammern auch noch weg. Wir schreiben daher immer abkürzend

$$\mathbb{P}(X < t), \quad \mathbb{P}(X \in (a, b]), \quad \mathbb{P}(X = a)$$

und so weiter. Das liest sich als „Wahrscheinlichkeit, dass X kleiner als t ist“ auch ziemlich natürlich.

Definition 4.1.4.  Zufallsvariablen X und Y , möglicherweise auf verschiedenen Wahrscheinlichkeitsräumen, heißen **identisch verteilt**, falls $F_X = F_Y$ bzw. $\mathbb{P}_X = \mathbb{P}_Y$. Man schreibt dann $X \sim Y$. Haben zwei Zufallsvariablen die gleiche Verteilungsfunktion, so sehen wir sie als gleichwertig an.

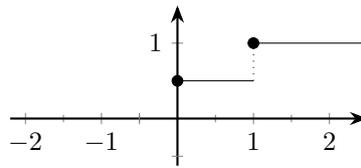
Zurück zum Münzwurf mit Auszahlung 1 für Zahl und 0 für Kopf. Wir schreiben dazu wieder zwei stochastische Modelle hin. Zum einen sei $\Omega = \{\text{Kopf}, \text{Zahl}\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ und

$$\mathbb{P}(\{\text{Kopf}\}) = \mathbb{P}(\{\text{Zahl}\}) = \frac{1}{2}, \quad \mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\emptyset) = 0,$$

mit Zufallsvariable (Auszahlungsfunktion) definiert als

$$X(\text{Kopf}) = 0, \quad X(\text{Zahl}) = 1.$$

Zum anderen sei $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ und $\mathbb{P} = \mathbb{P}_F$ mit der Verteilungsfunktion



Da wir diesmal die Auszahlung schon direkt im Modell modelliert haben, zahlen wir genau den Betrag aus, der zufällig gezogen wird. Das machen wir mit der Zufallsvariablen $Y(\omega) = \omega$. Berechnen wir nun die Verteilungen der Zufallsvariablen X und Y :

$$\begin{aligned} \mathbb{P}_X(B) &\stackrel{\text{Def.}}{=} \mathbb{P}(X \in B) = \begin{cases} \frac{1}{2} & : 0 \in B, 1 \notin B \text{ oder } 1 \in B, 0 \notin B \\ 1 & : 0, 1 \in B \\ 0 & : 0 \notin B, 1 \notin B \end{cases} \\ &= \frac{1}{2}\delta_0(B) + \frac{1}{2}\delta_1(B) \end{aligned}$$

sowie

$$\mathbb{P}_Y(B) \stackrel{\text{Def.}}{=} \mathbb{P}(Y \in B) = \mathbb{P}(\{\omega \in \mathbb{R} : Y(\omega) \in B\}) = \mathbb{P}(\{\omega \in \mathbb{R} : \omega \in B\}) = \mathbb{P}(B) \stackrel{\text{Def.}}{=} \mathbb{P}_F(B).$$

Weil $\mathbb{P}_F = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, gilt also $\mathbb{P}_X = \mathbb{P}_Y$ bzw. $F_X = F_Y$. Damit sind X und Y identische verteilte Zufallsvariablen und wir sehen die beiden stochastischen Modelle als gleichwertige Modelle für den Münzwurf an.

Vorlesung 17

Definition 4.1.5. ▶

- (i) Eine Zufallsvariable X heißt **absolutstetig** mit Dichte f , falls die Verteilungsfunktion F_X von X die Dichte f hat, es gilt also

$$\mathbb{P}(X \in (a, b]) = \mathbb{P}_X((a, b]) = F_X(b) - F_X(a) = \int_a^b f(x) dx.$$

- (ii) Eine Zufallsvariable X heißt **diskret**, falls F_X eine diskrete Verteilungsfunktion ist (oder \mathbb{P}_X ein diskretes Maß ist). In anderen Worten: X nimmt nur abzählbar viele Werte a_1, \dots, a_N mit Wahrscheinlichkeiten p_1, \dots, p_N an:

$$\mathbb{P}(X = a_k) = \mathbb{P}_X(\{a_k\}) = p_k.$$

Ihr merkt euch am besten folgende Rechenformeln:

$$\mathbb{P}(X \in A) = \begin{cases} \int_A f(x) dx & : X \text{ absolut stetig} \\ \sum_{a_k \in A} p_k & : X \text{ diskret} \end{cases},$$

man integriert oder summiert über die Werte, die X annehmen soll. Dumm ist nur, wenn wir nicht wissen, dass X absolutstetig oder diskret ist. Dann ist eine Zufallsvariable halt irgendeine reellwertige messbare Abbildung auf irgendeinem Wahrscheinlichkeitsraum mit irgendeiner Verteilungsfunktion F .

Beispiel 4.1.6. ▶ Wir kennen die meisten wichtigen Beispiele schon, manche kommen noch dazu:

Diskrete Zufallsvariablen	Absolutstetige Zufallsvariablen
$X \sim \text{Poi}(\lambda)$	$X \sim \mathcal{N}(\mu, \sigma^2)$
$X \sim \text{Bin}(n, p)$	$X \sim \text{Cauchy}(s, t)$
$X \sim \text{Ber}(p)$	$X \sim \text{Exp}(\lambda)$
$X \sim \text{Geo}(\lambda)$	$X \sim \mathcal{U}([a, b])$
$X \sim \text{Gleichverteilt auf endlichem } \Omega$	$X \sim \Gamma(\alpha, \beta)$
	$X \sim \text{Pareto}(k, a)$

Im Appendix gibt es eine Übersicht über die wichtigsten Verteilungen.

Um die Begriffe auszuprobieren, schauen wir uns eine kleine Rechnung an. Wir behaupten, dass $Y \sim \text{Exp}(1)$, wenn $Y = -\ln(U)$ mit $U \sim \mathcal{U}([0, 1])$. Probieren wir die Begriffe aus und berechnen definitionsgemäß die Verteilungsfunktion von Y durch Auflösen:

$$F_Y(t) = \mathbb{P}(Y \leq t) \stackrel{\text{Def.}}{=} \mathbb{P}(-\ln(U) \leq t) = \mathbb{P}(U \geq \exp(-t)) = 1 - \mathbb{P}(U \leq \exp(-t)).$$

Wenn wir jetzt die Verteilungsfunktion von $\mathcal{U}([0, 1])$ einsetzen, bekommen wir $F_Y(t) = (1 - e^{-t})\mathbf{1}_{[0, \infty)}(t)$ und das ist die Verteilungsfunktion der Exponentialverteilung.

Wir waren ein klein wenig ungenau weil Y den Wert $+\infty$ annehmen kann, eine Zufallsvariable per Definition aber nur reelle Werte annimmt. Das kann man einfach reparieren, indem man zum Beispiel $\log(0) = 0$ undefiniert. Alternativ nimmt man $U \sim \mathcal{U}((0, 1))$ statt $U \sim \mathcal{U}([0, 1])$, das diskutieren wir ausführlich in Abschnitt 4.3.1.

Satz 4.1.7.  [Existenz stochastischer Modelle - der „kanonische“ Wahrscheinlichkeitsraum]

Für jede Verteilungsfunktion F existiert eine Zufallsvariable X mit $X \sim F$. Genauer: Es existiert ein stochastisches Modell, d. h. ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und eine Zufallsvariable X auf $(\Omega, \mathcal{A}, \mathbb{P})$, mit $X \sim F$.

Beweis. Als Wahrscheinlichkeitsraum definieren wir $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mathbb{P}_F$ und darauf die Zufallsvariable $X(\omega) = \omega$. Beachte: $X(\omega) = \omega$ ist eine stetige Abbildung von \mathbb{R} nach \mathbb{R} und damit auch Borel-messbar. Berechnen wir die Verteilungsfunktion dieser konkreten Zufallsvariablen auf dem konkreten Wahrscheinlichkeitsraum:

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}_F(\{\omega : X(\omega) \leq t\}) = \mathbb{P}_F(\{\omega : \omega \leq t\}) = \mathbb{P}_F((-\infty, t]) = F(t).$$

Also gilt $X \sim F$, das war es schon! Zu beachten ist, dass die Konstruktion weit von trivial ist. Die Existenz von \mathbb{P}_F benötigt den Satz von Carathéodory und damit die komplette Maßtheorie. \square

Bemerkung 4.1.8.  Wenn wir uns nur für diskrete Zufallsvariablen interessieren würden, kämen wir komplett *ohne* Maßtheorie aus! Hier ist eine viel einfachere Konstruktion, die für alle diskrete Zufallsvariablen funktioniert. Sei dazu F eine diskrete Verteilungsfunktion, die an N -vielen Stellen a_k um p_k nach oben springt. Sei nun $\Omega = \{\omega_1, \dots, \omega_N\}$ eine beliebige Menge mit N Elementen (z.B. $\Omega = \{\text{Kopf}, \text{Zahl}\}$ beim Würfeln). Auf Ω wählen wir als σ -Algebra $\mathcal{A} = \mathcal{P}(\Omega)$. Das Maß definieren wir, indem wir es auf den Elementarereignissen als $\mathbb{P}(\{\omega_k\}) = p_k$ definieren und mit der σ -Additivität für beliebiges $A \in \mathcal{A}$ fortsetzen, d. h. $\mathbb{P}(A) = \sum_{\omega_k \in A} \mathbb{P}(\{\omega_k\}) = \sum_{\omega_k \in A} p_k$. An dieser Stelle nutzen wir die Abzählbarkeit. Als Zufallsvariable wählen wir die Abbildung $X(\omega_k) := a_k$. Weil auf dem Urbildraum die Potenzmenge gewählt wurde, ist natürlich jede Abbildung nach \mathbb{R} auch $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$ -messbar. Damit gilt $\mathbb{P}(X = a_k) = p_k$, also ist X gemäß F verteilt. Diese Konstruktion funktioniert nur für diskrete Verteilungsfunktionen so einfach (probiert es einfach mal für absolutstetige Verteilungsfunktionen aus, ihr werdet schnell den Fortsetzungssatz von Carathéodory brauchen). Im Allgemeinen kommen wir nicht umher, die Maßtheorie wie im Beweises von Satz 4.1.7 zu nutzen weil man Maße auf überabzählbaren Mengen nicht einfach auf den Elementarereignissen definieren kann.

Hier sind zwei konkrete Beispiele: Wenn jemand von euch ein stochastisches Modell für eine $\mathcal{N}(0, 1)$ -verteilte Zufallsvariable verlangt, so entgegnet ihr

$$\Omega = \mathbb{R}, \quad \mathcal{A} = \mathcal{B}(\mathbb{R}), \quad \mathbb{P} = \mathbb{P}_F, \quad X(\omega) = \omega,$$

wobei \mathbb{P}_F das eindeutige Maß auf $\mathcal{B}(\mathbb{R})$ mit Verteilungsfunktion $F \sim \mathcal{N}(0, 1)$ aus Carathéodory ist. Will jemand das gleiche für eine $\text{Poi}(\lambda)$ -verteilte Zufallsvariable haben, so entgegnet ihr entweder das gleiche, oder

$$\Omega = \mathbb{N}, \quad \mathcal{A} = \mathcal{P}(\mathbb{N}), \quad \mathbb{P}(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad X(\omega) = \omega.$$

Um ganz deutlich zu sein: Diskret ist einfach, weil die Potenzmenge von abzählbaren Mengen noch klein genug ist und Maße durch die abzählbare σ -Additivität von Maßen nur auf einpunktigen Mengen definiert werden müssen! Weil all das für überabzählbare Mengen wie \mathbb{R} schief geht, haben wir Maßtheorie gemacht. Auf die Normalverteilung können wir einfach nicht verzichten!

Definition 4.1.9. ▶ Ist X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$, so heißen, falls die Integrale wohldefiniert sind,

(i)

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega) \quad \text{Erwartungswert von } X,$$

(ii)

$$\mathbb{E}[X^k] := \int_{\Omega} X^k(\omega) \, d\mathbb{P}(\omega) \quad k\text{-tes Moment von } X,$$

(iii)

$$\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] \quad \text{Varianz von } X,$$

(iv)

$$\mathbb{E}[e^{\lambda X}] := \int_{\Omega} e^{\lambda X(\omega)} \, d\mathbb{P}(\omega) \quad \text{exponentielles Moment von } X.$$

Allgemein betrachten wir für $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ messbar die Erwartungswerte

$$\mathbb{E}[g(X)] := \int_{\Omega} g(X(\omega)) \, d\mathbb{P}(\omega),$$

falls das Integral wohldefiniert ist. Wir sagen die Erwartungswerte existieren, falls die Integrale existieren (also endlich sind).

Wir erinnern daran, dass ein wohldefiniertes Integral auch die Werte $+\infty$ oder $-\infty$ annehmen darf. Meistens werden wir aber davon sprechen, dass die Integrale existieren, also endlich sind. Wegen der allgemeinen Äquivalenzen

$$\int f \, d\mu \text{ existiert} \stackrel{\text{Def.}}{\Leftrightarrow} \int f^+ \, d\mu < \infty, \int f^- \, d\mu < \infty \Leftrightarrow \int |f| \, d\mu < \infty,$$

schreiben wir meistens bequemer „ $\mathbb{E}[|g(X)|] < \infty$ “ anstelle von „ $\mathbb{E}[g(X)]$ existiert“. Bei Erwartungswerten sagen wir also „der Erwartungswert von X existiert“, wenn der Erwartungswert wohldefiniert und endlich ist. Das wird oft in der Literatur genauso gehandhabt, manchmal aber auch anders. Manche sagen, der Erwartungswert existiert, wenn $\mathbb{E}[X]$ wohldefiniert ist (aber vielleicht unendlich) ist.

Bemerkung. Die Notation $\mathbb{E}[g(X)]$ ist etwas unglücklich weil das Integral nicht nur von g und X abhängt, sondern auch von \mathbb{P} . Daher sollte man eher $\mathbb{E}_{\mathbb{P}}[g(X)]$ schreiben, man lässt aber üblicherweise das \mathbb{P} aus Faulheit weg.

Lemma 4.1.10. ▶ Ist die Zufallsvariable X gemäß F verteilt, d. h. $X \sim F$, so gilt unabhängig von dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ auf dem X definiert ist,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) \, d\mathbb{P}_F(x)$$

für messbare numerische Abbildungen $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$. Wie immer gilt die Gleichheit, wenn eine Seite (und damit die die andere Seite) wohldefiniert ist.

Beweis. Mit dem Transformationssatz 3.1.16 (bzw. Korollar (3.1.17)) gilt in einem Schaubild

$$\begin{array}{ccc}
 (\Omega, \mathcal{A}, \mathbb{P}) & \xrightarrow{X} & (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X) \\
 & \searrow^{g \circ X} & \downarrow g \\
 & & (\overline{\mathbb{R}}, \overline{\mathcal{B}}(\overline{\mathbb{R}}))
 \end{array}$$

wobei \mathbb{P}_X der push-forward von X ist. Weil definitionsgemäß $\mathbb{P}_X = \mathbb{P}_F$ ist, gibt das sauber ausgeschrieben

$$\mathbb{E}[g(X)] \stackrel{\text{Def.}}{=} \int_{\Omega} g(X(\omega)) \, d\mathbb{P}(\omega) \stackrel{3.1.16}{=} \int_{\mathbb{R}} g(x) \, d\mathbb{P}_X(x) \stackrel{\mathbb{P}_X = \mathbb{P}_F}{=} \int_{\mathbb{R}} g(x) \, d\mathbb{P}_F(x).$$

□

Die Konsequenz ist natürlich, dass für beliebiges g , $\mathbb{E}[g(X)]$ gar nicht von dem kompletten Modell $(\Omega, \mathcal{A}, \mathbb{P}, X)$ abhängt, $\mathbb{E}[g(X)]$ hängt einfach nur von der Verteilung von X ab. Das ist der Grund, warum man sich üblicherweise nur für die Verteilungsfunktion von X , jedoch nicht für $(\Omega, \mathcal{A}, \mathbb{P})$ interessiert. Fassen wir die Beobachtung zusammen:

Bemerkung. Sind X, Y identisch verteilte Zufallsvariablen, die auf irgendwelchen Wahrscheinlichkeitsräumen definiert sind, so gilt

$$\mathbb{E}[g(X)] = \mathbb{E}[g(Y)]$$

für alle $g: \mathbb{R} \rightarrow \overline{\mathbb{R}}$ messbar.

Jetzt wissen wir auch schon, wie wir $\mathbb{E}[g(X)]$ berechnen können, das haben wir nämlich schon gemacht. Vieles was jetzt kommt sind Wiederholungen, indem wir Sätze für allgemeine Integrale nochmal für Erwartungswerte hinschreiben.

Satz 4.1.11.  [Berechnungsregeln]

Sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und $g: \mathbb{R} \rightarrow \overline{\mathbb{R}}$ messbar, so gelten:

- (i) Ist X absolutstetig mit Dichte f , so gilt

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) \, dx.$$

- (ii) Ist X diskret und nimmt die Werte $a_1, \dots, a_N \in \mathbb{R}$ mit Wahrscheinlichkeiten p_1, \dots, p_N an, so gilt

$$\mathbb{E}[g(X)] = \sum_{k=1}^N p_k g(a_k) = \sum_{k=1}^N \mathbb{P}(X = a_k) g(a_k).$$

Dabei ist wieder eine Seite genau dann wohldefiniert, wenn es die andere Seite ist.

Beweis. Dazu kombinieren wir nur die Formel aus Satz 4.1.10 mit den Formeln aus Satz 3.3.2 und Satz 3.3.3. □

Beispiel 4.1.12. 

- Für $X \sim \mathcal{N}(\mu, \sigma^2)$ gilt $\mathbb{E}[X] = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \mu$.
- Für $X \sim \text{Ber}(p)$ gilt $\mathbb{E}[X] = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p$.
- Für $X \sim \text{Poi}(\lambda)$ gilt $\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda$.

Wir sehen also: Ein großer Teil der Stochastik besteht aus dem Berechnen von Integralen und Summen bzw. Reihen.

Vorlesung 18

Beispiel 4.1.13.  Eine Webseite wird im Mittel pro Stunde zweimal geklickt. Wie groß ist die Wahrscheinlichkeit, dass die Webseite in einer Stunde mindestens fünfmal geklickt wird?

Um das Beispiel stochastisch zu behandeln, müssen wir zunächst ein Modell annehmen. Welche uns bekannte Zufallsvariable könnte die Anzahl der Klicks pro Stunde modellieren? Da die Ergebnisse der Zufallszahlen natürliche Zahlen sind, muss die Verteilung diskret sein. Nun ist die Anzahl nicht beschränkt, es sollte also eine Zufallsvariable sein, die alle Werte in \mathbb{N} annehmen kann. Dazu kennen wir bisher nur die geometrische- oder die Poissonverteilung. An der jetzigen Stelle können wir ohne weitere Annahmen keine von beiden ausschließen. Nehmen wir einfach mal an, dass $X \sim \text{Poi}(\lambda)$ ein gutes Modell ist. Aber was ist λ ? Weil wir wissen, dass $\mathbb{E}[X] = \lambda$ ist, schließen wir $\lambda = 2$ aus der Vorinformation (Das Gesetz der großen Zahlen wird uns in Satz 4.6.8 des letzten Kapitel die Rechtfertigung geben, warum wir aus dem Mittel auf den Erwartungswert schließen.) Um nun die Aufgabe zu lösen, müssen wir für eine $\text{Poi}(2)$ -verteilte Zufallsvariable $\mathbb{P}(X \geq 5)$ berechnen. Das geht ganz einfach:

$$\begin{aligned} \mathbb{P}(X \geq 5) &= 1 - \mathbb{P}(X \leq 4) \\ &= 1 - \sum_{k=0}^4 \mathbb{P}(X = k) \\ &= 1 - e^{-2} \left(\frac{1}{1} + \frac{2}{1} + \frac{4}{2} + \frac{8}{6} + \frac{16}{24} \right) \approx 0,053. \end{aligned}$$

Hierbei haben wir genutzt, dass für eine diskrete Zufallsvariable immer

$$\mathbb{P}(X \in A) = \sum_{a_k \in A} \mathbb{P}(X = a_k)$$

gilt. Das folgt natürlich aus der σ -Additivität von Maßen weil $A \mapsto \mathbb{P}(X \in A)$ ein Maß ist (die Verteilung von X).

Proposition 4.1.14.  [Rechenregeln für den Erwartungswert]

Seien X, Y Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\mathbb{E}[|X|], \mathbb{E}[|Y|] < \infty$ und $\alpha, \beta \in \mathbb{R}$, so gelten:

- (i) $\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$
- (ii) $X \geq 0$ \mathbb{P} -f.s. $\Rightarrow \mathbb{E}[X] \geq 0$ und $X \geq Y$ \mathbb{P} -f.s. $\Rightarrow \mathbb{E}[X] \geq \mathbb{E}[Y]$
- (iii) Ist $X = \alpha$ \mathbb{P} -f.s., so ist $\mathbb{E}[X] = \alpha$.
- (iv) $\mathbb{P}(X \in A) = \mathbb{E}[\mathbf{1}_A(X)]$, insbesondere gilt $F_X(t) = \mathbb{E}[\mathbf{1}_{(-\infty, t]}(X)]$, $t \in \mathbb{R}$.

Beweis. Wir müssen nur beachten, dass Erwartungswerte per Definition Integrale sind. Dann können wir die Rechenregeln für Integrale direkt anwenden. Zu beachten ist, dass wegen Satz 3.1.15 Änderungen auf Nullmengen Integrale nicht ändern.

- (i) Linearität von Integralen
- (ii) Monotonie von Integralen (die Nullmengen spielen keine Rolle)
- (iii) Nach Annahme gilt $X = \alpha \mathbf{1}_\Omega$ \mathbb{P} -f.s. Wegen Satz 3.1.15 können wir sofort die Definition des Integrals einsetzen:

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\Omega} \underbrace{\alpha \mathbf{1}_{\Omega}}_{\text{einfach}} d\mathbb{P} \stackrel{\text{Def.}}{=} \alpha \mathbb{P}(\Omega) = \alpha$$

(iv) Hier müssen wir nur die Definitionen im Kopf klar bekommen:

$$\mathbb{E}[\mathbf{1}_A(X)] = \int_{\Omega} \mathbf{1}_A(X(\omega)) d\mathbb{P}(\omega) \stackrel{\text{Trafo}}{=} \int_{\mathbb{R}} \underbrace{\mathbf{1}_A(x)}_{\text{einfach}} d\mathbb{P}_X(x) = \mathbb{P}_X(A) \stackrel{\text{Def.}}{=} \mathbb{P}(X \in A)$$

□

Natürlich ist eine Zufallsvariable, die fast sicher den selben Wert annimmt, gar keine interessante Zufallsvariable! Das modellierte Zufallsexperiment ist gar nicht zufällig, es passiert immer das gleiche! Beispiel: Jeden Tag um 7 Uhr wird die Zeit (Stunde) angeschaut. Es kommt immer 7 dabei raus, die beschreibende Zufallsvariable erfüllt also $\mathbb{P}(X = 7) = 1$, also $X = 7$ \mathbb{P} -fast sicher. Viel interessanter wäre zum Beispiel, jeden Tag um 7 Uhr die Temperatur zu messen. Die entsprechende Zufallsvariable wäre nicht fast sicher konstant.

Korollar 4.1.15.  **[Rechenregeln für die Varianz]**

Sei X eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$ (wir sagen auch, X ist quadratintegrierbar), so gelten:

(i) Es gilt $\mathbb{V}[X] < \infty$ und

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

(ii) Es gilt $\mathbb{V}[X] = 0$ genau dann, wenn X fast sicher den gleichen Wert annimmt, dieser ist dann $\mathbb{E}[X]$.

(iii) Für $a \in \mathbb{R}$ gilt $\mathbb{V}[aX] = a^2\mathbb{V}[X]$ und $\mathbb{V}[a + X] = \mathbb{V}[X]$.

Beweis. Übung. Beachte dazu: Ist $Y \geq 0$ \mathbb{P} -fast sicher und $\mathbb{E}[Y] = 0$, so ist $Y \equiv 0$ \mathbb{P} -fast sicher. Das gilt wegen Satz 3.1.15, der Erwartungswert ist schließlich ein Integral! □

Die Varianz misst als Kenngröße die Variabilität von Zufallsvariablen. Ist die Varianz null, so gibt es gar keine Variabilität, ist die Varianz groß, so nimmt X auch Werte an, die weit von dem Erwartungswert entfernt sind. Das ist mathematisch natürlich keine saubere Formulierung (dafür gibt es schließlich die Varianz), gibt aber das richtige Gefühl. Daher ist auch einleuchtend, dass sich die Variabilität beim Verschieben um einen festen Wert sich nicht ändert.

Satz 4.1.16.  **[Konvergenzsätze für Zufallsvariablen]**

Seien Y, X, X_1, X_2, \dots Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\lim_{n \rightarrow \infty} X_n = X$ \mathbb{P} -fast sicher.

(i) MCT: Gilt $0 \leq X_1 \leq X_2 \leq \dots \leq X$ \mathbb{P} -fast sicher, so gilt

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

(ii) DCT: Gilt $|X_n| \leq Y$ \mathbb{P} -fast sicher für alle $n \in \mathbb{N}$ mit $\mathbb{E}[|Y|] < \infty$, so gilt

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

(iii) Gilt $|X_n| \leq C$ \mathbb{P} -fast sicher für alle $n \in \mathbb{N}$ für ein $C > 0$, so gilt

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Beweis. Wegen $\mathbb{E}[X_n] \stackrel{\text{Def.}}{=} \int_{\Omega} X_n d\mathbb{P}$ und $\mathbb{E}[X] \stackrel{\text{Def.}}{=} \int_{\Omega} X d\mathbb{P}$ ist das gerade Satz 3.2.1, Satz 3.2.5 und Korollar 3.2.6. □

Definition 4.1.17.  Für eine Zufallsvariable X heißt

$$\mathcal{M}_X(t) := \mathbb{E}[e^{tX}], \quad t \in \mathbb{R},$$

die **momenterzeugende Funktion**. \mathcal{M}_X ist nur für die t definiert, für die $\mathbb{E}[e^{tX}] < \infty$ gilt.

Die momenterzeugenden Funktionen sind auf ihrem Definitionsbereich ganz normale Funktionen von \mathbb{R} nach \mathbb{R} . Wir können also über Ableitungen sprechen, Monotonie, und so weiter. In vielen Beispielen ist M_X eine ganz harmlose Funktion, manchmal ist \mathcal{M}_X aber auch gar nicht definiert.

Beispiel 4.1.18. 

- Sei $X \sim \text{Cauchy}(s, t)$, so ist \mathcal{M}_X nirgends definiert!
- Sei $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $\mathcal{M}_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$ für $t \in \mathbb{R}$, siehe Übungsaufgabe.
- Sei $X \sim \text{Poi}(\lambda)$, so ist

$$\mathcal{M}_X(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{\infty} e^{tX} \mathbb{P}(X = k) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{\lambda(e^t - 1)}$$

für alle $t \in \mathbb{R}$.

Noch viel mehr explizite Beispiele sind im Appendix gesammelt.

Das ganze ist ein so nützliches Konzept, weil wir viele Beispiele explizit ausrechnen können und mit dem nächsten Satz gleich noch alle Momente durch Ableiten ausrechnen können:

Satz 4.1.19.  Sei X eine Zufallsvariable, für die \mathcal{M}_X für ein $\epsilon > 0$ in $(-\epsilon, \epsilon)$ definiert ist. Dann ist \mathcal{M}_X an der Stelle 0 unendlich oft differenzierbar und es gilt für alle $n \in \mathbb{N}$

$$\mathbb{E}[X^n] = \mathcal{M}_X^{(n)}(0),$$

wobei $\mathcal{M}_X^{(n)}(0)$ die n -te Ableitung an der Stelle 0 ist.

Beweis.

- (a) Wir zeigen zunächst, dass \mathcal{M}_X in $(-\epsilon, \epsilon)$ eine Potenzreihe ist. Nach Analysis 1 ist

$$e^{tX(\omega)} = \sum_{k=0}^{\infty} \frac{(tX(\omega))^k}{k!} = \lim_{m \rightarrow \infty} \sum_{k=0}^m \frac{(tX(\omega))^k}{k!} =: \lim_{m \rightarrow \infty} S_m(\omega).$$

Die Zufallsvariable e^{tX} kann also als punktwiser Grenzwert der Folge $(S_m)_{m \in \mathbb{N}}$ von Zufallsvariablen geschrieben werden. Um gleich dominierte Konvergenz zu nutzen, brauchen wir eine integrierbare Majorante S für die Folge (S_m) . Das ist gar nicht so schwer:

$$|S_m| \stackrel{\text{Def.}}{=} \left| \sum_{k=0}^m \frac{(tX)^k}{k!} \right| \triangleq \sum_{k=0}^m \left| \frac{(tX)^k}{k!} \right| \leq \sum_{k=0}^{\infty} \left| \frac{(tX)^k}{k!} \right| =: S.$$

Wegen $S = e^{|tX|} \leq e^{tX} + e^{-tX}$ ist S eine integrierbare Majorante:

$$\mathbb{E}[|S|] = \mathbb{E}[e^{|tX|}] \stackrel{\text{Mon.}}{\leq} \mathbb{E}[e^{tX}] + \mathbb{E}[e^{-tX}] \stackrel{\text{Def.}}{=} \mathcal{M}_X(t) + \mathcal{M}_X(-t) < \infty$$

für $t \in (-\epsilon, \epsilon)$ nach Annahme. Jetzt kann also dominierte Konvergenz angewandt werden. Es gilt damit für $t \in (-\epsilon, \epsilon)$, dass

$$M_X(t) = \mathbb{E}\left[\lim_{m \rightarrow \infty} S_m\right] \stackrel{\text{DCT}}{=} \lim_{m \rightarrow \infty} \mathbb{E}[S_m] \stackrel{\text{Lin.}}{=} \lim_{m \rightarrow \infty} \sum_{k=0}^m \frac{t^k \mathbb{E}[X^k]}{k!} = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}[X^k]}{k!}.$$

Damit ist M_X in $(-\epsilon, \epsilon)$ eine Potenzreihe mit Koeffizienten $a_k = \frac{\mathbb{E}[X^k]}{k!}$ und Entwicklungspunkt $x_0 = 0$.

- (b) Aus Analysis 1 wisst ihr (so habt ihr die Taylor-Koeffizienten bestimmt!), dass M_X in $(-\epsilon, \epsilon)$ unendlich oft differenzierbar ist und die Reihe gliedweise differenziert werden kann. n -faches Ableiten und den Entwicklungspunkt $x_0 = 0$, einsetzen gibt dann $M_X^{(n)}(0) = \mathbb{E}[X^n]$.

□

Beispiel 4.1.20. 

- Für die Normalverteilungen $X \sim \mathcal{N}(\mu, \sigma^2)$ ergibt die explizite Formel der momenterzeugenden Funktion mit dem Satz

$$\begin{aligned}\mathbb{E}[X] &= \mathcal{M}'_X(0) = \exp\left(\mu \cdot 0 + \frac{\sigma^2 t^2}{2}\right) \cdot (\mu + \sigma^2 \cdot 0) = \mu, \\ \mathbb{E}[X^2] &= \mathcal{M}''_X(0) = \mu^2 + \sigma^2.\end{aligned}$$

Beides zusammen ergibt $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$. Die zwei Parameter der Normalverteilung sind also gerade Erwartungswert μ und Varianz σ^2 .

- Für $X \sim \text{Poi}(\lambda)$ ergibt die explizite Formel der momenterzeugenden Funktion mit dem Satz

$$\mathbb{E}[X] = \mathcal{M}'_X(0) = \lambda \quad \text{und} \quad \mathbb{E}[X^2] = \mathcal{M}''_X(0) = \lambda^2 + \lambda.$$

Beides zusammen ergibt $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda$.

Proposition 4.1.21.  **[Hölder und Jensen'sche Ungleichung]**

- (i) Für $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$ und Zufallsvariablen X, Y auf $(\Omega, \mathcal{A}, \mathbb{P})$ gilt

$$\mathbb{E}[|XY|] \leq (\mathbb{E}[|X|^p])^{1/p} (\mathbb{E}[|Y|^q])^{1/q}.$$

- (ii) Ist X eine Zufallsvariable mit $\mathbb{E}[|X|] < \infty$ und $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ *konvex* mit $\mathbb{E}[|\varphi(X)|] < \infty$, so gilt

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

Beweis. (i) Weil Erwartungswerte Integrale sind, ist das nur ein Spezialfall von Satz 3.4.1.

(ii) Wir geben den Beweis nur für differenzierbares φ . Wegen der Konvexität gibt es für jedes feste $x_0 \in \mathbb{R}$ ein $b \in \mathbb{R}$ mit

- $\varphi'(x_0)x + b \leq \varphi(x)$ für alle $x \in \mathbb{R}$,
- $\varphi'(x_0)x_0 + b = \varphi(x_0)$.

Wir wählen $x_0 = \mathbb{E}[X]$. Mit der zweiten Eigenschaft schreiben wir $\varphi(\mathbb{E}[X])$ wie folgt um:

$$\varphi(\mathbb{E}[X]) = \varphi(x_0) = \varphi'(x_0)x_0 + b = \varphi'(x_0)\mathbb{E}[X] + b.$$

Weil der Erwartungswert linear und monoton ist, sowie der Erwartungswert einer konstanten Zufallsvariable gerade die Konstante ist, können wir die rechte Seite wie folgt behandeln:

$$\varphi'(x_0)\mathbb{E}[X] + b = \mathbb{E}[\varphi'(x_0)X] + \mathbb{E}[b] = \mathbb{E}[\varphi'(x_0)X + b] \stackrel{\text{Mon.}}{\leq} \mathbb{E}[\varphi(X)].$$

Zusammen folgt die Behauptung. □

Beispiel. Als Merkregel für das „ \leq “ in Proposition 4.1.21 nimmt man $\varphi(x) = x^2$. Weil

$$0 \leq \mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \stackrel{\text{Üb.}}{=} \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

muss $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$ gelten. Also muss in 4.1.21 „ \leq “ und nicht „ \geq “ stehen.

Zum Abschluss nochmal die Markov- und Tschebyscheff-Ungleichungen, die wir für Integrale über beliebige Maße schon angeschaut haben. Weil Erwartungswerte Integrale sind, geht das in diesem Spezialfall natürlich genauso:

Satz 4.1.22. ▶ [Markov- und Tschebyscheff-Ungleichung]

Sei X eine Zufallsvariable, dann gelten für $a > 0$ folgende Ungleichungen:

- (i) Für $h: \mathbb{R} \rightarrow (0, \infty)$ wachsend gilt

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[h(X)]}{h(a)} \quad (\text{Markov-Ungleichung})$$

- (ii) Für $h: [0, \infty) \rightarrow (0, \infty)$ wachsend gilt

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[h(|X|)]}{h(a)} \quad (\text{Markov-Ungleichung})$$

- (iii)

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}[X]}{a^2} \quad (\text{Tschebyscheff-Ungleichung})$$

Beweis.

- (i) Definiere $A = [a, \infty)$, dann gilt weil h wachsend ist

$$\begin{aligned} \mathbb{E}[h(X)] &\stackrel{\text{Mon.}}{\geq} \mathbb{E}[h(X) \cdot \mathbf{1}_A(X)] \\ &\geq \mathbb{E}[h(a) \cdot \mathbf{1}_A(X)] \\ &\stackrel{\text{Lin.}}{=} h(a) \cdot \mathbb{E}[\mathbf{1}_A(X)] \\ &\stackrel{4.1.14.(iv)}{=} h(a) \cdot \mathbb{P}(X \geq a). \end{aligned}$$

Durchteilen gibt die Abschätzung. Hier haben wir den kleinen Trick genutzt, dass $\mathbf{1} \equiv \mathbf{1}_\Omega \equiv \mathbf{1}_A + \mathbf{1}_{A^c} \geq \mathbf{1}_A$ gilt. Der Trick wird jetzt immer wieder kommen!

- (ii) Genau wie (i).
 (iii) Benutze die Markov Ungleichung mit $h(x) = x^2$ und der „zentrierten“ Zufallsvariablen $X - \mathbb{E}[X]$.

□

Wie bei den Konzentrationsungleichungen für Maße, können wir durch Bildung von Gegenwahrscheinlichkeiten sofort Ungleichungen für $\mathbb{P}(X < a)$ oder $\mathbb{P}(|X| < a)$ bekommen.

Beispiel. Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so gilt $\mathbb{P}(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$.

4.2 Zufallsvektoren

Nachdem Zufallsvariablen jetzt hoffentlich einigermaßen klar geworden sind, gehen wir jetzt weiter zu Zufallsvektoren. Das sind Zufallsvariablen, deren Werte nicht reell, sondern aus dem \mathbb{R}^d sind. Weil wir den Namen Zufallsvariablen nur für den Fall $d = 1$ definiert haben, sprechen wir nun von Zufallsvektoren. Das Kapitel ist in großen Teilen eine Wiederholung, wir gehen durch die gleichen Schritte, die Notationen werden nur einen Tick aufwendiger. Das Vorgehen ist genau wie für reelle (eindimensionale) Zufallsvariablen:

- Lege σ -Algebra auf \mathbb{R}^d fest und zeige wichtige Eigenschaften für später.

- Charakterisiere Maße auf \mathbb{R}^d durch Verteilungsfunktionen.
- Definiere Zufallsvektoren als messbare Abbildungen und verbinde diese zu Maßen und Verteilungsfunktionen.
- Definiere Erwartungswerte und zeige Rechenregeln für diskrete und absolutstetige Zufallsvektoren.
- Rechnen, rechnen, rechnen.

(A) Borel- σ -Algebra auf \mathbb{R}^d

Definition 4.2.1. ▶ Wir wählen die Produkt- σ -Algebra auf dem \mathbb{R}^d , die aus d Kopien der Borel- σ -Algebra besteht:

$$\mathcal{B}(\mathbb{R}^d) := \underbrace{\mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})}_{d\text{-viele}} \stackrel{\text{Def.}}{=} \sigma(\{B_1 \times \dots \times B_d : B_i \in \mathcal{B}(\mathbb{R})\})$$

Wir hatten im ersten Kapitel schon erwähnt, dass die Definition der Borel- σ -Algebra als kleinste σ -Algebra erzeugt durch offene Mengen auch im \mathbb{R}^d funktioniert. Das ist in der Tat das selbe wie die gerade definierte Produkt- σ -Algebra $\mathcal{B}(\mathbb{R}^d)$.

Lemma 4.2.2. ▶ Es gilt

$$\begin{aligned} \mathcal{B}(\mathbb{R}^d) &= \sigma(\{O \subseteq \mathbb{R}^d : O \text{ offen}\}) \\ &= \sigma(\{(-\infty, t_1] \times \dots \times (-\infty, t_d] : t_i \in \mathbb{R}\}) \\ &= \sigma(\{(a_1, b_1] \times \dots \times (a_d, b_d] : a_i, b_i \in \mathbb{R}\}) \\ &= \sigma(\{(a_1, b_1) \times \dots \times (a_d, b_d) : a_i, b_i \in \mathbb{R}\}) \\ &= \dots, \end{aligned}$$

wobei ... bedeutet, dass ihr wie für $d = 1$ alle vorstellbaren Kombinationen von Intervallen nutzen könnt.

Beweis. Übungsaufgabe. □

Bemerkung 4.2.3. ▶ Wie in Dimension 1 gilt auch jetzt wieder, dass jede stetige Abbildung $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ ($\mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}^m)$)-messbar ist. Das gilt wieder weil wegen Proposition 2.1.4 Messbarkeit nur auf einem beliebigen Erzeuger getestet werden muss und für stetige Abbildungen Urbilder offener Mengen offen sind.

(B) Maße auf $\mathcal{B}(\mathbb{R}^d)$ und multivariate Verteilungsfunktionen

Wie für $d = 1$ definieren wir für Maße auf $\mathcal{B}(\mathbb{R}^d)$ Verteilungsfunktionen, der Unterschied ist nur die Anzahl der Variablen, d viele statt einer:

Definition 4.2.4. ▶ Ist \mathbb{P} ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}^d)$, so heißt

$$F_{\mathbb{P}}(t_1, \dots, t_d) = \mathbb{P}((-\infty, t_1] \times \dots \times (-\infty, t_d]), \quad t_1, \dots, t_d \in \mathbb{R},$$

(multivariate) Verteilungsfunktion von \mathbb{P} .

Für die Vorstellung nehmen wir immer den Fall $d = 2$. Dann ist $F(t_1, t_2)$ gerade das Maß des „unendlichen Rechtecks unten links“ unter dem Punkt (t_1, t_2) , also das Maß von $(-\infty, t_1] \times (-\infty, t_2]$. Wie für $d = 1$ (nichtfallend, rechtsstetig, Grenzwerte 1 und 0 bei unendlich) können wir aus den Eigenschaften des Maßes Eigenschaften der Verteilungsfunktion ableiten:

Proposition 4.2.5. ▶ Ist F die Verteilungsfunktion eines Wahrscheinlichkeitsmaßes auf $\mathcal{B}(\mathbb{R}^d)$, so gelten

(i) $F : \mathbb{R}^d \rightarrow [0, 1]$

(ii) F konvergiert gegen 0, wenn eine Koordinate nach $-\infty$ läuft:

$$\lim_{t_i \rightarrow -\infty} F(t_1, \dots, t_d) = \dots = \lim_{t_d \rightarrow -\infty} F(t_1, \dots, t_d) = 0.$$

(iii) F konvergiert gegen 1, wenn alle Koordinaten gemeinsam nach $+\infty$ laufen:

$$\lim_{t_i \rightarrow -\infty, i=1, \dots, d} F(t_1, \dots, t_d) = 1.$$

(iv) F ist **rechtsstetig** in jeder Koordinate.

(v) F ist **rechtecksmonoton**, d. h. für alle $a^1, a^2 \in \mathbb{R}^d$ mit $a^1 \leq a^2$ (d. h. $a_1^1 \leq a_1^2, \dots, a_d^1 \leq a_d^2$) gilt

$$\Delta_{a^1}^{a^2} F := \sum_{i_1, \dots, i_d \in \{1, 2\}} (-1)^{i_1 + \dots + i_d} F(a_1^{i_1}, \dots, a_d^{i_d}) \geq 0.$$

Eine Funktion mit den Eigenschaften (i)-(v) nennt man (multivariate) Verteilungsfunktion.

Beweis. Sei \mathbb{P} ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R}^d)$ und $F = F_{\mathbb{P}}$ die zugehörige Verteilungsfunktion. Die erste Eigenschaften von F ist klar, die weiteren drei Eigenschaften folgen aus der Stetigkeit von Maßen, genau wie für $d = 1$. Interessanter ist die Rechtecksmonotonie, die wir uns nur für $d = 1$ und $d = 2$ veranschaulichen.

$d = 1$: Einsetzen gibt hier $F(a^2) - F(a^1) \geq 0$ für $a^2 \geq a^1$, und das ist gerade die Monotonie, die wir schon kennen aus der Diskussion von Verteilungsfunktionen in einer Variablen.

$d = 2$: Einsetzen in die Formel (es gibt $2^d = 4$ Summanden) ergibt

$$F(a_1^2, a_2^2) - F(a_1^2, a_2^1) - F(a_1^1, a_2^2) + F(a_1^1, a_2^1) \geq 0.$$

Doch was soll das bedeuten? Dazu ist zu beachten, dass die zwei Punkte $a^1 \leq a^2$ ein Rechteck R „aufspannen“. Die Eckpunkte von R sind gerade (siehe Bildchen)

- (a_1^1, a_2^1) , unten links
- (a_1^2, a_2^1) , unten rechts
- (a_1^2, a_2^2) , oben rechts
- (a_1^1, a_2^2) , oben links

Weil \mathbb{P} ein Maß ist, gilt $\mathbb{P}(R) \geq 0$. Jetzt schreiben wir durch Zerlegung von R in „unendliche Rechtecke unten links“, unter Berücksichtigung der σ -Additivität von \mathbb{P} , $\mathbb{P}(R)$ als

$$\begin{aligned} \mathbb{P}(R) &= \mathbb{P}((-\infty, a_1^2] \times (-\infty, a_2^2]) - \mathbb{P}((-\infty, a_1^1] \times (-\infty, a_2^2]) \\ &\quad - \mathbb{P}((-\infty, a_1^2] \times (-\infty, a_2^1]) + \mathbb{P}((-\infty, a_1^1] \times (-\infty, a_2^1]) \\ &\stackrel{\text{Def. } F}{=} F(a_1^2, a_2^2) - F(a_1^1, a_2^2) - F(a_1^2, a_2^1) + F(a_1^1, a_2^1). \end{aligned}$$

Die Bedingung $\Delta_{a^1}^{a^2} F \geq 0$ gilt also weil $\Delta_{a^1}^{a^2} F$ nur ein komplizierter Ausdruck für die Wahrscheinlichkeit des von a^1 und a^2 aufgespannten Rechtecks ist! \square

In Analogie zum eindimensionalen Fall fragen wir nun, ob die Umkehrung auch gilt. Gibt es also für jede multivariate Verteilungsfunktion F ein Wahrscheinlichkeitsmaß \mathbb{P}_F auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, dessen Verteilungsfunktion F ist. In anderen Worten, gibt es eine bijektive Abbildung zwischen den multivariaten Verteilungsfunktionen und den Wahrscheinlichkeitsmaßen auf $\mathcal{B}(\mathbb{R}^d)$?

Satz 4.2.6.  [Analogie zu 1.4.2]

Für jede multivariate Verteilungsfunktion $F : \mathbb{R}^d \rightarrow \mathbb{R}$ gibt es genau ein Wahrscheinlichkeitsmaß \mathbb{P}_F auf $\mathcal{B}(\mathbb{R}^d)$ mit

$$\mathbb{P}_F((-\infty, t_1] \times \dots \times (-\infty, t_d]) = F(t_1, \dots, t_d), \quad t_i \in \mathbb{R}. \quad (4.1)$$

Man sagt wieder „ \mathbb{P} ist gemäß F verteilt.“

Beweis. Wir führen den Beweis nicht vollständig aus, die Argumente gehen im Prinzip wie für $d = 1$.

Eindeutigkeit: Wie immer nutzten wir für die Eindeutigkeit Dynkin-Systeme. Weil (4.1) das Maß auf \cap -stabilem Erzeuger festlegt, kann es aufgrund von Satz 1.2.12 nur ein Wahrscheinlichkeitsmaß mit der Eigenschaft (4.1) geben.

Existenz: Zur Konstruktion haben wir den Fortsetzungssatz von Carathéodory, Satz 1.3.7. Hier nur eine Skizze, die für das Verständnis der komischen Rechtecksmonotonie hilfreich ist, formuliert für $d = 2$. Zunächst muss eine σ -additive Mengenfunktion auf einem Erzeuger definiert werden. Dazu nehmen wir die Rechtecke der Form $(a_1^1, a_1^2] \times (a_2^1, a_2^2]$ und definieren

$$\mu((a_1^1, a_1^2] \times (a_2^1, a_2^2]) := \Delta_{a_1^1}^{a_1^2} F \geq 0.$$

Die Definition ist motiviert durch den Beweis von Proposition 4.2.5, $\Delta_{a_1^1}^{a_1^2} F$ war dort ja gerade die Wahrscheinlichkeit des Rechtecks $(a_1^1, a_1^2] \times (a_2^1, a_2^2]$. Nun muss man wie für $d = 1$ zeigen, dass μ eine σ -additive Mengenfunktion auf den Rechtecken ist. Das ist wieder etwas hässlich, funktioniert aber wie im Beweis von Satz 1.4.2. Hat man das geschafft, so existiert eine Fortsetzung von μ auf $\mathcal{B}(\mathbb{R}^d)$ und die tut es. \square

Vorlesung 20

Proposition 4.2.7.  [Spezialfall Produktmaß]

Sind F_1, \dots, F_d reelle Verteilungsfunktionen, so ist

$$F(t_1, \dots, t_d) := F_1(t_1) \cdot \dots \cdot F_d(t_d), \quad t_i \in \mathbb{R},$$

eine multivariate Verteilungsfunktion. Es gilt: $\mathbb{P}_F = \mathbb{P}_{F_1} \otimes \dots \otimes \mathbb{P}_{F_d}$.

Beweis. Variante 1: F ist eine multivariate Verteilungsfunktion \rightsquigarrow Große Übung. Das ist ein gutes Beispiel, um die Eigenschaften mal selber nachzurechnen. Mit Satz 4.2.6 gibt es ein dazugehöriges Maß auf $\mathcal{B}(\mathbb{R}^d)$. Um zu checken, dass dieses Maß das Produktmaß ist, berechnet man mit der Formel für $\Delta_{a_1^1}^{a_1^2} F$ (Produktform von F einsetzen!) die Wahrscheinlichkeit von Quadern aus und findet gerade die benötigte Formel (3.8) aus der Definition des Produktmaßes.

Variante 2: Das Produktmaß $\mathbb{P}_{F_1} \otimes \dots \otimes \mathbb{P}_{F_d}$ existiert auf $\mathcal{B}(\mathbb{R}^d)$ nach Korollar 3.5.3.

Behauptung: F ist die (multivariate) Verteilungsfunktion von $\mathbb{P}_{F_1} \otimes \dots \otimes \mathbb{P}_{F_d}$. Checken wir also, dass das Produktmaß die richtige Verteilungsfunktion hat:

$$\begin{aligned} \mathbb{P}_{F_1} \otimes \dots \otimes \mathbb{P}_{F_d}((-\infty, t_1] \times \dots \times (-\infty, t_d]) &\stackrel{\text{Def.}}{=} \mathbb{P}_{F_1}((-\infty, t_1]) \cdot \dots \cdot \mathbb{P}_{F_d}((-\infty, t_d]) \\ &= F_1(t_1) \cdot \dots \cdot F_d(t_d) \\ &= F(t_1, \dots, t_d), \quad t_i \in \mathbb{R}. \end{aligned}$$

 \square

Genau wie für reellwertige Zufallsvariablen gibt es absolutstetige und diskrete Wahrscheinlichkeitsmaße auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$:

Definition 4.2.8. 

- (i) Eine multivariate Verteilungsfunktion F (bzw. das zugehörige Maß \mathbb{P}_F) heißt **absolutstetig** mit Dichte $f : \mathbb{R}^d \rightarrow [0, +\infty]$, falls f messbar ist und

$$F(t_1, \dots, t_d) = \int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f(x) \, dx, \quad t_i \in \mathbb{R}.$$

- (ii) Ein multivariate Verteilungsfunktion F (bzw. das zugehörige Maß \mathbb{P}_F) heißt **diskret**, falls für ein $N \in \mathbb{N} \cup \{+\infty\}$ Vektoren $a_1, \dots, a_N \in \mathbb{R}^d$ und Wahrscheinlichkeitsgewichte $p_1, \dots, p_N \geq 0$ existieren, sodass

$$F(t_1, \dots, t_d) = \sum_{k=1}^N p_k \mathbf{1}_{[a_{k,1}, \infty) \times \dots \times [a_{k,d}, \infty)}(t_1, \dots, t_d), \quad t_i \in \mathbb{R}.$$

Die Definition ist wie für Zufallsvariablen, nur aufgrund mehrerer Koordinaten unübersichtlicher. Im diskreten Fall merkt ihr euch wie im eindimensionalen Fall einfach, dass \mathbb{P}_F Masse auf den Vektoren a_1, \dots, a_N hat und die Wahrscheinlichkeiten p_1, \dots, p_N sind.

Wenn wir besonders Wert auf die einzelnen Koordinaten legen wollen, schreiben wir das Integral auch als $\int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f(x_1, \dots, x_d) d(x_1, \dots, x_d)$. Wir meinen mit beiden Notationen immer das Lebesgue Integral bezüglich des d -dimensionalen Lebesgue-Maßes λ auf $\mathcal{B}(\mathbb{R}^d)$, also $\int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f d\lambda$. Aufgrund von Fubini gilt für absolutstetige Maße immer auch die Darstellung durch Mehrfachintegrale

$$F(t_1, \dots, t_d) = \int_{-\infty}^{t_1} \dots \left(\int_{-\infty}^{t_d} f(x_1, \dots, x_d) dx_d \right) \dots dx_1, \quad t_i \in \mathbb{R}$$

und das werden wir zum Rechnen auch meistens benutzen. Das einfachste Beispiel solche iterierten Integrale zu berechnen, tritt auf, wenn f faktorisiert, d. h. die einzelnen Koordinaten sich nur durch Produkte bedingen. In dem Fall wird die Verteilungsfunktion auch faktorisiert:

Beispiel 4.2.9. ▶ Sind f_1, \dots, f_d Dichten von reellen Verteilungsfunktionen F_1, \dots, F_d . Dann ist $f(x) = f(x_1, \dots, x_d) := f_1(x_1) \cdot \dots \cdot f_d(x_d)$ eine Dichte von $F = F_1 \cdot \dots \cdot F_d$ aus Proposition 4.2.7. Das können wir sofort mit Fubini zeigen:

$$\begin{aligned} F(t_1, \dots, t_d) &\stackrel{\text{Def.}}{=} F_1(t_1) \cdot \dots \cdot F_d(t_d) \\ &\stackrel{\text{Def.}}{=} \int_{-\infty}^{t_1} f_1(x_1) dx_1 \cdot \dots \cdot \int_{-\infty}^{t_d} f_d(x_d) dx_d \\ &\stackrel{\text{Lin.}}{=} \int_{-\infty}^{t_1} \dots \left(\int_{-\infty}^{t_d} f_1(x_1) \cdot \dots \cdot f_d(x_d) dx_d \right) \dots dx_1 \\ &\stackrel{3.5.8}{=} \int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f_1(x_1) \cdot \dots \cdot f_d(x_d) d(x_1, \dots, x_d) \\ &\stackrel{\text{Notation}}{=} \int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f(x) dx. \end{aligned}$$

Natürlich müssen wir für Fubini die Messbarkeit von f zeigen. Zum Glück ist f das Produkt messbarer Funktionen und damit wieder messbar.

(C) Zufallsvektoren

Nachdem wir die Maße auf $\mathcal{B}(\mathbb{R}^d)$ verstanden haben, kommen wir jetzt analog zum reellen Fall zu den Zufallsvektoren.

Definition 4.2.10. ▶ Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, so heißt eine $(\mathcal{A}, \mathcal{B}(\mathbb{R}^d))$ -messbar Abbildung $X: \Omega \rightarrow \mathbb{R}^d$ **Zufallsvektor**.

Proposition 4.2.11. ▶

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} : \Omega \rightarrow \mathbb{R}^d$$

ist ein Zufallsvektor genau dann, wenn $X_1, \dots, X_d: \Omega \rightarrow \mathbb{R}$ Zufallsvariablen sind.

Die Proposition ist eine reine Messbarkeitseigenschaft. Sie besagt nur, dass eine vektorwertige Abbildung messbar ist, genau dann, wenn jede Koordinatenabbildung messbar ist. Das ist ein wenig wie in Analysis 2, als immer $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ auf die Koordinatenabbildungen $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ reduziert wurde.

Beweis. „ \Rightarrow “: Für $B \in \mathcal{B}(\mathbb{R})$ gilt

$$X_k^{-1}(B) = X^{-1}(\underbrace{\mathbb{R} \times \dots \times \mathbb{R} \times B \times \mathbb{R} \times \dots \times \mathbb{R}}_{k\text{-te Stelle}}) \in \mathcal{A}.$$

„ \Leftarrow “: Messbarkeit muss nur auf einem Erzeuger gezeigt werden, wir wählen dazu $\mathcal{S} = \{B_1 \times \dots \times B_d : B_i \in \mathcal{B}(\mathbb{R})\}$:

$$\begin{aligned} X^{-1}(B_1 \times \dots \times B_d) &= \left\{ \omega \in \Omega : \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_d(\omega) \end{pmatrix} \in B_1 \times \dots \times B_d \right\} \\ &= \bigcap_{k=1}^d \{\omega \in \Omega : X_k(\omega) \in B_k\} \in \mathcal{A} \end{aligned}$$

Damit ist X messbar. □

Diskussion 4.2.12. ▶ Wegen Proposition 4.2.11 gibt es jetzt zwei Interpretationen von Zufallsvektoren:

- (i) Ein Zufallsvektor beschreibt d -viele Eigenschaften („Feature-Vektor“) **einer** zufälligen Beobachtung.
- (ii) X beschreibt die Beobachtungen von **d -vielen** zufälligen eindimensionalen Experimenten.

Wir verbalisieren (i) und (ii) unterschiedlich auch wenn es sich mathematisch eigentlich um das gleiche Objekt handelt:

- (i) „Sei $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ ein Zufallsvektor auf $(\Omega, \mathcal{A}, \mathbb{P})$.“
- (ii) „Seien X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$.“

Weiter geht's mit der Verallgemeinerung von Verteilungsfunktionen von Zufallsvariablen auf Zufallsvektoren. Analog zum eindimensionalen Fall definieren wir die gleichen Begriffe:

Definition 4.2.13. ▶

- (i) Für einen Zufallsvektor X auf $(\Omega, \mathcal{A}, \mathbb{P})$ heißt

$$F_X(t_1, \dots, t_d) := \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d), \quad t_i \in \mathbb{R},$$

Verteilungsfunktion von X . Dabei steht das Komma für „und“ (also Durchschnitt), wir lesen also „Wahrscheinlichkeit, dass $X_1 \leq t_1$ und ... und $X_d \leq t_d$ “, formell steht da aber der schlechter lesbare Ausdruck $\mathbb{P}(\bigcap_{k=1}^d \{X_k \leq t_k\})$ oder noch schlimmer $\mathbb{P}(\bigcap_{k=1}^d \{\omega \in \Omega : X_k(\omega) \leq t_k\})$. F heißt auch **gemeinsame Verteilungsfunktion** der Zufallsvariablen X_1, \dots, X_d . Wir nutzen wieder die Schreibweise $X \sim F$, wenn F die Verteilungsfunktion von X ist.

- (ii) Das Bildmaß

$$\mathbb{P}_X(B) := \mathbb{P}(X \in B), \quad B \in \mathcal{B}(\mathbb{R}^d),$$

heißt **Verteilung von X** oder die **gemeinsame Verteilung** der Zufallsvariablen X_1, \dots, X_d . \mathbb{P}_X ist ein Maß auf $\mathcal{B}(\mathbb{R}^d)$.

(iii) Zwei Zufallsvektoren X und Y heißen **identisch verteilt**, falls $F_X = F_Y$ gilt.

(iv) Für $k = 1, \dots, d$ heißt

$$\mathbb{P}_{X_k}(B) = \mathbb{P}(X_k \in B) = \mathbb{P}(\{\omega : X_k(\omega) \in B\}), \quad B \in \mathcal{B}(\mathbb{R}),$$

die **(eindimensionale) Randverteilung** von X_k und

$$F_{X_k}(t) = \mathbb{P}(X_k \leq t), \quad t \in \mathbb{R},$$

die **(eindimensionale) Randverteilungsfunktion** von X_k .

Wir hatten im eindimensionalen Fall erst die Verteilung \mathbb{P}_X und dann daraus die Verteilungsfunktion F_X definiert. Hier haben wir die Verteilungsfunktion direkt definiert und anschließend die Verteilung \mathbb{P}_X . Um mit den Definitionen zu spielen überlegt mal, warum wieder $\mathbb{P}_X = \mathbb{P}_{F_X}$ gilt. Wegen der Gleichheit ist es egal, ob wir die Begriffe wie hier oder wie in Definition 4.1.3 einführen.

Die Notationen werden hier etwas unübersichtlich, diskutieren wir sie also ein wenig.

Bemerkung 4.2.14. ▶ Weil $\mathbb{P}(X_i \leq t) = \mathbb{P}(X_1 \in \mathbb{R}, \dots, X_i \leq t, \dots, X_d \in \mathbb{R})$ gilt, folgt aus der Stetigkeit von Maßen

$$F_{X_i}(t_i) = \lim_{\substack{t_k \rightarrow \infty, \\ \forall k \neq i}} F(t_1, \dots, t_d) =: F_X(+\infty, \dots, t_i, \dots, +\infty).$$

Mit dieser Formel ist klar, wie aus der gemeinsamen Verteilungsfunktion aller X_i die Verteilungsfunktion eines einzelnen X_i berechnet werden kann: Man schickt einfach in der gemeinsamen Verteilungsfunktion alle anderen Variablen t_k nach $+\infty$.

Analog zum eindimensionalen Fall jetzt auch noch die kanonische Konstruktion von Zufallsvektoren, die funktioniert fast wörtlich wie die Konstruktion im Beweis von Satz 4.1.7.

Satz 4.2.15. ▶ **[Kanonische Konstruktion von Zufallsvektoren]**

Für jede multivariate Verteilungsfunktion F gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und einen Zufallsvektor $X: \Omega \rightarrow \mathbb{R}^d$ mit $X \sim F$.

Beweis. Als Wahrscheinlichkeitsraum definieren wir $\Omega = \mathbb{R}^d$, $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P} = \mathbb{P}_F$ aus Satz 4.2.6 und darauf den Zufallsvektor $X(\omega) = \omega$, also $X_i(\omega) = \omega_i$. Beachte: Die Identitätsabbildung $X(\omega) = \omega$ ist eine stetige Abbildung von \mathbb{R}^d nach \mathbb{R}^d und damit auch messbar. Berechnen wir die Verteilungsfunktion dieses konkreten Zufallsvektors:

$$\begin{aligned} \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d) &= \mathbb{P}_F(\{\omega \in \mathbb{R}^d : X_1(\omega) \leq t_1, \dots, X_d(\omega) \leq t_d\}) \\ &= \mathbb{P}_F(\{\omega \in \mathbb{R}^d : \omega_1 \leq t_1, \dots, \omega_d \leq t_d\}) \\ &= \mathbb{P}_F((-\infty, t_1] \times \dots \times (-\infty, t_d]) \\ &= F(t_1, \dots, t_d), \quad t_i \in \mathbb{R}. \end{aligned}$$

Das war es schon! Zu beachten ist, dass die Konstruktion weit von trivial ist. Die Existenz von \mathbb{P}_F benötigt den Satz von Carathéodory und damit die komplette Maßtheorie. □

Ab jetzt werden wir immer die Interpretation eines Zufallsvektors als d -viele Zufallsvariablen nutzen, damit wir uns langsam an Folgen von Zufallsvariablen gewöhnen.

Definition 4.2.16. ▶ Seien X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$.

- (i) X_1, \dots, X_d haben die **gemeinsame Dichte** f , falls die gemeinsame Verteilungsfunktion F_X absolutstetig ist und Dichte f hat.
- (ii) X_1, \dots, X_d heißen **diskret**, falls die gemeinsame Verteilungsfunktion F_X diskret ist.

Der diskrete Fall ist viel einfacher, als es aussieht. Die Definition bedeutet einfach nur, dass der Vektor X abzählbar viele Werte $a_1, \dots, a_N \in \mathbb{R}^d$ mit Wahrscheinlichkeiten p_1, \dots, p_N annimmt. Oder, anders ausgedrückt, dass alle Zufallsvariablen X_1, \dots, X_d nur abzählbar viele Werte annehmen.

Aus der Definition folgt sofort, dass eine gemeinsame Dichte nicht-negativ und messbar ist, sowie $\int_{\mathbb{R}^d} f(x) dx = 1$ erfüllt. Andersrum zeigt ihr in den Übungsaufgaben, dass für solch eine Funktion $F(t_1, \dots, t_d) := \int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f(x) dx$ die Eigenschaften einer multivariaten Verteilungsfunktion erfüllt.

Proposition 4.2.17. ▶ Seien X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$.

- (i) Haben X_1, \dots, X_d die gemeinsame Dichte f , so haben X_1, \dots, X_d Dichten f_1, \dots, f_d und es gilt

$$f_i(x_i) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{(d-1)\text{-viele}} \underbrace{f(x_1, \dots, x_d)}_{x_i \text{ fest}} \underbrace{dx_1 \dots dx_d}_{\text{ohne } x_i}, \quad x_i \in \mathbb{R},$$

ist eine Dichte von X_i für $i = 1, \dots, d$. In Worten: Ist X absolutstetig, so sind alle X_i absolutstetig und die Dichten der X_i entstehen durch Ausintegrieren aller anderen Variablen.

- (ii) Die Rückrichtung gilt im Allgemeinen nicht. Es gibt also absolutstetige Zufallsvariablen, die keine gemeinsame Dichte haben.

Beweis. (i) Rechnen wir die Verteilungsfunktion von X_i aus:

$$\begin{aligned} F_{X_i}(t_i) &\stackrel{\text{Def.}}{=} \mathbb{P}(X_i \leq t_i) \\ &\stackrel{\text{Trick}}{=} \mathbb{P}(X_1 \in \mathbb{R}, \dots, X_i \leq t_i, \dots, X_d \in \mathbb{R}) \\ &\stackrel{\text{Stet. Maße}}{=} \lim_{\substack{t_k \rightarrow \infty, \\ k \neq i}} \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d) \\ &\stackrel{\text{Dichte}}{=} \lim_{\substack{t_k \rightarrow \infty, \\ k \neq i}} \underbrace{\int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_d}}_{d\text{-mal}} f(x_1, \dots, x_d) dx_d \dots dx_1 \\ &\stackrel{3.2.1}{=} \int_{-\infty}^{t_i} f_i(x_i) dx_i. \end{aligned}$$

Im letzten Gleichheitszeichen haben wir fröhlich die Reihenfolge der iterierten Integrale getauscht, das war natürlich der Satz von Fubini.

(ii) Als Beispiel kann man $X_1 \sim \mathcal{U}([0, 1])$ und $X_2 = X_1$ betrachten. Der Vektor (X_1, X_2) nimmt nur Werte in $A = \{(x_1, x_2) \in [0, 1] \times [0, 1] : x_1 = x_2\}$ an und das ist eine Lebesgue-Nullmenge. In den Übungen sollt ihr euch überlegen, dass in so einem Fall keine Dichte existieren kann. \square

Die Situation ist einfacher für diskrete Zufallsvariablen. Wir sehen hier sehr deutlich, warum diskrete Stochastik so viel einfacher ist.

Proposition 4.2.18. ▶ Seien X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, dann gilt:

$$X \text{ ist ein diskreter Zufallsvektor} \iff X_1, \dots, X_d \text{ sind diskrete Zufallsvariablen.}$$

Beweis. „ \Rightarrow “: Wenn der Vektor nur abzählbar viele Werte im \mathbb{R}^d annimmt, kann natürlich auch jeder Eintrag nur abzählbar viele Werte (die Koordinaten der abzählbar vielen Werte) annehmen. Damit sind die Koordinaten X_1, \dots, X_d diskrete Zufallsvariablen. Wie im absolutstetigen Fall

ergeben sich die Wahrscheinlichkeiten durch Ausintegrieren, was hier Aussummieren bedeutet. Wenn X die Vektoren a_1, \dots, a_N annimmt, so gilt

$$\mathbb{P}(X_i = a_{k,i}) = \underbrace{\sum_{k_1=1}^N \dots \sum_{k_d=1}^N}_{(d-1)\text{-viele}} \mathbb{P}(X_1 = a_{k_1,1}, \dots, X_i = a_{k,i}, \dots, X_d = a_{k_d,d}),$$

wobei nur $(d-1)$ -viele Koordinaten aussummiert werden.

„ \Leftarrow “: Wenn X_1 die reellen Werte $a_{1,1}, \dots, a_{N_1,1}, \dots$, annimmt, X_2 die Werte $a_{1,2}, \dots, a_{N_2,2}$ annimmt (und so weiter), so kann der Zufallsvektor $X = (X_1, \dots, X_d)$ nur die $N_1 \cdot \dots \cdot N_d$ Vektoren annehmen, die sich aus den Werten ergeben. Damit ist X diskret. \square

Vorlesung 21

Bisher ist in diesem Kapitel kaum neues passiert. Nur die Rechtecksmonotonie einer multivariaten Verteilungsfunktion ist als neue Idee hinzugekommen. Das ändert sich jetzt allerdings mit dem Konzept der Unabhängigkeit. Euch ist sicher intuitiv von irgendwo bekannt, was zum Beispiel die Unabhängigkeit von drei Würfelwürfeln sein soll, insbesondere werdet ihr alle sofort sagen, dass Wahrscheinlichkeiten durch Produktbildung von Wahrscheinlichkeiten entstehen. Genau das machen wir jetzt mathematisch präzise:

Definition 4.2.19.  Seien X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$.

- (i) X_1, \dots, X_d heißen **unabhängig**, falls die gemeinsame Verteilungsfunktion in die Randverteilungsfunktionen faktorisiert, d. h.

$$F_X(t_1, \dots, t_d) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_d}(t_d), \quad t_i \in \mathbb{R}$$

oder mit Wahrscheinlichkeiten geschrieben

$$\mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d) = \mathbb{P}(X_1 \leq t_1) \cdot \dots \cdot \mathbb{P}(X_d \leq t_d), \quad t_i \in \mathbb{R}.$$

- (ii) X_1, \dots, X_d heißen **abhängig**, falls sie nicht unabhängig sind.
- (iii) X_1, \dots, X_d heißen **unabhängig und identisch verteilt (u.i.v.)**, falls sie unabhängig und identisch verteilt ($F_{X_1} = \dots = F_{X_d}$) sind. Weil die gemeinsame Verteilungsfunktion F bei u.i.v. Zufallsvariablen schon eindeutig durch jede Randverteilungsfunktion festgelegt ist, gibt man oft nur die Verteilung von X_1 an.

Was soll das abstrakte Konzept der Unabhängigkeit eigentlich bedeuten? Unabhängigkeit ist die mathematische Formulierung der Idee, dass der Wert von einer Zufallsvariablen keinen Einfluss auf den Wert der anderen Zufallsvariablen hat. Die Temperaturen in Heidelberg und Mannheim morgen um 12 Uhr sind vermutlich nicht unabhängig (ist es in Heidelberg kalt, so ist es vermutlich auch in Mannheim kalt). Andererseits hat die Temperatur morgen in Peking vermutlich keinen Einfluss darauf, wie groß der Kaffeeleck auf meiner Hose übermorgen ist.

Klingt vielleicht blöd, aber warum gibt es überhaupt u.i.v. Zufallsvariablen? Natürlich wegen des Produktmaßes!

Satz 4.2.20.  Nehmt eure Lieblingsverteilungsfunktion F , so gibt es u.i.v. X_1, \dots, X_d auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit $X_1 \sim F$. Es gilt $\mathbb{P}_X = \mathbb{P}_F \otimes \dots \otimes \mathbb{P}_F$.

Beweis. Zunächst sieht man sofort (oder nutzt 4.2.7), dass $\bar{F}(t_1, \dots, t_d) = F(t_1) \cdot \dots \cdot F(t_d)$ für $t_i \in \mathbb{R}$ eine Verteilungsfunktion ist. Mit der kanonischen Konstruktion 4.2.15 gibt es also Zufallsvariablen X_1, \dots, X_d , deren gemeinsame Verteilungsfunktion \bar{F} ist und deren gemeinsame Verteilung ist nach Proposition 4.2.7 gerade das Produktmaß. Schauen wir uns nun die Randverteilung der X_i an. Wegen Bemerkung 4.2.14 gilt also

$$\bar{F}_{X_i}(t_i) = \lim_{\substack{t_k \rightarrow \infty, \\ \forall k \neq i}} F(t_1) \cdot \dots \cdot F(t_d) = F(t_i), \quad t_i \in \mathbb{R},$$

weil für Verteilungsfunktionen $\lim_{t \rightarrow +\infty} F(t) = 1$ gilt. Damit sind X_1, \dots, X_d also identisch verteilt mit $X_1 \sim F$. Die Unabhängigkeit folgt damit direkt aus der Definition von \bar{F} :

$$F_X(t_1, \dots, t_d) = \bar{F}(t_1, \dots, t_d) = F(t_1) \cdot \dots \cdot F(t_d) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_d}(t_d) \quad t_i \in \mathbb{R}.$$

□

Beispiel 4.2.21. ▶ Sei $X_1 \sim \mathcal{N}(0, 1)$ und $X_2 := -X_1$. Dann sind X_1, X_2 identisch verteilt, jedoch nicht unabhängig. Bestimmen wir dazu zunächst die Verteilungsfunktionen:

$$F_{X_1}(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

und

$$F_{X_2}(t) = \mathbb{P}(-X_1 \leq t) = \int_{-t}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \stackrel{\text{subst.}}{=} \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Die letzte Gleichheit gilt natürlich weil $\int_{-\infty}^t f(x) dx = \int_{-t}^{+\infty} f(x) dx$ für jede symmetrische integrierbare Funktion gilt. Also gilt $X_1 \sim \mathcal{N}(0, 1)$, $X_2 \sim \mathcal{N}(0, 1)$ und damit sind X_1, X_2 identisch verteilt. Um zu zeigen, dass sie nicht unabhängig sind, berechnen wir die gemeinsame Verteilungsfunktion an einer Stelle und zeigen, dass diese nicht faktorisiert. Es gelten

$$F_X(0, 0) = \mathbb{P}(X_1 \leq 0, X_2 \leq 0) = \mathbb{P}(X_1 \leq 0, X_1 \geq 0) = \mathbb{P}(X_1 = 0) \stackrel{\text{abs. st.}}{=} 0$$

und

$$F_{X_1}(0) = F_{X_2}(0) = \mathbb{P}(X_1 \leq 0) = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{2},$$

also gilt

$$F_X(0, 0) = 0 \neq \frac{1}{4} = F_{X_1}(0) \cdot F_{X_2}(0)$$

und damit sind X_1, X_2 abhängig. Natürlich war intuitiv sowieso klar, dass X_1, X_2 nicht unabhängig sind. Unabhängig bedeutet schließlich, dass X_1 keinen Einfluss auf X_2 hat. Bei der Beziehung $X_1 = -X_2$ haben wir natürlich eine extreme Abhängigkeit: Kennen wir den Wert von X_1 , so kennen wir auch den Wert von X_2 .

Um mit gemeinsamen Verteilungen rumzurechnen, ist das nächste Korollar nützlich.

Korollar 4.2.22. ▶ Sind X_1, \dots, X_d Zufallsvariablen mit gemeinsamer Dichte f , dann gilt:

$$X_1, \dots, X_d \text{ sind unabhängig} \quad \Leftrightarrow \quad f(x) = f_1(x_1) \cdot \dots \cdot f_d(x_d) \quad \text{Lebesgue-fast überall,}$$

wobei f_1, \dots, f_d Dichten von X_1, \dots, X_d sind.

Beweis. Zuerst erinnern wir daran, dass die Existenz der gemeinsamen Dichte die Absolutstetigkeit der einzelnen Zufallsvariablen impliziert (nicht andersrum).

„ \Leftarrow “: Um die Unabhängigkeit zu prüfen, rechnen wir die Verteilungsfunktion aus und zeigen dabei, dass sie faktorisiert:

$$\begin{aligned} F_X(t_1, \dots, t_d) &\stackrel{\text{Annahme}}{=} \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_d} f_1(x_1) \cdot \dots \cdot f_d(x_d) dx_d \dots dx_1 \\ &\stackrel{\text{Lin.}}{=} \int_{-\infty}^{t_1} f_1(x_1) dx_1 \cdot \dots \cdot \int_{-\infty}^{t_d} f_d(x_d) dx_d \\ &\stackrel{\text{Def.}}{=} F_{X_1}(t_1) \cdot \dots \cdot F_{X_d}(t_d), \quad t_i \in \mathbb{R}. \end{aligned}$$

Also sind X_1, \dots, X_d nach Definition unabhängig.

„ \Rightarrow “: Rechnen wir andersrum mit der gemeinsamen Verteilungsfunktion los:

$$\begin{aligned}
 \int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f(x) dx &\stackrel{\text{Dichte}}{=} F_X(t_1, \dots, t_d) \\
 &\stackrel{\text{Ann.}}{=} F_{X_1}(t_1) \cdot \dots \cdot F_{X_d}(t_d) \\
 &= \int_{-\infty}^{t_1} f_1(x_1) dx_1 \cdot \dots \cdot \int_{-\infty}^{t_d} f_d(x_d) dx_d \\
 &\stackrel{\text{Lin.}}{=} \int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_d} f_1(x_1) \cdot \dots \cdot f_d(x_d) dx_d \dots dx_1 \\
 &\stackrel{\text{Fubini}}{=} \int_{(-\infty, t_1] \times \dots \times (-\infty, t_d]} f_1(x_1) \cdot \dots \cdot f_d(x_d) d(x_1, \dots, x_d).
 \end{aligned}$$

Beachtet wieder, dass wir sowohl mit dx als auch mit $d(x_1, \dots, dx_d)$ das Lebesguemaß meinen. Wir haben also gezeigt, dass sowohl f als auch das Produkt der f_i Dichten von F sind.

Es fehlt jetzt noch die Aussage, dass zwei Dichten automatisch fast überall gleich sind. Das ist ein bisschen hübsche Integrationstheorie. Hier ist ein kleiner Hinweis, die Details können diejenigen zusammen basteln, die aktuell noch genug Kraft übrig haben.

- Wir haben schon gesehen, dass $\nu(A) = \int_A f(x_1, \dots, x_d) d(x_1, \dots, x_d)$ und $\mu(A) = \int_A f_1(x_1) \cdot \dots \cdot f_d(x_d) d(x_1, \dots, x_d)$ Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R}^d)$ sind. Das Stichwort ist MCT.
- Nach obiger Rechnung gilt $\nu(A) = \mu(A)$ für die Rechteckmengen. Weil die Rechteckmengen ein schnittstabiler Erzeuger sind, gilt also aufgrund des Eindeutigkeitsatzes $\mu = \nu$ auf $\mathcal{B}(\mathbb{R}^d)$.
- Wir wissen aus einer Übungsaufgabe, dass für Integrale auch strikte Monotonie gilt: Wenn $h < g$ fast überall gilt, so ist das Integral über h auch strikt kleiner als das Integral über g .
- Durch die Wahl der messbaren Mengen $A := \{f \neq g\} = \{f > g\} \cup \{f < g\} =: A_1 \cup A_2$ können wir aus den letzten zwei Schritten direkt die Aussage bekommen. Warum?

□

Wir kennen jetzt Zufallsvektoren und deren Verteilungen, fehlen noch Erwartungswerte von Zufallsvektoren.

(D) Erwartungswerte

Die Definition ist analog zu der Definition für eine Zufallsvariable:

Definition 4.2.23. ▶ Seien X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $g: \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$ ($\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}})$)-messbar. Dann sei

$$\mathbb{E}[g(X_1, \dots, X_d)] := \int_{\Omega} g(X_1(\omega), \dots, X_d(\omega)) d\mathbb{P}(\omega),$$

falls das Integral wohldefiniert ist. Wir sprechen von $\mathbb{E}[g(X_1, \dots, X_d)]$ als Erwartungswert, weil $Y := g(X_1, \dots, X_d)$ eine Zufallsvariable ist (zumindest wenn g endlich ist).

Die Berechnungstheorie geht jetzt komplett analog zu dem Fall einer Zufallsvariablen. Erst der Trafosatz, dann die Rechenregeln für absolutstetige und diskrete Zufallsvektoren.

Lemma 4.2.24. ▶ Seien X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ und sei \mathbb{P}_X die gemeinsame Verteilung von $X = (X_1, \dots, X_d)$. Dann gilt

$$\mathbb{E}[g(X_1, \dots, X_d)] = \int_{\mathbb{R}^d} g(x_1, \dots, x_d) d\mathbb{P}_X(x_1, \dots, x_d),$$

wobei eine Seite wohldefiniert ist, wenn es die andere Seite ist.

Beweis. Das ist nichts anderes als der Trafosatz, genau wie in Lemma 4.1.10:

$$\begin{array}{ccc} (\Omega, \mathcal{A}, \mathbb{P}) & \xrightarrow{X} & (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X) \\ & \searrow^{g \circ X} & \downarrow g \\ & & (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}})) \end{array}$$

□

Wie für eine Zufallsvariable in Satz 4.1.11 kommen nun Rechenregeln für Wahrscheinlichkeiten und Integrale. Zusätzlich zu den diskreten und absolutstetigen Fällen gibt es jetzt auch noch Regeln für unabhängige Zufallsvariablen.

Satz 4.2.25. ▶ [Berechnungsregeln, allgemeiner Fall]

Sind X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, so gelten:

(i) Für $A \in \mathcal{B}(\mathbb{R}^d)$ gilt $\mathbb{E}[\mathbf{1}_A(X)] = \mathbb{P}(X \in A)$.

(ii) Haben X_1, \dots, X_d eine gemeinsame Dichte f , so gilt

$$\mathbb{E}[g(X_1, \dots, X_d)] = \int_{\mathbb{R}^d} g(x_1, \dots, x_d) f(x_1, \dots, x_d) \, d(x_1, \dots, x_d).$$

(iii) Sind X_1, \dots, X_d diskret und nimmt der Zufallsvektor $X = (X_1, \dots, X_d)$ die Vektoren $a_1, \dots, a_N \in \mathbb{R}^d$ mit Wahrscheinlichkeiten p_1, \dots, p_N an, so gilt

$$\mathbb{E}[g(X_1, \dots, X_d)] = \sum_{k=1}^N p_k g(a_k) = \sum_{k=1}^N \mathbb{P}(X = a_k) g(a_k).$$

Wie für $d = 1$ gilt in (ii) und (iii), dass die Erwartungswerte wohldefiniert sind (oder existieren), genau dann, wenn die Integrale bzw. Summen wohldefiniert (oder endlich) sind.

Beweis. (i) Wir müssen dazu nur die Transformationsformel und die Definition der gemeinsam Verteilung \mathbb{P}_X nutzen:

$$\mathbb{E}[\mathbf{1}_A(X)] = \int_{\mathbb{R}^d} \mathbf{1}_A(x_1, \dots, x_d) \mathbb{P}_X(x_1, \dots, x_d) = \mathbb{P}_X(A) = \mathbb{P}(X \in A).$$

(ii) und (iii) beweisen wir nicht. Der Beweis ist etwas länglich, aber exakt wie im eindimensionalen Fall bewiesen, siehe Beweis von Satz 4.1.11. □

Nach diesen allgemeinen Regeln, wenden wir uns jetzt konkret dem Fall von unabhängigen Zufallsvariablen zu. Hier werden viele Rechnungen einfacher weil Dichten und Wahrscheinlichkeiten faktorisieren.

Satz 4.2.26. ▶ Sind X_1, \dots, X_d unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, so gilt

$$\mathbb{E}[g_1(X_1) \cdot \dots \cdot g_d(X_d)] = \mathbb{E}[g_1(X_1)] \cdot \dots \cdot \mathbb{E}[g_d(X_d)]$$

für alle messbaren $g_1, \dots, g_d : \mathbb{R} \rightarrow \overline{\mathbb{R}}$. Insbesondere gilt auch

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{E}[\mathbf{1}_{A_1}(X_1)] \cdot \dots \cdot \mathbb{E}[\mathbf{1}_{A_d}(X_d)] = \mathbb{P}(X_1 \in A_1) \cdot \dots \cdot \mathbb{P}(X_d \in A_d)$$

für alle $A_1, \dots, A_d \in \mathcal{B}(\mathbb{R})$.

Beweis. Wir schreiben den Beweis nur für $d = 2$, sonst wird die Notation zu hässlich. Wir wissen schon, dass Unabhängigkeit gerade bedeutet, dass die gemeinsame Verteilung ein Produktmaß der Randverteilungen ist, d. h. $\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$. Berechnen wir damit den Erwartungswert mit dem Trafosatz und der Funktion $g(x_1, x_2) := g_1(x_1)g_2(x_2)$:

$$\begin{aligned} \mathbb{E}[g_1(X_1) \cdot g_2(X_2)] &= \mathbb{E}[g(X_1, X_2)] \\ &\stackrel{4.2.24}{=} \int_{\mathbb{R}^2} g(x_1, x_2) \, d\mathbb{P}_{(X_1, X_2)}(x_1, x_2) \\ &= \int_{\mathbb{R}^2} g(x_1, x_2) \, d\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}(x_1, x_2) \\ &\stackrel{\text{Fubini}}{=} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} g_1(x_1)g_2(x_2) \, d\mathbb{P}_{x_1}(x_1) \right) d\mathbb{P}_{x_2}(x_2) \\ &\stackrel{\text{Lin.}}{=} \int_{\mathbb{R}} g_1(x_1) \, d\mathbb{P}_{X_1}(x_1) \int_{\mathbb{R}} g_2(x_2) \, d\mathbb{P}_{X_2}(x_2) \\ &\stackrel{2 \times 4.1.10}{=} \mathbb{E}[g_1(X_1)] \cdot \mathbb{E}[g_2(X_2)]. \end{aligned}$$

Die zweite Aussage folgt aus der ersten mit den messbaren Abbildungen $g_1 = \mathbf{1}_{A_1}, \dots, g_d = \mathbf{1}_{A_d}$ sowie die wichtige Verbindung von Wahrscheinlichkeiten und Erwartungswerten aus Satz 4.2.25 mit der Menge $A = A_1 \times \dots \times A_d$ beziehungsweise Satz 4.1.14 (iv). \square

Der diskrete Fall (z. B. drei Mal Würfeln) geht in der allgemeinen Theorie leicht unter, daher schreiben wir es sicherheitshalber explizit hin:

Bemerkung. Sind X_1, \dots, X_d diskret und unabhängig, so gilt

$$\mathbb{P}(X_1 = a_1, \dots, X_d = a_d) = \mathbb{P}(X_1 = a_1) \cdot \dots \cdot \mathbb{P}(X_d = a_d).$$

Das ist einfach nur der vorherige Satz mit den Einpunktmengen $A_i = \{a_{k,i}\}$. Wenn wir also zweimal Würfeln, stände da zum Beispiel

$$\mathbb{P}(X_1 = 3, X_2 = 2) = \mathbb{P}(X_1 = 3)\mathbb{P}(X_2 = 2) = \frac{1}{6} \frac{1}{6} = \frac{1}{36}.$$

Zum Abschluss können wir die Faktorisierungen von Dichten und Wahrscheinlichkeiten noch in die Berechnungsregeln einsetzen, um einfachere Formen im Fall unabhängiger Zufallsvariablen zu bekommen. Im Prinzip ist der Satz schon in obigen Formeln enthalten, wir wollen ihn aber nochmal deutlich hinschreiben. Diese Formeln werden im nächsten Abschnitt immer wieder zum konkreten Rechnen mit Zufallsvariablen benutzt.

Satz 4.2.27. ▶ **[Berechnungsregeln, unabhängiger Fall]**

Sind X_1, \dots, X_d unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, so gelten:

(i) Haben X_1, \dots, X_d Dichten f_1, \dots, f_d , so gilt

$$\mathbb{E}[g(X_1, \dots, X_d)] = \int_{\mathbb{R}^d} g(x_1, \dots, x_d) f_1(x_1) \cdot \dots \cdot f_d(x_d) \, d(x_1, \dots, x_d).$$

(ii) Sind X_1, \dots, X_d diskret und nimmt der Zufallsvektor $X = (X_1, \dots, X_d)$ die Vektoren $a_1, \dots, a_N \in \mathbb{R}^d$ an, so gilt

$$\mathbb{E}[g(X_1, \dots, X_d)] = \sum_{k=1}^N \mathbb{P}(X_1 = a_{k,1}) \cdot \dots \cdot \mathbb{P}(X_d = a_{k,d}) g(a_{k,1}, \dots, a_{k,d}).$$

Wie für $d = 1$ gilt in (i) und (iii), dass die Erwartungswerte wohldefiniert sind (oder existieren), genau dann, wenn die Integrale bzw. Summen wohldefiniert (oder endlich) sind.

Eine direkte Folgerung aus obigen Regeln ist die Folgerung, dass Unabhängigkeit erhalten bleibt, wenn messbare Abbildungen angewandt werden.

Korollar 4.2.28. ▶ Sind X_1, \dots, X_d unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $f_1, \dots, f_d : \mathbb{R} \rightarrow \mathbb{R}$ messbar. Dann sind auch $f_1(X_1), \dots, f_d(X_d)$ unabhängige Zufallsvariablen.

Beweis. Zunächst sind die $f_i(X_i)$ auch Zufallsvariablen weil die Verknüpfung messbarer Abbildungen wieder messbar ist. Wir schreiben den Beweis nur für $d = 2$, sonst wird die Notation zu hässlich. Mit den vorherigen Sätzen (geht die Rechnung durch und sucht nach den passenden Sätzen als Wiederholung!) folgt

$$\begin{aligned} \mathbb{P}(f_1(X_1) \leq t_1, f_2(X_2) \leq t_2) &= \mathbb{P}(X_1 \in f_1^{-1}((-\infty, t_1]), X_2 \in f_2^{-1}((-\infty, t_2])) \\ &= \mathbb{E}[\mathbf{1}_{f_1^{-1}((-\infty, t_1])}(X_1) \mathbf{1}_{f_2^{-1}((-\infty, t_2])}(X_2)] \\ &= \mathbb{E}[\mathbf{1}_{f_1^{-1}((-\infty, t_1])}(X_1)] \mathbb{E}[\mathbf{1}_{f_2^{-1}((-\infty, t_2])}(X_2)] \\ &= \mathbb{P}(X_1 \in f_1^{-1}((-\infty, t_1])) \mathbb{P}(X_2 \in f_2^{-1}((-\infty, t_2])) \\ &= \mathbb{P}(f_1(X_1) \leq t_1) \mathbb{P}(f_2(X_2) \leq t_2). \end{aligned}$$

Also sind $f_1(X_1)$ und $f_2(X_2)$ unabhängig. □

Wir schließen unsere Diskussion von mehreren Zufallsvariablen mit einer wichtigen Kenngröße, die in einem gewissen Sinne die Abhängigkeit zweier Zufallsvariablen misst. In der Wahrscheinlichkeitstheorie spielen die Begriffe keine sehr große Rolle, in der Statistik jedoch eine sehr große.

Definition 4.2.29. ▶

- (i) Sind X, Y quadratintegrierbare Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, d. h. $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$, dann heißt

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Kovarianz von X und Y .

- (ii) Sind $\mathbb{V}[X], \mathbb{V}[Y] \neq 0$, das bedeutet X und Y sind nicht fast sicher konstant, so heißt

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}[X] \mathbb{V}[Y]}}$$

Korrelation von X, Y .

- (iii) Ist $\rho(X, Y) = 0$, so heißen X, Y **unkorreliert**.

In der großen Übung wurden folgende Eigenschaften diskutiert:

Bemerkung 4.2.30. ▶

- Fall X und Y endliche zweite Momente haben, so existiert die Kovarianz und es gilt $\rho(X, Y) \in [-1, 1]$. Das ist einfach nur Cauchy-Schwarz (d. h. Hölder mit $p = q = 2$):

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &\leq \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]} \sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]} = \sqrt{\mathbb{V}[X] \mathbb{V}[Y]}. \end{aligned}$$

Durchteilen gibt $\rho(X, Y) \in [-1, 1]$.

- Sind X und Y unabhängig, so gilt $\text{Cov}(X, Y) = \rho(X, Y) = 0$. Unabhängige Zufallsvariablen sind also auch unkorreliert! Das folgt sofort aus Satz 4.2.26 angewandt auf die Funktionen $g_1(x) = x - \mathbb{E}[X]$ und $g_2(y) = y - \mathbb{E}[Y]$ und Ausmultiplizieren.

- Die Korrelation wird in der Statistik genutzt, um Abhängigkeiten zu beschreiben. Positive Korrelation bedeutet, dass X und Y eher gleiches Vorzeichen haben, negative Korrelation bedeutet, dass X und Y eher ungleiches Vorzeichen haben. Je näher $\rho(X, Y)$ an ± 1 ist, desto stärker ist dieser Effekt. Je näher $\rho(X, Y)$ bei 0 ist, desto weniger wissen wir über den Zusammenhang von X und Y . Am besten sieht man das an den Extremfällen: Für $X = Y$ gilt $\rho(X, Y) = 1$, für $X = -Y$ gilt $\rho(X, Y) = -1$, für unabhängige Zufallsvariablen gilt $\rho(X, Y) = 0$.

Satz 4.2.31. ▶ [Bienaymé]

Sind X_1, \dots, X_d Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\mathbb{E}[X_1^2], \dots, \mathbb{E}[X_d^2] < \infty$, so gelten:

(i)

$$\mathbb{V}\left[\sum_{k=1}^d X_k\right] = \sum_{k=1}^d \mathbb{V}[X_k] + \sum_{\substack{i,j=1 \\ i \neq j}}^d \text{Cov}(X_i, X_j).$$

(ii)

$$\mathbb{V}\left[\sum_{k=1}^d X_k\right] = \sum_{k=1}^d \mathbb{V}[X_k],$$

falls X_1, \dots, X_d paarweise unkorreliert sind, d. h. wenn $\text{Cov}(X_i, X_j) = 0$ für alle $i, j = 1, \dots, d$, $i \neq j$.

Beweis. Übung, das ist einfach nur Ausmultiplizieren und Definitionen einsetzen. □

4.3 Rechnen mit Zufallsvariablen

Vorlesung 22

In diesem Abschnitt wollen wir endlich mal mit Zufallsvariablen konkret rechnen. Wir schauen uns an, wie man eine Zufallsvariable zu einer anderen transformiert, wie man in konkreten Beispielen mehrere Zufallsvariablen zu einer neuen transformiert und dann noch, was mit Summen von unabhängigen Zufallsvariablen passiert.

4.3.1 Inverse Transformations Methode

Schauen wir uns als erstes an, wie man einzelne Zufallsvariablen zu anderen Zufallsvariablen transformieren kann. Den Trick haben wir schon gesehen, als wir nach Beispiel 4.1.6 eine uniforme Zufallsvariable in eine exponentielle Zufallsvariable transformiert haben. Das geht auch allgemeiner. Wir brauchen dazu die Pseudoinverse, die auch in Stochastik 2 und Monte Carlo Methoden sehr prominent auftauchen wird.

Definition 4.3.1. ▶ Ist F eine Verteilungsfunktion, so heißt

$$F^{-1}(x) := \inf\{s \in \mathbb{R} : F(s) \geq x\}, \quad x \in [0, 1],$$

Pseudoinverse (oder **verallgemeinerte Inverse**) von F . Mit $\inf(\emptyset) := +\infty$ und $\inf(\mathbb{R}) := -\infty$ ist F^{-1} eine Abbildung von $[0, 1]$ nach $\bar{\mathbb{R}}$.

Der Begriff Pseudoinverse taucht auf, weil die Umkehrfunktion (inverse Funktion) nur für bijektive Funktionen definiert ist. Für Verteilungsfunktionen muss F allerdings nicht surjektiv sein (Sprünge) oder nicht injektiv sein (stückweise konstant). Wenn F stetig und streng monoton wachsend ist (z. B. wenn F eine positive Dichte hat), dann ist die Pseudoinverse einfach nur die Umkehrfunktion (Spiegelung an der Winkelhalbierenden) aus Analysis 1!

Bemerkung. Sprünge in F werden konstante Stücke in F^{-1} , konstante Stücke in F werden Sprünge in F^{-1} .

Folgende elementare Eigenschaften sind essentiell, um mit F^{-1} zu arbeiten:

Lemma 4.3.2. 

- (i) Ist F bijektiv, so ist F^{-1} die Umkehrfunktion aus Analysis 1.
- (ii) F^{-1} ist nicht-fallend.
- (iii) Es gilt $F^{-1}(y) \leq t \Leftrightarrow y \leq F(t)$ für alle $t \in \mathbb{R}$ und $y \in (0, 1)$.

Beweis.

- (i) Definition.
- (ii) Definition.
- (iii) Nach Definition gilt

$$F^{-1}(y) \leq t \Leftrightarrow \inf\{s: F(s) \geq y\} \leq t \Leftrightarrow F(t) \geq y.$$

□

Der Trick an der verallgemeinerten Inversen ist, dass man damit alle (ja, wirklich alle!) Zufallsvariablen aus einer uniformen Zufallsvariablen basteln kann, indem man diese in die verallgemeinerte Inverse einsetzt. Für die Theorie ist das unglaublich nützlich.

Satz 4.3.3.  **[Inverse Transformation Methode]**

Ist $U \sim \mathcal{U}((0, 1))$ und F eine beliebige Verteilungsfunktion, so gilt $X := F^{-1}(U) \sim F$.

Beweis. Um die Verteilungsfunktion von X zu berechnen, setzen wir einfach die Definition ein und führen die Wahrscheinlichkeit mit (iii) aus dem letzten Lemma auf die uniforme Verteilung zurück:

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(F^{-1}(U) \leq t) = \mathbb{P}(U \leq F(t)) = F(t), \quad t \in \mathbb{R}.$$

Im letzten Schritt haben wir die Verteilungsfunktion von $\mathcal{U}((0, 1))$ eingesetzt (siehe Übungsaufgaben). □

Eine kleine Bemerkung zur uniformen Verteilung in der Trafomethode. Oft sieht man $V \sim \mathcal{U}([0, 1])$ statt $U \sim \mathcal{U}((0, 1))$. Im Prinzip ist es egal, was man nimmt, die beiden Zufallsvariablen U und V sind nämlich identisch verteilt (siehe Übungsaufgaben), der Unterschied liegt auf einer Nullmenge. Weil aber $F^{-1}(0) = -\infty$ und $F^{-1}(1) = +\infty$ sein kann, müssten wir uns dann mit V Gedanken über die Definition einer Zufallsvariablen machen. Die nehmen gemäß unserer Definition nur endliche reelle Werte an.

Die Trafomethode sieht auf den ersten Blick klasse aus. Wenn man eine uniforme Zufallsvariable simulieren kann (das nennt man auch *samplen*), so kann man automatisch alle Zufallsvariablen simulieren! In der Monte Carlo Vorlesung wird der erste Simulationsalgorithmus für Zufallsvariablen daher wie folgt funktionieren. Simuliere eine $\mathcal{U}((0, 1))$ Zufallsvariable (das ist eigentlich Zahlentheorie) und setze diese in F^{-1} ein. Das gibt dann eine Simulation einer Zufallsvariablen mit Verteilungsfunktion F . Leider ist das zu schön, um wahr zu sein. Der Grund ist, dass leider F^{-1} selten explizit bekannt ist, dann bringt die Methode natürlich nicht viel für praktische Anwendungen. Es gibt aber einige Beispiele, in denen F^{-1} explizit bekannt ist.

Beispiel 4.3.4. 

- Für den diskreten Fall solltet ihr euch Beispiele aufs Papier malen. Ist F eine diskrete Verteilungsfunktion mit Werten $a_1 < \dots < a_N$ und Wahrscheinlichkeiten p_1, \dots, p_N , so ist F^{-1} eine Treppenfunktion. Die Trafomethode gibt folgenden intuitiven Algorithmus, eine diskrete Zufallsvariable zu simulieren: Zerlege das Intervall $(0, 1)$ in die Teilintervalle $I_1 = (0, p_1]$, $I_2 = (p_1, p_1 + p_2]$ bis $I_N = (p_1 + \dots + p_{N-1}, 1)$ und ziehe uniform eine Zahl U aus $(0, 1)$, U liegt also in einem der Intervalle I_k . Liegt U in I_k , so gebe den Wert a_k aus. Die so gewonnene Zufallsvariable nimmt die Werte a_k mit Wahrscheinlichkeiten p_k an, ist also diskret mit Verteilungsfunktion F .

- Für die Exponentialverteilung $\text{Exp}(1)$ ist $F^{-1}(t) = -\log(1-t), t \in [0, 1]$. Also gilt $X = -\ln(1-U) \sim \text{Exp}(1)$, sofern U eine uniforme Zufallsvariable ist.
- Für die Normalverteilung funktioniert die Methode nicht. Wir haben keine explizite Formel für F , also auch nicht für F^{-1} . Die Methode kann aber trotzdem für theoretische Betrachtungen genutzt werden, wie wir bei der Konvergenz von Zufallsvariablen sehen werden.

Mehr explizite Beispiele kennen wir leider noch nicht.

4.3.2 Ein paar konkrete Beispiele

Als nächstes wollen wir an ein paar Beispielen zeigen, wie man aus verschiedenen Zufallsvariablen neue Zufallsvariablen basteln kann. Ihr sollt damit ein besseres Gefühl für verschiedene Zufallsvariablen bekommen und lernen, wie man mit den Erwartungswerten von unabhängigen Zufallsvariablen rechnet. Im nächsten Abschnitte betrachten wir dann den Spezialfall von Summen von Zufallsvariablen.

Beispiel 4.3.5. [gespiegelte Zufallsvariable]

Eine Zufallsvariable X heißt symmetrisch, wenn $\mathbb{P}(X \leq t) = \mathbb{P}(X \geq -t)$ für alle $t \in \mathbb{R}$ gilt. Das bedeutet, dass X gleichwahrscheinlich Werte in an 0 gespiegelten Mengen annimmt. Eine äquivalente Formulierung ist, dass $X \sim -X$ gilt. Checken wir dafür einmal schnell die Verteilungsfunktionen:

$$F_{-X}(t) = \mathbb{P}(-X \leq t) = \mathbb{P}(X \geq -t) = \mathbb{P}(X \leq t) = F_X(t), \quad t \in \mathbb{R}.$$

Symmetrie ist einfach im diskreten oder absolutstetigen Fall zu erkennen. Im diskreten bedeutet dies, dass die „Zähldichte“ immer an Werten a_k und $-a_k$ die gleichen Werte p_k annimmt. Im absolutstetigen Fall muss die Dichte achsensymmetrisch sein, so wie zum Beispiel bei $\mathcal{N}(0, 1)$.

Beispiel 4.3.6. [Uniform zu uniform]

Ist $U \sim \mathcal{U}([0, 1])$ eine uniform verteilte Zufallsvariable, so ist auch $V := 1 - U \sim \mathcal{U}([0, 1])$. Dazu berechnen wir einfach die Verteilungsfunktion von V , weil wir die Verteilungsfunktion von U kennen:

$$F_V(t) = \mathbb{P}(V \leq t) = \mathbb{P}(U \geq 1 - t) = 1 - \mathbb{P}(U \leq 1 - t) = \begin{cases} 0 & : t < 0 \\ 1 - (1 - t) & : t \in (0, 1) \\ 1 & : t > 1 \end{cases},$$

für $t \in \mathbb{R}$. Damit hat V die Verteilungsfunktion einer $\mathcal{U}([0, 1])$ -verteilten Zufallsvariable. Beachtet, dass wir $\mathbb{P}(U \leq 1 - t) = \mathbb{P}(U < 1 - t)$, also $\mathbb{P}(U = 1 - t) = 0$ genutzt haben. Das liegt daran, dass U eine absolutstetige Zufallsvariable ist.

Beispiel 4.3.7. [Minimum von Exp-verteilten Zufallsvariablen]

Sind $X \sim \text{Exp}(\alpha)$ und $Y \sim \text{Exp}(\beta)$ unabhängige Zufallsvariablen, so gilt $\min\{X, Y\} \sim \text{Exp}(\alpha + \beta)$. Wir beachten dazu, dass das Minimum zweier messbarer Abbildungen wieder messbar ist, also ist auch $Z := \min\{X, Y\}$ eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Berechnen wir nun die Verteilungsfunktion von Z (das Minimum ist größer als t genau dann, wenn beide größer als t sind):

$$\begin{aligned} \mathbb{P}(Z \leq t) &= 1 - \mathbb{P}(\min\{X, Y\} > t) \\ &\stackrel{\text{Unab.}}{=} 1 - \mathbb{P}(X > t, Y > t) \\ &= 1 - \mathbb{P}(X > t)\mathbb{P}(Y > t) \\ &= 1 - e^{-\alpha t}e^{-\beta t} = 1 - e^{-(\alpha+\beta)t}, \quad t > 0. \end{aligned}$$

Wenn man $t \leq 0$ einsetzt, kommt überall 0 raus weil X und Y nicht-negative Zufallsvariablen sind. Damit hat Z die Verteilungsfunktion von $\text{Exp}(\alpha + \beta)$. Exakt die gleiche Rechnung funktioniert

auch mit geometrischen Zufallsvariablen, das Minimum zweier unabhängiger geometrischer Zufallsvariablen ist wieder geometrisch - probiert das mal aus!

Beispiel 4.3.8.  Seien $X, Y \sim \text{Exp}(1)$ unabhängig, so gilt

$$\frac{X}{X+Y} \sim \mathcal{U}([0, 1]).$$

Diese Übungsaufgabe braucht ein gutes Verständnis vom Rechnen mit Zufallsvektoren, daher ein Tipp. Zu berechnen ist die Verteilungsfunktion $\mathbb{P}(\frac{X}{X+Y} \leq t)$. Die einzige Information ist die gemeinsame Dichte von (X, Y) , die aufgrund der Unabhängigkeit faktorisiert. Wir schreiben die Verteilungsfunktion mit geeignetem g wieder als $\mathbb{E}[g(X, Y)]$ um, und nutzen dann die Berechnungsregel im absolutstetigen Fall. In diesem Fall schreibt man

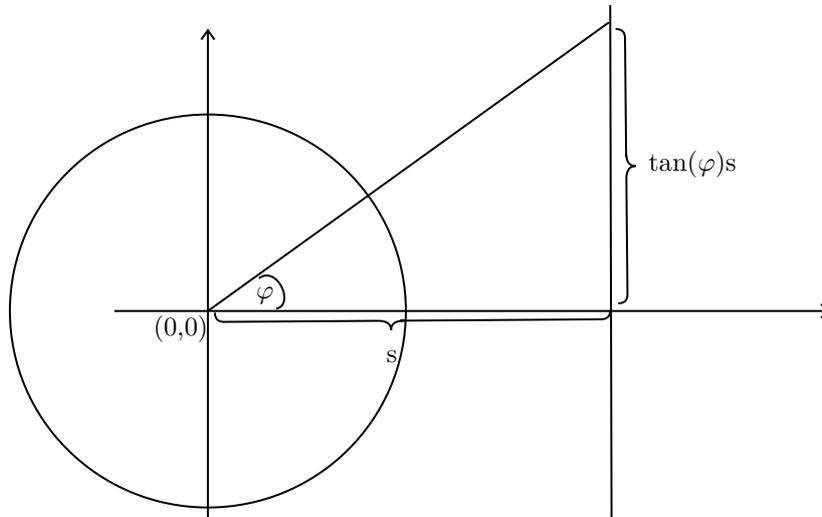
$$\mathbb{P}\left(\frac{X}{X+Y} \leq t\right) = \mathbb{P}\left(X \leq Y \frac{t}{1-t}\right) = \mathbb{E}\left[\mathbf{1}_{(-\infty, Y \frac{t}{1-t}]}(X)\right].$$

Die rechte Seite kann man durch Einsetzen ausrechnen, auf geht's!

Einen Haufen weiterer Verbindungen verschiedener Verteilungen kann man hier (klicken) verlinkt finden. Sollte der Link bei eurem pdf-viewer nicht funktionieren, sucht einfach nach „Wikipedia relationships among probability distributions“. Auch witzig ist zum Beispiel folgende Aussage: Sind $X, Y \sim \mathcal{N}(0, 1)$ und unabhängig, so ist $Z := \frac{X}{Y}$ Cauchyverteilt!

Beispiel 4.3.9.  [Discokugel]

Was hat eine „Discokugel“ (blinkende Kugel in der Mitte eines Raumes, die mittels kleiner Laser bunte Punkte an die Wand wirft) mit der Cauchyverteilung zu tun? Die Punkte an der Wand sind Realisierungen einer Cauchyverteilung! Schauen wir uns zunächst eine Skizze an:



Der Kreis ist ein zweidimensionaler Schnitt der Kugel. Es wird ein Punkt auf der rechten Hälfte des Kreisrandes uniform gezogen (d. h. der Winkel φ wird uniform aus $(-\frac{\pi}{2}, \frac{\pi}{2})$ gezogen, φ hat also Dichte $\frac{1}{\pi} \mathbf{1}_{(-\frac{\pi}{2}, \frac{\pi}{2})}$). Dann wird der Laser vom Ursprung in Richtung des gezogenen Punktes auf dem Kreisrand geschossen und bis zur Mauer verlängert, wo dann ein bunter Punkt erscheint. Aus der Schule sollte noch bekannt sein, dass der Treffpunkt der Mauer $(s, \tan(\varphi)s)$ ist. Berechnen wir nun die Verteilung des Treffpunktes Y (y -Achsen Abstand) auf der Mauer. Weil $\varphi \sim \mathcal{U}((-\frac{\pi}{2}, \frac{\pi}{2}))$ ist, gilt

$$\mathbb{P}(Y \leq t) = \mathbb{P}(\tan(\varphi)s \leq t) = \mathbb{P}(\varphi \leq \arctan(t/s)) = \frac{1}{2} + \frac{1}{\pi} \arctan(t/s),$$

wobei in der letzten Gleichheit die Verteilungsfunktion von $\mathcal{U}((-\frac{\pi}{2}, \frac{\pi}{2}))$ eingesetzt wurde (beachte, dass $\arctan(t/s) \in (-\frac{\pi}{2}, \frac{\pi}{2})$ für alle $t, s \in \mathbb{R}$). Damit ist die Verteilungsfunktion von Y gerade die Verteilungsfunktion von $\text{Cauchy}(s, 0)$.

Wir hatten angemerkt, dass die inverse Transformations Methode für die Normalverteilung nicht funktioniert weil F^{-1} nicht explizit bekannt ist. Das nächste Beispiel ist daher ziemlich verblüffend, die sogenannte Box-Muller Methode. Man kann aus einer uniformen Zufallsvariable zwar nicht einfach eine normalverteilte Zufallsvariable bekommen, man kann aber ganz einfach aus zwei uniformen Zufallsvariablen zwei (und damit auch eine) normalverteilte bekommen! Sind U_1, U_2 unabhängige $\mathcal{U}([0, 1])$ verteilte Zufallsvariablen, so sind

$$X_1 = \cos(2\pi U_1)\sqrt{-\log(U_2)} \quad \text{und} \quad X_2 = \sin(2\pi U_1)\sqrt{-\log(U_2)}$$

zwei unabhängige $\mathcal{N}(0, 1)$ Zufallsvariablen. Wer gerade noch Energie übrig hat, kann mal versuchen, $\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2)$ mit Polarkoordinaten auszurechnen (alle anderen können sich das in der Monte Carlo Methoden Vorlesung anschauen). Um die Verteilung zu berechnen, solltet ihr euch an Korollar 4.2.22 erinnern. Das Korollar besagt, dass U_1, U_2 und X_1, X_2 gemeinsame Dichten haben, nämlich jeweils das Produkt der einzelnen, also $\mathbf{1}_{[0,1] \times [0,1]}$ bzw. $\frac{1}{2\pi}e^{-(x_1^2+x_2^2)/2}$. Damit wisst ihr, wie ihr $\mathbb{P}(X_1 \leq t_1, X_2 \leq t_2)$ durch Einsetzen der Definition ausrechnet und was rauskommen sollte.

4.3.3 Summen von unabhängigen Zufallsvariablen

Wir berechnen nun die Verteilungen von Summen unabhängiger Zufallsvariablen. Was ist zum Beispiel die Verteilung der Summe zweier exponentialverteilter Zufallsvariablen? Oder wie ist die Summe zweier Normalverteilungen verteilt? Allgemein ist das ziemlich schwierig zu sagen, wenn die Zufallsvariablen aber unabhängig sind, gibt es schöne Rechenricks.

Starten wir mit dem diskreten Fall, der ist einfacher. Die Idee ist einfach: Durch welche Kombination von Werten, kann die Summe $X + Y$ einen gegebenen Wert annehmen? Natürlich indem Y irgendeinen Wert annimmt, und X gerade die Differenz zum gegebenen Wert. Wenn man alle solche Möglichkeiten addiert (und die Unabhängigkeit ausnutzt), bekommt man die diskrete Faltungsformel:

Satz 4.3.10. [Diskrete Faltungsformel]

Sind X, Y diskrete Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann ist auch $X + Y$ diskret mit möglichen Werten in $X(\Omega) + Y(\Omega) := \{a + b : a \in X(\Omega), b \in Y(\Omega)\}$, wobei $X(\Omega), Y(\Omega)$ die möglichen Werte von X, Y sind. Sind X und Y unabhängig, so können die Wahrscheinlichkeiten mit der Faltungsformel berechnet werden:

$$\mathbb{P}(X + Y = k) = \sum_{b \in Y(\Omega)} \mathbb{P}(X = k - b)\mathbb{P}(Y = b).$$

Beweis. Der Trick ist es, ein diskretes Ereignis in eine disjunkte Vereinigung von Teilereignissen (alle Möglichkeiten) zu zerlegen und darauf die σ -Additivität anzuwenden. Wir schreiben den Beweis einmal knapp auf, so schreibt man Argumente mit diskreten Zufallsvariablen fast immer auf, und dann noch einmal super ausführlich, um die Maßtheorie dahinter zu erkennen:

$$\begin{aligned} \mathbb{P}(X + Y = k) &\stackrel{\sigma\text{-add.}}{=} \sum_{b \in Y(\Omega)} \mathbb{P}(X + Y = k, Y = b) \\ &= \sum_{b \in Y(\Omega)} \mathbb{P}(X = k - b, Y = b) \\ &\stackrel{\text{unab.}}{=} \sum_{b \in Y(\Omega)} \mathbb{P}(X = k - b)\mathbb{P}(Y = b). \end{aligned}$$

Die zweite Gleichheit nutzt σ -Additivität weil alle Möglichkeiten von Y ausprobiert wurden.

Ganz ausführlich sieht das gleiche Argument wie folgt aus:

$$\begin{aligned}
\mathbb{P}(X + Y = k) &\stackrel{\text{Notation}}{=} \mathbb{P}(\{\omega : X(\omega) + Y(\omega) = k\}) \\
&\stackrel{\text{Trick}}{=} \mathbb{P}\left(\bigcup_{b \in Y(\Omega)} \{\omega : X(\omega) + Y(\omega) = k, Y(\omega) = b\}\right) \\
&\stackrel{\sigma\text{-add.}}{=} \sum_{b \in Y(\Omega)} \mathbb{P}(\{\omega : X(\omega) + Y(\omega) = k, Y(\omega) = b\}) \\
&= \sum_{b \in Y(\Omega)} \mathbb{P}(\{\omega : X(\omega) = k - b, Y(\omega) = b\}) \\
&\stackrel{\text{Notation}}{=} \sum_{b \in Y(\Omega)} \mathbb{P}(X = k - b, Y = b) \\
&\stackrel{\text{unab.}}{=} \sum_{b \in Y(\Omega)} \mathbb{P}(X = k - b)\mathbb{P}(Y = b).
\end{aligned}$$

Wir müssen in der Mathematik immer vorsichtig sein, einfache Dinge nicht zu sehr zu verkomplizieren. Diskrete Zufallsvariablen sind ganz sicher so ein Beispiel! \square

Als Anmerkung zum Freuen auf die strahlende Zukunft: Genau wegen dieses kleinen Tricks, kann man so schön mit Markovketten rumrechnen!

Die Formel sieht vielleicht abstrakt und unhandlich aus, wie können aber wirklich ganz einfach in konkrete Beispielen damit rechnen:

Beispiel 4.3.11.  Sind X_1, \dots, X_n u.i.v mit $X_1 \sim \text{Ber}(p)$ für ein $p \in (0, 1)$, dann gilt $X_1 + \dots + X_n \sim \text{Bin}(n, p)$. Interpretation: Wenn $\text{Ber}(p)$ die erfolgreiche Ausführung eines Versuchs (1=Erfolg, 0=Misserfolg) beschreibt, so beschreibt $\text{Bin}(n, p)$ die Anzahl der erfolgreichen Ausführungen von n unabhängigen Versuchen.

Beweis. Induktion über n :

IA: $n = 1$: \checkmark $\text{Ber}(p) = \text{Bin}(1, p)$ nach Definition beider Verteilungen.

IV: Die Behauptung gelte für ein *beliebiges*, aber festes $n \in \mathbb{N}$

IS: Indem wir $X_1 + \dots + X_{n+1} = (X_1 + \dots + X_n) + X_{n+1}$ klammern, wenden wir die diskrete Faltungsformel an und nutzen dann die Induktionsvoraussetzung:

$$\begin{aligned}
\mathbb{P}(X_1 + \dots + X_{n+1} = k) &\stackrel{4.3.10}{=} \mathbb{P}(X_1 + \dots + X_n = k - 1)\mathbb{P}(X_{n+1} = 1) \\
&\quad + \mathbb{P}(X_1 + \dots + X_n = k)\mathbb{P}(X_{n+1} = 0) \\
&= \binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} p + \binom{n}{k} p^k (1-p)^{n-k} (1-p) \\
&= \left(\binom{n}{k-1} + \binom{n}{k} \right) p^k (1-p)^{n-k+1} \\
&= \binom{n+1}{k} p^k (1-p)^{n-k+1}.
\end{aligned}$$

Im letzten Schritt haben wir eine Rechenregel für Binomialkoeffizienten benutzt, die kennt ihr vermuthlich aus Analysis 1. Also ist $X_1 + \dots + X_n \sim \text{Bin}(n, p)$ gezeigt. \square

Beispiel 4.3.12.  Seien $X \sim \text{Poi}(\lambda), Y \sim \text{Poi}(\beta)$ unabhängig. Dann ist auch $X + Y$ Poissonverteilt, und zwar mit Parameter $\lambda + \beta$. In den Übungen setzt ihr einfach in die Faltungsformel ein, um $X + Y \sim \text{Poi}(\lambda + \beta)$ zu zeigen. Easy.

Definition 4.3.13. 

- (i) Sind μ_1, \dots, μ_n Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$, so heißt das Bildmaß vom Produktmaß $\mu_1 \otimes \dots \otimes \mu_n$ unter der Abbildung $h_d(x_1, \dots, x_d) = x_1 + \dots + x_d$ **Faltung** der Maße. Wir schreiben $\mu_1 * \dots * \mu_n$ für die Faltung.
- (ii) Sind X_1, \dots, X_d unabhängige Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, so heißt $\mathbb{P}_{X_1} * \dots * \mathbb{P}_{X_d}$ **Faltung** von X_1, \dots, X_d .

Bemerkung. Natürlich ist die Definition der Faltung abstrakt, andererseits ist sie auch konkret. Die Faltung ist nichts weiter als die Verteilung der Summe unabhängiger Zufallsvariablen:

$$X_1 + \dots + X_n \sim \mathbb{P}_{X_1} * \dots * \mathbb{P}_{X_n}.$$

Wir müssen uns dafür nur daran erinnern, dass die Verteilung unabhängiger Zufallsvariablen das Produktmaß ist:

$$\begin{aligned} \mathbb{P}(X_1 + \dots + X_d \in B) &= \mathbb{P}(X_1, \dots, X_d \in h_d^{-1}(B)) \\ &\stackrel{\text{unab.}}{=} \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_d}(h_d^{-1}(B)) \\ &\stackrel{\text{Def. push-forw.}}{=} \mathbb{P}_{X_1} * \dots * \mathbb{P}_{X_d}(B). \end{aligned}$$

Mit der Definition der Faltung können wir erstmal nicht viel anstellen, wir haben die Faltung schließlich einfach als das definiert, was wir berechnen wollen. Was wir gerne hätten, wäre ein analog zu der diskreten Faltungsformel, weil wir damit in konkreten Beispielen rechnen können. Hier ist die allgemeine Formel, danach konkretisieren wir diese für den Fall mit Dichten.

Proposition 4.3.14.  Seien μ_1, μ_2 Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$ mit Verteilungsfunktionen F_1, F_2 , dann gelten:

- (i) Mit $B - y := \{x - y : x \in B\}$, d. h. die Verschiebung der Menge B um y nach links, gilt

$$\mu_1 * \mu_2(B) = \int_{\mathbb{R}} \mu_1(B - y) \, d\mu_2(y)$$

für alle $B \in \mathcal{B}(\mathbb{R})$.

- (ii) Für alle $t \in \mathbb{R}$ gilt

$$F_{\mu_1 * \mu_2}(t) = \int_{\mathbb{R}} F_1(t - y) \, d\mu_2(y).$$

Beweis.

- (i) Wegen der Manipulation

$$\mathbf{1}_{B-y}(x) = \begin{cases} 1 & : x \in B - y \\ 0 & : \text{sonst} \end{cases} = \begin{cases} 1 & : x + y \in B \\ 0 & : \text{sonst} \end{cases} = \mathbf{1}_{h_2^{-1}(B)}(x, y)$$

folgt

$$\begin{aligned} \mu_1 * \mu_2(B) &\stackrel{\text{Def.}}{=} \mu_1 \otimes \mu_2(h_2^{-1}(B)) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_{h_2^{-1}(B)}(x, y) \, d\mu_1 \otimes \mu_2(x, y) \\ &= \int_{\mathbb{R}^2} \mathbf{1}_{B-y}(x) \, d\mu_1 \otimes \mu_2(x, y) \\ &\stackrel{\text{Fubini}}{=} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{B-y}(x) \, d\mu_1(x) \right) \mu_2(y) \\ &= \int_{\mathbb{R}} \mu_1(B - y) \, d\mu_2(y). \end{aligned}$$

Das ist die erste Aussage.

- (ii) Wir setzen in (i) die Mengen $B = (-\infty, t]$ mit $B - y = (-\infty, t - y]$ ein, und nutzen die Definition der Verteilungsfunktion.

□

Nun zur konkreten Rechenvorschrift, um die Verteilung der Summe unabhängiger Zufallsvariablen mit Dichten zu berechnen:

Satz 4.3.15.  [Stetige Faltungsformel]

Sind X, Y unabhängige absolutstetige Zufallsvariablen mit Dichten f_X, f_Y , dann ist auch $X + Y$ absolutstetig und hat Dichte

$$f_{X+Y}(x) = \int_{\mathbb{R}} f_X(x-y)f_Y(y) dy, \quad x \in \mathbb{R}.$$

Beweis. Mit vorheriger Proposition kennen wir die Verteilungsfunktion der Summe. Wir rechnen diese mit der Formel aus und setzen dabei die bekannten Dichten ein:

$$\begin{aligned} \mathbb{P}_{X+Y}((-\infty, t]) &\stackrel{\text{Def. Faltung}}{=} \mathbb{P}_X * \mathbb{P}_Y((-\infty, t]) \\ &\stackrel{4.3.14}{=} \int_{\mathbb{R}} \mathbb{P}_X((-\infty, t-y]) d\mathbb{P}_Y(y) \\ &\stackrel{3.3.2}{=} \int_{\mathbb{R}} \int_{-\infty}^{t-y} f_X(x) dx f_Y(y) dy \\ &\stackrel{\text{Subst.}}{=} \int_{\mathbb{R}} \int_{-\infty}^t f_X(x-y) dx f_Y(y) dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{(-\infty, t]}(x) f_X(x-y) f_Y(y) dx dy \\ &\stackrel{\text{Fubini}}{=} \int_{-\infty}^t \int_{\mathbb{R}} f_X(x-y) f_Y(y) dy dx \\ &= \int_{-\infty}^t f_{X+Y}(x) dx. \end{aligned}$$

Also ist $X + Y$ absolutstetig mit behaupteter Dichte f_{X+Y} . □

Das Konzept der Faltung kommt nicht aus der Stochastik, daher können wir mit dem Begriff „Faltung“ auch nichts anfangen. Die Faltung zweier integrierbarer Funktionen wird in der Fourieranalysis als

$$f * g(x) = \int_{\mathbb{R}} f(x-y)g(y) dy, \quad x \in \mathbb{R},$$

definiert und zum Beispiel in der Signalverarbeitung studiert. Dass die Faltung bei uns für Summen unabhängiger Zufallsvariablen auftaucht, ist ein Zufall der Mathematik. Es gibt also keinen Grund sich Gedanken über die Begrifflichkeit Faltung zu machen, für uns ist es einfach nur eine Berechnungsformel.

Vorlesung 23

Beispiel 4.3.16. 

- (i) Sind $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ unabhängig, dann ist auch die Summe wieder normalverteilt, genauer, es gilt $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Das kann man mit der stetigen Faltungsformel ausrechnen:

$$\begin{aligned} f_{X_1+X_2}(x) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-y-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}} dy \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} e^{-\frac{(x-(\mu_1+\mu_2))^2}{2(\sigma_1^2 + \sigma_2^2)}}, \quad x \in \mathbb{R}. \end{aligned}$$

Auf den ersten Blick ist nicht so klar, ob die Berechnung einfach oder nicht so einfach ist. Im Prinzip muss man nur richtig quadratisch ergänzen und Substituieren. In der Tat ist die Rechnung ziemlich böse, klappt aber. Weil wir gleich eine viel einfachere Methode kennenlernen, lassen wir die Rechnung weg.

- (ii) Sind $X_1 \sim \text{Exp}(\lambda)$, $X_2 \sim \text{Exp}(\lambda)$ unabhängig, so ist $X_1 + X_2 \sim \Gamma(2, \lambda)$. Das kann man direkt mit der stetigen Faltungsformel ausrechnen:

$$\begin{aligned} f_{X_1+X_2}(x) &= \int_{\mathbb{R}} f_{X_1}(x-y)f_{X_2}(y) dy \\ &= \int_{\mathbb{R}} \mathbf{1}_{[0,\infty)}(x-y)\lambda e^{-\lambda(x-y)} \mathbf{1}_{[0,\infty)}(y)\lambda e^{-\lambda y} dy \\ &= \lambda^2 \int_0^x e^{-\lambda x} dy = \lambda^2 x e^{-\lambda x}. \end{aligned}$$

Jetzt noch eine ganz andere Methode zur Berechnung der Verteilung der Summe von unabhängigen Zufallsvariablen. Dafür nutzen wir einen Satz der Wahrscheinlichkeitstheorie, den Eindeutigkeitsatz für momenterzeugende Funktionen. Der Satz besagt, dass Verteilungen eindeutig durch ihre momenterzeugenden Funktionen festgelegt sind, falls diese um 0 existieren.

Satz 4.3.17. ▶ Seien X und Y Zufallsvariablen für die die momenterzeugenden Funktionen $\mathcal{M}_X, \mathcal{M}_Y$ in $(-\varepsilon, \varepsilon)$ existieren für ein $\varepsilon > 0$. Falls $\mathcal{M}_X(t) = \mathcal{M}_Y(t)$ für alle $t \in (-\varepsilon, \varepsilon)$ gilt, so gilt $X \sim Y$.

Beweis. Siehe fortgeschrittene Stochastikvorlesung - hart! □

Für Verteilungen mit expliziten momenterzeugenden Funktionen ist diese ein extrem nützliches Hilfsmittel.

Beispiel 4.3.18. ▶ Mit der momenterzeugenden Funktion können wir ganz einfach die Skalierungseigenschaft der Normalverteilung zeigen: Ist $X \sim \mathcal{N}(0, 1)$, $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, so gilt $Y := \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$. Rechnen wir dazu die momenterzeugende Funktion von Y aus:

$$M_Y(t) = \mathbb{E}[e^{(\sigma X + \mu)t}] = \mathbb{E}[e^{\sigma t X}] e^{\mu t} = M_X(\sigma t) \cdot e^{\mu t} = e^{\frac{\sigma^2 t^2}{2} + \mu t}, \quad t \in \mathbb{R},$$

und die rechte Seite ist gerade die momenterzeugende Funktion einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariable. Also folgt die Behauptung aus Satz 4.3.17.

Gemeinsam mit folgender Proposition ist Satz 4.3.17 ein mächtiges Hilfsmittel um die Verteilung Summen unabhängiger Zufallsvariablen zu berechnen:

Proposition 4.3.19. ▶ Sind X und Y unabhängige Zufallsvariablen mit $\mathcal{M}_X(t), \mathcal{M}_Y(t) < \infty$ für ein $t \in \mathbb{R}$, so gilt auch $\mathcal{M}_{X+Y}(t) = \mathcal{M}_X(t)\mathcal{M}_Y(t)$.

Beweis. Das folgt direkt daraus, dass Erwartungswerte von Produkten unabhängiger Zufallsvariablen faktorisieren, siehe Satz 4.2.26:

$$\mathcal{M}_{X+Y}(t) = \mathbb{E}[e^{t(X+Y)}] = \mathbb{E}[e^{tX} \cdot e^{tY}] = \mathbb{E}[e^{tX}] \cdot \mathbb{E}[e^{tY}] = \mathcal{M}_X(t)\mathcal{M}_Y(t).$$

□

Die Anwendung der momenterzeugenden Funktion zur Bestimmung der Verteilung von Summen unabhängiger Zufallsvariablen versteht man am besten mit folgendem einfachen Beispiel. Beachte: Das Beispiel ist einfach, die Aussage aber nicht. Hätten wir nicht die Kanone aus der Wahrscheinlichkeitstheorie ausgepackt, hätten wir das mit der stetigen Faltungsformel nachrechnen müssen.

Beispiel 4.3.20. ▶ Kommen wir zurück zu Beispiel 4.3.16 (i) und berechnen mit der Proposition die momenterzeugende Funktion der Summe der zwei unabhängigen Normalverteilungen. Wegen Proposition 4.3.19 und der in den Übungen berechneten Formel für die momenterzeugende Funktion einer Normalverteilung gilt

$$\mathcal{M}_{X_1+X_2}(t) = \mathcal{M}_{X_1}(t)\mathcal{M}_{X_2}(t) = e^{\mu_1 t + \frac{\sigma_1^2}{2} t^2} e^{\mu_2 t + \frac{\sigma_2^2}{2} t^2} = e^{(\mu_1+\mu_2)t + \frac{\sigma_1^2+\sigma_2^2}{2} t^2}, \quad t \in \mathbb{R}.$$

Nun wissen wir aber auch, dass

$$\mathcal{M}_Y(t) = e^{(\mu_1+\mu_2)t + \frac{\sigma_1^2+\sigma_2^2}{2} t^2}, \quad t \in \mathbb{R},$$

für $Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Also gilt $X_1 + X_2 \sim Y$ nach Satz 4.3.17 und damit gilt für die Summe $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Mit ähnlichen Rechnungen kann man (siehe Übungen) mit momenterzeugenden Funktionen auch zeigen, dass

- die Summe unabhängiger Exponentialverteilungen gammaverteilt ist,
- die Summe unabhängiger Gammaverteilungen wieder gammaverteilt ist,
- die Summe unabhängiger Bernoulliverteilungen binomialverteilt ist (haben wir oben schon mit der Faltungsformel berechnet),

Das Argument ist immer identisch: Berechne die momenterzeugende Funktion der Summe als Produkt der momenterzeugenden Funktionen der einzelnen (Unabhängigkeit) und hoffe, dass das Produkt eine bereits bekannt momenterzeugende Funktion ist. Dann kann die Verteilung der Summe mittels Satz 4.3.17 identifiziert werden.

Bemerkung 4.3.21. ▶

- Der Trick mit der Normalverteilung funktioniert aufgrund der Faktorisierungseigenschaft der Exponentialfunktion. Wir sehen also, dass der Trick auf Situationen beschränkt ist, wenn in einer nützlichen Form Potenzen auftauchen.
- Die Vorlesung ist hier etwas irreführend. Satz 4.3.17 ist kein Allheilmittel. In der (mathematischen) Realität sind exponentielle Momente sehr oft unendlich und wir können nicht mit der momenterzeugenden Funktion arbeiten. In dieser Vorlesung sind alle Verteilungen außer Cauchy und Exponentiell unproblematisch, weil alle exponentiellen Momente existieren. In der Wahrscheinlichkeitstheorie werden wir das Problem lösen, indem wir die Momenterzeugende Funktionen durch sogenannte charakteristische Funktionen $\varphi(t) = \mathbb{E}[e^{itX}]$ ersetzen.

4.4 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

Wir kommen nun zu bedingten Wahrscheinlichkeiten, die aus der Schule vielleicht bekannt sind. In dieser Vorlesung spielen bedingte Wahrscheinlichkeiten noch keine zentrale Rolle, das ändert sich in späteren Vorlesungen sehr, wenn zum Beispiel Markovketten angeschaut werden.

Definition 4.4.1. ▶ Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$. Dann heißt

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

bedingte Wahrscheinlichkeit von A gegeben B .

Lemma 4.4.2. ▶ Für $B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ ist $A \mapsto \mathbb{P}(A|B) =: \mathbb{P}_B(A)$ ein Maß auf (Ω, \mathcal{A}) .

Beweis. Wir rechnen die definierenden Eigenschaften eines Maßes nach:

- $\mathbb{P}_B(A) \geq 0$ für alle $A \in \mathcal{A}$ ist klar, weil \mathbb{P} ein Maß ist.
- Weil \mathbb{P} ein Maß ist, gilt auch

$$\mathbb{P}_B(\emptyset) = \mathbb{P}(\emptyset|B) = \frac{\mathbb{P}(\emptyset \cap B)}{\mathbb{P}(B)} = 0.$$

- σ -Additivität: Seien $A_1, A_2, \dots \in \mathcal{A}$ disjunkt, so gilt wegen der σ -Additivität von \mathbb{P}

$$\begin{aligned} \mathbb{P}_B\left(\bigcup_{k=1}^{\infty} A_k\right) &\stackrel{\text{Def.}}{=} \frac{\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k \cap B\right)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}\left(\bigcup_{k=1}^{\infty} (A_k \cap B)\right)}{\mathbb{P}(B)} \\ &\stackrel{\sigma\text{-Add.}}{=} \frac{\sum_{k=1}^{\infty} \mathbb{P}(A_k \cap B)}{\mathbb{P}(B)} = \sum_{k=1}^{\infty} \mathbb{P}_B(A_k). \end{aligned}$$

□

Wir kommen nun zu extrem wichtigen Rechenregeln, obwohl diese ganz einfach aus der Definition folgen:

Satz 4.4.3. 

- (i) **Multiplikationsregel:** Für $A_1, \dots, A_n \in \mathcal{A}$ mit $\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) > 0$ gilt

$$\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \dots \cdot \mathbb{P}\left(A_n \mid \bigcap_{k=1}^{n-1} A_k\right).$$

- (ii) **Formel der totalen Wahrscheinlichkeit:** Ist $B_1, \dots, B_n \in \mathcal{A}$ eine disjunkte Zerlegung von Ω , d. h. $\bigcup_{k=1}^n B_k = \Omega$, mit $\mathbb{P}(B_k) > 0$ für alle $k = 1, \dots, n$, so gilt

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(B_k) \mathbb{P}(A|B_k), \quad \forall A \in \mathcal{A}.$$

- (iii) **Bayes-Formel:** Mit B_1, \dots, B_n aus (ii) gilt für $A \in \mathcal{A}$ mit $\mathbb{P}(A) > 0$

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)}, \quad \forall B \in \mathcal{A},$$

oder

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B) \cdot \mathbb{P}(B)}{\sum_{k=1}^n \mathbb{P}(B_k) \mathbb{P}(A|B_k)}, \quad \forall B \in \mathcal{A}.$$

Beweis. (i) Induktion über n :

IA: $n = 2$ folgt direkt aus der Definition der bedingten Wahrscheinlichkeit.

IV: Die Behauptung gelte für ein *beliebiges*, aber festes $n \in \mathbb{N}$.

IS: Wenn wir nun die Induktionsvoraussetzung und den Fall $n = 2$ nutzen, so bekommen wir

$$\begin{aligned} \mathbb{P}\left(\bigcap_{k=1}^{n+1} A_k\right) &= \mathbb{P}\left(\bigcap_{k=1}^n A_k \cap A_{n+1}\right) \\ &= \mathbb{P}\left(\bigcap_{k=1}^n A_k\right) \mathbb{P}\left(A_{n+1} \mid \bigcap_{k=1}^n A_k\right) \\ &\stackrel{\text{IV}}{=} \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \dots \cdot \mathbb{P}\left(A_{n+1} \mid \bigcap_{k=1}^n A_k\right) \end{aligned}$$

und das ist gerade die Aussage.

(ii) Zunächst folgt direkt aus der Definition

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

Setzen wir dies in folgende Rechnung ein, so bekommen wir die Aussage:

$$\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}\left(A \cap \bigcup_{k=1}^n B_k\right) = \mathbb{P}\left(\bigcup_{k=1}^n (A \cap B_k)\right) \stackrel{\sigma\text{-Add.}}{=} \sum_{k=1}^n \mathbb{P}(A \cap B_k).$$

(iii) Die einfache Bayes Formel folgt direkt durch Einsetzen der Definition und Erweitern:

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

Die zweite Formel folgt aus der ersten Formel durch Ersetzen des Nenners durch die Formel der totalen Wahrscheinlichkeit. \square

Kommen wir nun zu zwei klassischen Anwendungen. Hier ist allerdings etwas Vorsicht angesagt, das ist alles etwas wild (funktioniert aber). Gemäß Definition brauchen wir für bedingte Wahrscheinlichkeiten einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. In vielen Anwendungen außerhalb der Mathematik werden die Formeln allerdings auch genutzt, ohne einen Wahrscheinlichkeitsraum hinzuschreiben. Nennen wir das vielleicht „heuristischen Gebrauch von Wahrscheinlichkeiten“. Das ist natürlich nicht sehr schön, allerdings sind solche Aussagen extrem wichtig.

Beispiel 4.4.4.

(i) Wir ziehen aus vier blauen und drei weißen Kugeln zweimal ohne Zurücklegen. Was ist die Wahrscheinlichkeit, zwei blaue zu ziehen? Machen wir das ganze zunächst „heuristisch“. Das ist ein zweistufiges Experiment. Im ersten Versuch ziehen wir blau mit Wahrscheinlichkeit $\frac{4}{7}$. Mit dem Wissen im ersten Zug blau gezogen zu haben, ist das „bedingte Ziehen“ im zweiten Schritt ein Ziehen aus drei blauen und drei weißen Kugeln. Die bedingte Wahrscheinlichkeit im zweiten Schritt blau zu ziehen, gegeben das erste Ziehen gab eine blaue Kugel, ist also $\frac{3}{6}$. Mit der Multiplikationsregel folgt

$$\begin{aligned} & \mathbb{P}(\text{beide Kugeln blau}) \\ &= \mathbb{P}(\text{erste Kugel blau}) \cdot \mathbb{P}(\text{zweite Kugel blau} \mid \text{erste Kugel blau}) \\ &= \frac{4}{7} \cdot \frac{3}{6} = \frac{2}{7}. \end{aligned}$$

Das funktioniert ganz entspannt, ist aber schon etwas kritisch. Wir nutzen hier ganz intuitiv den Begriff der bedingten Wahrscheinlichkeit in einem interpretierten Sinn (Änderung des Modells, eine Kugel weniger). Dann haben wir die Multiplikationsregel genutzt, die aufgrund unseres Beweises und damit aufgrund der mathematischen Definition der bedingten Wahrscheinlichkeit stimmt. Nun ist allerdings nicht so ganz klar, warum der intuitive Begriff der bedingten Wahrscheinlichkeit (geändertes Experiment) mit der mathematischen Definition $\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ überhaupt etwas zu tun hat. Machen wir das ganze zur Beruhigung also nochmal mathematisch penibel genau. Schreiben wir zunächst ein Modell hin, das wir als Modell für das zweifache Würfeln plausibel finden. Dazu sei

$$\Omega = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{\text{blau}, \text{weiß}\}\}$$

und $\mathcal{A} = \mathcal{P}(\Omega)$. Als Wahrscheinlichkeiten definieren wir

$$\mathbb{P}(\{\omega_1, \omega_2\}) = \begin{cases} \frac{4}{7} \cdot \frac{3}{6} & : \omega_1 = \omega_2 = \text{blau} \\ \frac{3}{7} \cdot \frac{2}{6} & : \omega_1 = \omega_2 = \text{weiß} \\ \frac{4}{7} \cdot \frac{3}{6} & : \omega_1 = \text{blau}, \omega_2 = \text{weiß} \\ \frac{3}{7} \cdot \frac{4}{6} & : \omega_1 = \text{weiß}, \omega_2 = \text{blau} \end{cases}.$$

Mit den Ereignissen $A = \{(\omega_1, \omega_2): \omega_1 = \text{blau}\}$, $B = \{(\omega_1, \omega_2): \omega_2 = \text{blau}\}$ wollen wir die Wahrscheinlichkeit von $A \cap B$ bestimmen. Mit der Multiplikationsregel folgt

$$\mathbb{P}(\text{ziehe zweimal blau}) = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) = \frac{4}{7} \cdot \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{2}{7}.$$

Das macht natürlich auch wieder nicht so richtig viel Sinn, wir hätten die Wahrscheinlichkeit von $A \cap B$ auch ohne die Multiplikationsregel „ablesen“ können. Dennoch ist das vielleicht eine gute Beruhigung: Wenn wir wollen, können wir ein rigoroses Modell hinschreiben, in dem die Multiplikationsformel rigoros gemacht werden kann. Für die Schulanwendungen ist das natürlich viel zu kompliziert.

- (ii) Nun ein Beispiel für die Bayes Formel, ein medizinischer Test, z. B. ein Aidstest. Wir machen das jetzt wieder in der „heuristischen“ Art, wer will, kann sich wieder ein sauberes Modell definieren. Wir nehmen an, dass die Wahrscheinlichkeiten folgender Ereignisse bekannt sind:

- 1% der Bevölkerung ist tatsächlich krank.
- Test ist mit 98% positiv, wenn eine Person krank ist.
- Test ist mit 5% positiv, wenn eine Person nicht krank ist, das nennt man false-positive (Fehlalarm).

Gesucht ist die Wahrscheinlichkeit gesund zu sein, obwohl der Test positiv ist. Mit der Bayes Formel gilt

$$\begin{aligned} & \mathbb{P}(\text{krank} \mid \text{Test positiv}) \\ &= \mathbb{P}(\text{Test positiv} \mid \text{krank}) \cdot \frac{\mathbb{P}(\text{krank})}{\mathbb{P}(\text{Test positiv})} \\ &= \frac{\mathbb{P}(\text{Test positiv} \mid \text{krank})\mathbb{P}(\text{krank})}{\mathbb{P}(\text{Test positiv} \mid \text{krank})\mathbb{P}(\text{krank}) + \mathbb{P}(\text{Test positiv} \mid \text{gesund})\mathbb{P}(\text{gesund})} \\ &= \frac{0,98 \cdot 0,01}{0,98 \cdot 0,01 + 0,05 \cdot 0,99} \\ &= 0,165. \end{aligned}$$

Das ergibt dann

$$\mathbb{P}(\text{gesund} \mid \text{Test positiv}) = 1 - 0,165 = 0,835,$$

was erschreckend hoch ist! Die wichtige take-home message ist also: Wird etwas sehr unwahrscheinliches getestet, dominieren falsche Tests und man muss bei positiven Tests unbedingt weitere Tests machen. Ein weiteres sehr wichtiges Beispiel sind pränatale Tests bei ungeborenen Kindern, z. B. Tests auf Trisomie 21 (auf die moralische Frage wollen wir hier natürlich nicht eingehen!).

Weiter geht es jetzt mit einer mathematischen Definition von Unabhängigkeit von Ereignissen. Ihr habt sicherlich eine naive Vorstellung „Hat nichts mit einander zu tun“. Beispielsweise wären die Ereignisse „Meine Kaffeemaschine geht morgen kaputt“ und „Es regnet morgen in Thailand“ vermutlich unabhängig. Hier ist eine Definition:

Definition 4.4.5.  Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A, B \in \mathcal{A}$. Die Ereignisse A und B heißen **unabhängig**, falls $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Aufgrund der Definition der bedingten Wahrscheinlichkeit sind, sofern $\mathbb{P}(B) > 0$ gilt, A und B unabhängig genau dann, wenn $\mathbb{P}(A|B) = \mathbb{P}(A)$. Das passt also zur Begriffsbildung: Zwei Ereignisse sind genau dann unabhängig, falls die Wahrscheinlichkeit des einen sich nicht ändert, wenn das Eintreten des anderen gegeben ist.

Oft braucht man auch die Unabhängigkeit mehrerer Ereignisse. Für endlich viele ist klar was man macht, die Wahrscheinlichkeit des endlichen Schnittes soll natürlich das endlich Produkt sein. Für unendlich viele ist das problematisch, darum führt man alles auf endlich viele zurück:

Definition 4.4.6. ▶ Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $A_i \in \mathcal{A}$, $i \in I$, und I eine beliebige Indexmenge.

(i) Die Ereignisse $(A_i)_{i \in I}$ heißen **unabhängig**, falls

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i), \quad \forall J \subseteq I \text{ mit } \#J < \infty.$$

(ii) Die Ereignisse $(A_i)_{i \in I}$ heißen **paarweise unabhängig**, falls

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j), \quad \forall i \neq j.$$

Selbstverständlich impliziert Unabhängigkeit die paarweise Unabhängigkeit, statt aller endlicher Teilmengen J von I werden schließlich nur alle Teilmengen mit $\#J = 2$ gewählt. Die Umkehrung gilt nicht:

Warnung 4.4.7. ▶ Paarweise Unabhängigkeit impliziert im Allgemeinen nicht Unabhängigkeit. Kleine Mengen reichen schon aus, um Gegenbeispiel anzugeben, siehe Übungsblatt.

In Worten ausgedrückt heißt die Unabhängigkeit auch „Die Ereignisse $(A_i)_{i \in I}$ sind unabhängig, falls jede Wahl von endlich vielen Ereignissen A_{i_1}, \dots, A_{i_n} unabhängig ist“.

Als nächstes wollen wir die Unabhängigkeit von σ -Algebren und Zufallsvariablen thematisieren. Dazu zunächst eine allgemeinere Definition:

Definition 4.4.8. ▶ Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{E}_i)_{i \in I}$ eine Familie von Teilmengen $\mathcal{E}_i \subseteq \mathcal{A}$ der σ -Algebra für eine beliebige Indexmenge I . Dann heißen die $(\mathcal{E}_i)_{i \in I}$ unabhängig, falls die Ereignisse $(A_i)_{i \in I}$ unabhängig sind, und zwar für alle $A_i \in \mathcal{E}_i$.

Die Definition wird insbesondere für σ -Algebren verwendet. Wie bei der Messbarkeit fragen wir uns hier, ob wir die Unabhängigkeit auf Erzeuger reduzieren können. Wie immer funktioniert das, zumindest wenn der Erzeuger \cap -stabil ist:

Proposition 4.4.9. ▶ Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathcal{E}_i \subseteq \mathcal{A}$ für alle $i \in I$. Sind alle \mathcal{E}_i \cap -stabil, so gilt:

$$(\mathcal{E}_i)_{i \in I} \text{ unabhängig} \quad \Leftrightarrow \quad (\sigma(\mathcal{E}_i))_{i \in I} \text{ unabhängig.}$$

Beweis.

„ \Leftarrow “: Klar nach Definition, weil $\mathcal{E}_i \subseteq \sigma(\mathcal{E}_i)$ gilt. Die Unabhängigkeit von Mengensystemen bedeutet schließlich, dass die Unabhängigkeit für alle Auswahlen von Teilmengen gilt. Gilt dies für mehr Möglichkeiten, so natürlich auch für weniger Möglichkeiten.

„ \Rightarrow “: Ohne Einschränkung sei I endlich weil die Definition der Unabhängigkeit nur auf endlichen Teilmengen $J \subseteq I$ beruht. Nennen wir die Mengen $\mathcal{E}_1, \dots, \mathcal{E}_n$. Wir zeigen, dass dann auch $\sigma(\mathcal{E}_1), \dots, \sigma(\mathcal{E}_n)$ unabhängig sind. Dazu zeigen wir zunächst:

$$\mathcal{D} := \{E \in \mathcal{A} : \{E\}, \mathcal{E}_2, \dots, \mathcal{E}_n \text{ sind unabhängig}\} \text{ ist ein Dynkin-System.}$$

Um das zu zeigen, checken wir die definierenden Eigenschaften eines Dynkin-Systems:

(i) Aufgrund der Definition 4.4.8 müssen wir zeigen, dass für alle $A_2 \in \mathcal{E}_2, \dots, A_n \in \mathcal{E}_n$ die Mengen Ω, A_2, \dots, A_n unabhängig sind, die Wahrscheinlichkeit vom Schnitt also zur Wahrscheinlichkeiten der einzelnen Ereignisse faktorisiert. Das folgt aber direkt aus der angenommenen Unabhängigkeit von $\mathcal{E}_1, \dots, \mathcal{E}_n$:

$$\begin{aligned} \mathbb{P}(\Omega \cap A_2 \cap \dots \cap A_n) &= \mathbb{P}(A_2 \cap \dots \cap A_n) \\ &= \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n) \\ &= 1 \cdot \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n) = \mathbb{P}(\Omega) \cdot \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n). \end{aligned}$$

Damit ist $\Omega \in \mathcal{D}$.

- (ii) Nun zur Abgeschlossenheit bezüglich Komplementbildung. Wir argumentieren wie im ersten Schritt. Sei dazu $E \in \mathcal{D}$ und seien $A_2 \in \mathcal{E}_2, \dots, A_n \in \mathcal{E}_n$ beliebig. Weil $\Omega = E \cup E^C$ ergibt sich

$$\begin{aligned} & \mathbb{P}(E^C \cap A_2 \cap \dots \cap A_n) \\ & \stackrel{\sigma\text{-Add.}}{=} \mathbb{P}(\Omega \cap A_2 \cap \dots \cap A_n) - \mathbb{P}(E \cap A_2 \cap \dots \cap A_n) \\ & \stackrel{\Omega, E \in \mathcal{D}}{=} \mathbb{P}(\Omega) \cdot \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n) - \mathbb{P}(E) \cdot \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n) \\ & = (\mathbb{P}(\Omega) - \mathbb{P}(E)) \cdot \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n) \\ & = \mathbb{P}(E^C) \cdot \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n). \end{aligned}$$

Also sind E^C, A_2, \dots, A_n unabhängige Ereignisse und damit ist $E^C \in \mathcal{D}$.

- (iii) Das Argument für die Vereinigungen geht genau wie für die Komplemente.

Jetzt beenden wir den Beweis. Aufgrund der Annahme $\mathcal{E}_1, \dots, \mathcal{E}_n$ unabhängig, gilt für jede Menge $E \in \mathcal{E}_1$ auch $E \in \mathcal{D}$. Also gilt $\mathcal{E}_1 \subseteq \mathcal{D}$. Daraus folgt mit dem Hauptsatz für Dynkinsysteme, Satz 1.2.11, wie immer (Bildung des kleinsten Dynkin-Systems ist monoton)

$$\sigma(\mathcal{E}_1) \stackrel{\cap\text{-stabil}}{=} d(\mathcal{E}_1) \subseteq d(\mathcal{D}) = \mathcal{D}.$$

Weil also $\sigma(\mathcal{E}_1) \subseteq \mathcal{D}$ gilt, folgt aus der Definition von \mathcal{D} , dass $\sigma(\mathcal{E}_1), \mathcal{E}_2, \dots, \mathcal{E}_n$ unabhängig sind. Iterativ ersetzen wir nun Schritt für Schritt mit einem analogen Argument ein \mathcal{E}_k nach dem anderen durch $\sigma(\mathcal{E}_k)$, indem wir genau wie oben zeigen, dass alle

$$\mathcal{D}_k := \{E \in \mathcal{A} : \sigma(\mathcal{E}_1), \sigma(\mathcal{E}_2), \dots, \sigma(\mathcal{E}_{k-1}), \{E\}, \mathcal{E}_{k+1}, \dots, \mathcal{E}_n \text{ sind unabhängig}\}$$

Dynkin-Systeme sind.

□

Vorlesung 24

Nach der Unabhängigkeit von Ereignissen und Zufallsvariablen kommen wir nun zu einer alternativen Definition der Unabhängigkeit von Zufallsvariablen. Anstatt die Faktorisierung der gemeinsamen Verteilung zu fordern, kann man auch fordern, dass die erzeugten σ -Algebren (siehe Definition 2.1.8) unabhängig sind:

Definition 4.4.10. Für Zufallsvariablen $(X_i)_{i \in I}$ auf $(\Omega, \mathcal{A}, \mathbb{P})$ definiert man: $(X_i)_{i \in I}$ sind unabhängig, falls die erzeugten σ -Algebren $(\sigma(X_i))_{i \in I}$ unabhängig sind.

Weil $\sigma(X_i) \stackrel{\text{Def.}}{=} \{X_i^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\} = \sigma(\{X_i^{-1}((-\infty, t]) : t \in \mathbb{R}\})$, können wir direkt folgern, dass die Unabhängigkeit auch durch die gemeinsame Verteilungsfunktion definiert werden kann.

Korollar 4.4.11. Für Zufallsvariablen X_1, \dots, X_d auf $(\Omega, \mathcal{A}, \mathbb{P})$ stimmt die neue Definition der Unabhängigkeit mit der alten überein. Wir können Unabhängigkeit also auf verschiedene Arten beschreiben:

$$\begin{aligned} & X_1, \dots, X_d \text{ sind unabhängig} \\ \Leftrightarrow & \sigma(X_1), \dots, \sigma(X_d) \text{ sind unabhängige } \sigma\text{-Algebren} \\ \Leftrightarrow & F_X(t_1, \dots, t_d) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_d}(t_d), \quad \forall t_i \in \mathbb{R} \\ \Leftrightarrow & \mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \mathbb{P}(X_1 \in A_1) \cdot \dots \cdot \mathbb{P}(X_d \in A_d), \quad \forall A_i \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

Beweis. Um die vorherige Proposition anzuwenden, seien $\mathcal{E}_i := \{X_i \leq t\} : t \in \mathbb{R}$ für $i = 1, \dots, d$. Also gilt $\sigma(\mathcal{E}_i) = \sigma(X_i)$ und die \mathcal{E}_i sind \cap -stabil. Überlegen wir einmal schnell, warum $\sigma(\mathcal{E}_i) = \sigma(X_i)$ gilt. Die Richtung „ \subseteq “ gilt, weil $\mathcal{E}_i \subseteq \sigma(X_i)$ aufgrund der Definition von $\sigma(X_i)$. Die Richtung „ \supseteq “ gilt, weil wegen Proposition 2.1.4 X_i ($\sigma(\mathcal{E}_i), \mathcal{B}(\mathbb{R})$)-messbar ist und $\sigma(X_i)$ die kleinste σ -Algebra ist, bezüglich derer X_i messbar ist.

Damit gilt aufgrund der Definitionen der Unabhängigkeit von Mengensystemen und Ereignissen

$$\begin{aligned}
 & \sigma(X_1), \dots, \sigma(X_d) \text{ unabhängig} \\
 \stackrel{4.4.9}{\Leftrightarrow} & \mathcal{E}_1, \dots, \mathcal{E}_d \text{ unabhängig} \\
 \Leftrightarrow & E_1, \dots, E_d \text{ unabhängig für alle } E_1 \in \mathcal{E}_1, \dots, E_d \in \mathcal{E}_d \\
 \Leftrightarrow & \mathbb{P}(\{X_1 \leq t_1\} \cap \dots \cap \{X_d \leq t_d\}) = \mathbb{P}(\{X_1 \leq t_1\}) \cdots \mathbb{P}(\{X_d \leq t_d\}), \quad t_1, \dots, t_d \in \mathbb{R} \\
 \stackrel{\text{Notation}}{\Leftrightarrow} & \mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d) = \mathbb{P}(X_1 \leq t_1) \cdots \mathbb{P}(X_d \leq t_d), \quad t_1, \dots, t_d \in \mathbb{R} \\
 \stackrel{\text{Def. VF}}{\Leftrightarrow} & F_X(t_1, \dots, t_d) = F_{X_1}(t_1) \cdots F_{X_d}(t_d), \quad t_1, \dots, t_d \in \mathbb{R}.
 \end{aligned}$$

□

Bitte beachtet die Notation im letzten Beweis. In der Stochastik bevorzugen wir immer die Notation $\mathbb{P}(X_1 \leq t_1, \dots, X_d \leq t_d)$, wir nutzen praktisch nie die ausführliche Schreibweise $\mathbb{P}(\{X_1 \leq t_1\} \cap \dots \cap \{X_d \leq t_d\})$. Das liegt einfach nur daran, dass sich die erste Notation viel natürlicher lesen lässt. Am besten gewöhnt ihr euch direkt die kompakte Schreibweise an.

Im nächsten Abschnitt besprechen wir Konvergenzen von Folgen von Zufallsvariablen. Wir werden insbesondere Folgen unabhängiger Zufallsvariablen nutzen. Unsere ursprüngliche Definition war nur für endlich viele Zufallsvariablen, die Definition dieses Abschnittes funktioniert auch für unendlich viele Zufallsvariablen (man testet die Eigenschaft einfach für alle endlichen Teilmengen). Genau so wollen wir ab jetzt die Unabhängigkeit einer ganzen Folgen von Zufallsvariablen definieren.

Definition 4.4.12.  Ist X_1, X_2, \dots eine (unendliche) Folge von Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$, so heißt die Folge unabhängig, falls eine der äquivalenten Eigenschaften gilt:

- (i) Für alle $n \in \mathbb{N}$ sind $\sigma(X_1), \dots, \sigma(X_n)$ unabhängige σ -Algebren.
- (ii) Für alle $n \in \mathbb{N}$ gilt: $\mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) = \prod_{k=1}^n \mathbb{P}(X_k \leq t_k)$, $t_1, \dots, t_n \in \mathbb{R}$.

Wie für Zufallsvariablen und Zufallsvektoren ist es auch für Folgen von Zufallsvariablen nicht klar, dass es diese überhaupt gibt. In der Tat kann man auch in diesem Fall eine kanonische Konstruktion angeben, die uns die Existenz von Folgen unabhängiger Zufallsvariablen gibt. Der kanonische Wahrscheinlichkeitsraum besteht ganz analog aus den Werten, die angenommen werden. Dies war zunächst \mathbb{R} , dann \mathbb{R}^d und ist nun \mathbb{R}^∞ (die Menge der reellen Folgen). Die kanonische σ -Algebra ist die passende „Borel“- σ -Algebra und die Folge der Zufallsvariablen ist durch die Identitätsabbildung gegeben. Das Thema gehört eigentlich nicht in die Stochastik 1 sondern in die Wahrscheinlichkeitstheorie 1. Daher skizzieren wir die Konstruktion nur ganz kurz. Ihr solltet euch jedoch merken, dass es eine kanonische Konstruktion gibt und insbesondere Folgen von unabhängigen Zufallsvariablen existieren

Satz 4.4.13.  [Kanonische Konstruktion von Folgen unabhängiger Z.V.]

Seien F_1, F_2, \dots Verteilungsfunktionen, so existieren ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und unabhängige Zufallsvariablen X_1, X_2, \dots auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit $X_i \sim F_i$.

Beweis. Man kann ganz analog zu \mathbb{R} und zum \mathbb{R}^d eine kanonische Konstruktion angeben. Hier nur eine Skizze für die übermotivierten Studis, die Konstruktion wird in Ruhe in einer der Vorlesungen im Master thematisiert. Alle anderen merken sich aber bitte die Aussage des Satzes, sonst wäre die Vorlesung an dieser Stelle vorbei!

- $\Omega := \{(\omega_n)_{n \in \mathbb{N}} : \omega_n \in \mathbb{R}\}$, die Menge der „reellen Folgen“. Die ω sind also Folgen, oder unendlich lange Vektoren. Als Analogie zum \mathbb{R}^d schreibt man auch \mathbb{R}^∞ .
- $\mathcal{A} := \mathcal{B}(\mathbb{R}^\infty) := \sigma(\{B_1 \times \dots \times B_d \times \mathbb{R} \times \dots : d \in \mathbb{N}, B_1, \dots, B_d \in \mathcal{B}(\mathbb{R})\})$
- $\mathbb{P} := \mathbb{P}_{F_1} \otimes \mathbb{P}_{F_2} \otimes \dots$ sei das unendliche Produktmaß, das auf dem Erzeuger von $\mathcal{B}(\mathbb{R}^\infty)$ festgelegt ist durch $\mu(B_1 \times \dots \times B_d \times \mathbb{R} \times \dots) = \mathbb{P}_{F_1}(B_1) \cdots \mathbb{P}_{F_d}(B_d)$.

$$\bullet (X_1(\omega), X_2(\omega), \dots) := (\omega_1, \omega_2, \dots) = \omega$$

Um zu zeigen, dass es ein unendliches Produktmaß auf $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ auch gibt, kann man den Fortsetzungssatz von Carathéodory anwenden. Das ist etwas hässlich. Wir haben nun also einen Wahrscheinlichkeitsraum und eine Folge von Zufallsvariablen, die einfach nur für ein ω die Koordinaten ausgibt (genau wie im \mathbb{R}^d , vergleiche den Beweis von Satz 4.2.15). Die Unabhängigkeit der konstruierten Folge folgt direkt aus der Produkteigenschaft des Produktmaßes, weil aufgrund der Definition der Zufallsvariablen als Koordinatenabbildungen

$$\begin{aligned} \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n) &\stackrel{\text{Def } X}{=} \mathbb{P}((-\infty, t_1] \times \dots \times (-\infty, t_n] \mathbb{R} \times \dots) \\ &\stackrel{\text{Produktmaß}}{=} \mathbb{P}_{F_1}((-\infty, t_1]) \cdots \mathbb{P}_{F_n}((-\infty, t_n]) \\ &= \mathbb{P}(X_1 \leq t_1) \cdots \mathbb{P}(X_n \leq t_n). \end{aligned}$$

Auch sofort folgt durch Einsetzen, dass die Randverteilungen $\mathbb{P}(X_i \leq t) = F_{X_i}(t)$ erfüllen. \square

Definition 4.4.14. ▶ Sind alle F_i gleich, so nennen wir die Folge unabhängiger Zufallsvariablen X_1, X_2, \dots aus Satz 4.4.13 eine u.i.v. Folge mit $X_1 \sim F_1$.

In den nächsten Kapitel schauen wir uns Konvergenzeigenschaften von u.i.v. Folgen an, insbesondere das Gesetz der großen Zahlen und den zentralen Grenzwertsatz.

4.5 Konvergenz von Folgen von Zufallsvariablen

Bevor wir zu den zentralen Konvergenzsätzen kommen, müssen wir uns überlegen, was Konvergenz von Folgen von Zufallsvariablen überhaupt bedeutet. Das ist in der Tat gar nicht so klar, es gibt verschiedene Begriffe:

Definition 4.5.1. ▶ [Vier Konvergenzarten in der Stochastik]

Für eine Zufallsvariable X und eine Folge X_1, X_2, \dots von Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ definiert man

(i) „ X_n konvergiert **stochastisch** (oder **in Wahrscheinlichkeit**) gegen X “, man schreibt

$$X_n \xrightarrow{P} X, \quad n \rightarrow \infty,$$

falls für alle $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

(ii) „ X_n konvergiert **in \mathcal{L}^p** gegen X “ (oder **im p -ten Mittel**) für $p \geq 1$, man schreibt

$$X_n \xrightarrow{\mathcal{L}^p} X, \quad n \rightarrow \infty,$$

falls

$$\mathbb{E}[|X_n - X|^p] \rightarrow 0, \quad n \rightarrow \infty.$$

(iii) „ X_n konvergiert **fast sicher** gegen X “, man schreibt

$$X_n \xrightarrow{\text{f.s.}} X, \quad n \rightarrow \infty,$$

falls

$$\mathbb{P}(X_n \rightarrow X) := \mathbb{P}(\{\omega : X_n(\omega) \rightarrow X(\omega), n \rightarrow \infty\}) = 1.$$

(iv) „ X_n konvergiert **in Verteilung** (oder **schwach**) gegen X “, man schreibt

$$X_n \xrightarrow{(d)} X, \quad n \rightarrow \infty,$$

falls für alle $f : \mathbb{R} \rightarrow \mathbb{R}$ stetig und beschränkt

$$\mathbb{E}[f(X_n)] \rightarrow \mathbb{E}[f(X)], \quad n \rightarrow \infty.$$

Nur die ersten drei Konvergenzbegriffe benötigen wirklich, dass die Zufallsvariablen auf dem selben Wahrscheinlichkeitsraum definiert sind. Die Konvergenz in Verteilung ist strukturell anders weil die Zufallsvariablen X_n nicht direkt mit der Grenzzufallsvariablen X „verglichen“ werden, es wird nicht $|X_n(\omega) - X(\omega)|$ berechnet. Die Konvergenz in Verteilung hängt nur von den Verteilungen ab, vergleiche dazu die Berechnungsformel des Erwartungswertes mit dem Transformationssatz, Lemma 4.1.10.

Bemerkung 4.5.2. 

- (i) Warnung: Die Konvergenzen sind nicht durch Metriken definiert worden, d. h. übliche Tricks aus der Analysis (z. B. Δ -Ungleichung, Eindeutigkeit von Grenzwerten, ...) gelten nicht einfach so! Konkretes Beispiel: Nur wenn man einen Quotientenraum mit fast sicher gleichen Zufallsvariablen bildet, ist Konvergenz im p -ten Mittel eine Normenkonvergenz und die Eindeutigkeit von Grenzwerten gilt (siehe Übungsaufgaben).
- (ii) Zwei Konvergenzarten sind uns schon bekannt:

- fast sichere Konvergenz ist schon von messbaren Funktionen bekannt,
- p -tes Mittel ist schon von $(\mathcal{L}^p, \|\cdot\|_p)$ bekannt.

Um ein Gefühl für die Konvergenzarten zu bekommen, sind Beispiele äusserst nützlich. Viele nützliche Beispiele können ganz explizit hingeschrieben werden.

Beispiel 4.5.3.  Seien X_1, X_2, \dots Zufallsvariablen auf irgendeinem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit

$$\mathbb{P}(X_n = e^n) = \frac{1}{n}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n},$$

so gelten:

$\mathbf{X}_n \xrightarrow{\mathbb{P}} \mathbf{0}, n \rightarrow \infty$, weil

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(X_n = e^n) = \frac{1}{n} \rightarrow 0, \quad n \rightarrow \infty.$$

$\mathbf{X}_n \not\xrightarrow{\mathcal{L}^p} \mathbf{0}, n \rightarrow \infty$, für alle $p \geq 1$, weil

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}[|X_n|^p] = e^{pn} \frac{1}{n} + 0^p \left(1 - \frac{1}{n}\right) = \frac{e^{pn}}{n} \rightarrow +\infty, \quad n \rightarrow \infty.$$

Das nächste Beispiel ist sehr anschaulich. Dazu beachten wir, dass die Einschränkung des Lebesguemaßes auf $[0, 1]$ ein Wahrscheinlichkeitsmaß ist. Wir können also viele Beispiele basteln, wenn wir uns als Zufallsvariablen einfach messbare Funktionen (z. B. Indikatorfunktionen oder stetige Funktionen) auf $[0, 1]$ wählen. Das hat den Vorteil, dass die Begriffe durch Skizzen sehr anschaulich gemacht werden können. So ist zum Beispiel die fast sichere Konvergenz die punktweise Konvergenz (bis auf eine Nullmenge) und die \mathcal{L}^p -Konvergenz ist die Konvergenz der Flächeninhalte der Differenzfunktion (hoch p) weil $\mathbb{E}[X] = \int_0^1 X(\omega) d\omega$. Sieht wegen $d\omega$ vielleicht blöd aus, ist aber einfach nur das ganz normale Integral auf $[0, 1]$.

Beispiel 4.5.4.  Sei $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ und $\mathbb{P} = \lambda_{[0,1]}$ das Lebesgue Maß auf $[0, 1]$. Schauen wir uns als Beispiel $X \equiv 0$ (Nullfunktion) und die Folge

$$X_n = \mathbf{1}_{\left(\frac{m}{2^k}, \frac{m+1}{2^k}\right]}$$

an, wobei $m, k \in \mathbb{N}$ die eindeutigen natürlichen Zahlen mit $n = 2^k + m$ und $m < 2^k$ sind. In Worten (am besten skizziert ihr die Funktionen) schieben wir für wachsendes n einfach nur Indikatorfunktionen von links nach rechts durch $[0, 1]$, wobei die Breite der Indikatorfunktionen schmaler wird: $\mathbf{1}_{(0,1)}$, $\mathbf{1}_{(0, \frac{1}{2}]}$, $\mathbf{1}_{(\frac{1}{2}, 1]}$, $\mathbf{1}_{(0, \frac{1}{4}]}$, $\mathbf{1}_{(\frac{1}{4}, \frac{1}{2}]}$, ... Mit dieser Folge gelten:

$\mathbf{X}_n \xrightarrow{\mathcal{L}^p} \mathbf{0}, n \rightarrow \infty$, weil

$$\mathbb{E}[|X_n - X|^p] = \mathbb{E}[|X_n|^p] = \int_{\Omega} X_n^p(\omega) d\mathbb{P}(\omega) = \int_0^1 \mathbf{1}_{(\frac{m}{2^k}, \frac{m+1}{2^k}]}(\omega) d\omega = \frac{1}{2^k}.$$

Weil mit $n \rightarrow \infty$ auch $k \rightarrow \infty$ gilt, konvergiert die Folge also im p -ten Mittel gegen 0. Warum ist das auch anschaulich klar? $\mathbb{E}[|X_n|^p]$ ist der Flächeninhalt zwischen der Indikatorfunktion und der x -Achse. Weil die Breite des Indikators gegen 0 konvergiert, konvergiert das Integral und damit (in diesem Wahrscheinlichkeitsraum) der Erwartungswert gegen 0.

$\mathbf{X}_n \not\xrightarrow{\text{f.s.}} \mathbf{0}, n \rightarrow \infty$, das ist klar, weil

$$\mathbb{P}(\{\omega : X_n(\omega) \rightarrow 0\}) = 0.$$

Man beachte: In diesem Wahrscheinlichkeitsraum ist die fast sichere Konvergenz gerade die punktweise Konvergenz auf einer Menge mit Maß 1. Weil die Folge (X_n) ausgewertet in beliebigem $\omega \in [0, 1]$ unendlich oft zwischen 0 und 1 wechselt (1 wenn das kleine Intervall ω enthält, 0 sonst), konvergiert sie fast sicher nicht.

Beispiel 4.5.5. ▶ Wie im vorherigen Beispiel sei $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ und $\mathbb{P} = \lambda_{|[0,1]}$, das Lebesguemaß auf $[0, 1]$. Wir schauen uns die konkrete Folge

$$X_n = n \cdot \mathbf{1}_{[0, \frac{1}{n}]}$$

an, und schauen, in welchem Sinne sie gegen die Grenzzufallsvariable $X = 0$ (Nullfunktion) konvergiert.

$\mathbf{X}_n \xrightarrow{\text{f.s.}} \mathbf{0}, n \rightarrow \infty$: Das ist klar, weil in diesem Beispiel die fast sichere Konvergenz einfach nur die übliche punktweise Konvergenz von Funktionen auf einer Menge von Maß 1 bedeutet und unsere Folge auf $(0, 1]$ punktweise gegen 0 konvergiert. Weil ein einzelner Punkt im Lebesguemaß eine Nullmenge ist, konvergiert die Folge fast sicher:

$$\mathbb{P}(X_n \rightarrow 0) = \mathbb{P}((0, 1]) = 1.$$

$\mathbf{X}_n \xrightarrow{\mathbf{P}} \mathbf{0}, n \rightarrow \infty$: Die stochastische Konvergenz gilt, weil für $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n| > \varepsilon) = \mathbb{P}(n\mathbf{1}_{[0, \frac{1}{n}]} > \varepsilon) = \frac{1}{n} \rightarrow 0, \quad n \rightarrow \infty.$$

$\mathbf{X}_n \not\xrightarrow{\mathcal{L}^p} \mathbf{0}, n \rightarrow \infty$: Weil wir nur Erwartungswerte von Indikatoren berechnen müssen, folgt alles direkt aus den Rechenregeln für Erwartungswerte:

$$\mathbb{E}[X_n - X]^p = \mathbb{E}[n^p \cdot \mathbf{1}_{[0, \frac{1}{n}]}^p] = n^p \mathbb{E}[\mathbf{1}_{[0, \frac{1}{n}]}] = n^p \frac{1}{n} = n^{p-1} \not\rightarrow 0, \quad n \rightarrow \infty,$$

weil wir bei \mathcal{L}^p -Konvergenz immer $p \geq 1$ annehmen.

Beispiel 4.5.6. ▶ Sei $X \sim \mathcal{N}(0, 1)$ und $X_n = (-1)^n X$ für $n \in \mathbb{N}$. Es gilt aufgrund der Symmetrie der Normalverteilung $X_n \sim \mathcal{N}(0, 1)$ für alle $n \in \mathbb{N}$. Weil Erwartungswerte nur von der Verteilung abhängen, gilt also $\mathbb{E}[f(X)] = \mathbb{E}[f(X_n)]$ für alle $n \in \mathbb{N}$, die Folge der Erwartungswerte ist also konstant und konvergiert daher gegen $\mathbb{E}[f(X)]$. Also gilt $X_n \xrightarrow{(d)} X, n \rightarrow \infty$. Andere Konvergenzarten gelten für diese Folge nicht.

Beispiel 4.5.7. ▶ Seien X_1, X_2, \dots unabhängige Zufallsvariablen mit

$$\mathbb{P}(X_n = 1) = \frac{1}{n}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n},$$

d. h. $X_n \sim \text{Ber}(\frac{1}{n}), n \in \mathbb{N}$. Man stelle sich z. B. unabhängige Münzwürfe vor (1 bedeutet „Zahl“, 0 bedeutet „Kopf“), bei denen die Wahrscheinlichkeit für „Zahl“ immer kleiner wird. Alternativ

kann man sich irgendwelche unabhängigen Versuche vorstellen, bei denen 1 „Erfolg“ und 0 „Misserfolg“ bedeutet. Mit dieser Folge gelten:

$\mathbf{X}_n \xrightarrow{\mathbf{P}} \mathbf{0}, \mathbf{n} \rightarrow \infty$: Für $\varepsilon > 0$ gilt

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(X_n > \varepsilon) = \begin{cases} \mathbb{P}(X_n = 1) & : \varepsilon \leq 1 \\ 0 & : \varepsilon > 1 \end{cases} = \begin{cases} \frac{1}{n} & : \varepsilon \leq 1 \\ 0 & : \varepsilon > 1 \end{cases} \rightarrow 0, \quad n \rightarrow \infty.$$

$\mathbf{X}_n \xrightarrow{\text{f.s.}} \mathbf{0}, \mathbf{n} \rightarrow \infty$: Das Argument ist nicht einfach, taucht aber im nächsten Kapitel in verschiedenen Formen auf. Wir schreiben die Konvergenz um in Schnitte und Vereinigungen um. Unter der Beachtung, dass eine Folge mit den Werten 0 und 1 nur gegen 0 konvergiert, wenn sie irgendwann nur noch den Wert 0 annimmt, ist das

$$\begin{aligned} \{\omega \in \Omega \mid X_n(\omega) \rightarrow 0\} &= \{\omega \in \Omega \mid \exists n_0 \in \mathbb{N}: X_n(\omega) = 0 \forall n \geq n_0\} \\ &= \bigcup_{n_0=1}^{\infty} \bigcap_{n \geq n_0} \{\omega \in \Omega \mid X_n(\omega) = 0\}. \end{aligned} \quad (4.2)$$

Zur Erinnerung, wie hatten wir Vereinigungen und Schnitte in Analysis 1 definiert?

$$\begin{aligned} \bigcup_{i \in I} A_i &:= \{\omega \in \Omega \mid \exists i \in I : \omega \in A_i\}, \\ \bigcap_{i \in I} A_i &:= \{\omega \in \Omega \mid \forall i \in I : \omega \in A_i\}. \end{aligned}$$

Diese kleine Überlegung ist unglaublich wichtig. Sie wird später der Weg sein, das starke Gesetz der großen Zahlen zu beweisen. Damit kann fast sichere Konvergenz immer in Vereinigungen über Schnitte umformuliert werden und diese können wir immer mit Subadditivität und Stetigkeit von Maßen attackieren. Mit (4.2) bekommen wir also

$$\begin{aligned} \mathbb{P}(X_n \rightarrow 0) &= \mathbb{P}\left(\bigcup_{n_0=1}^{\infty} \bigcap_{n \geq n_0} \{\omega \in \Omega \mid X_n(\omega) = 0\}\right) \\ &\stackrel{\text{Subadd.}}{\leq} \sum_{n_0=1}^{\infty} \mathbb{P}\left(\bigcap_{n \geq n_0} \{\omega \in \Omega \mid X_n(\omega) = 0\}\right) \\ &\stackrel{\text{Stet. Maße}}{=} \sum_{n_0=1}^{\infty} \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcap_{n=n_0}^m \{\omega \in \Omega \mid X_n(\omega) = 0\}\right) \\ &\stackrel{\text{unab.}}{=} \sum_{n_0=1}^{\infty} \lim_{m \rightarrow \infty} \prod_{n=n_0}^m \mathbb{P}(X_n = 0) \\ &= \sum_{n_0=1}^{\infty} \lim_{m \rightarrow \infty} \left(1 - \frac{1}{n_0}\right) \cdots \left(1 - \frac{1}{m}\right) \\ &= \sum_{n_0=1}^{\infty} \lim_{m \rightarrow \infty} \left(\frac{n_0 - 1}{n_0}\right) \cdots \left(\frac{m - 1}{m}\right) \\ &\stackrel{\text{kürzen}}{=} \sum_{n_0=1}^{\infty} \lim_{m \rightarrow \infty} \frac{n_0 - 1}{m} = 0. \end{aligned}$$

Folglich ist die Wahrscheinlichkeit der Konvergenz sogar 0 und daher gilt fast sichere Konvergenz nicht.

Mit den Beispielen haben wir schon ein paar Beispiele gesammelt, die zeigen, dass die Konvergenzarten nicht äquivalent sein können. Den Zusammenhang der Konvergenzarten wollen wir nun genauer beleuchten.

Zunächst schauen wir die Konvergenz in Verteilung etwas genauer an. Dazu zunächst ein Hilfslemma:

Lemma 4.5.8. ▶ Seien F, F_1, F_2, \dots Verteilungsfunktionen. Wenn $\lim_{n \rightarrow \infty} F_n(t) = F(t)$ für alle Stetigkeitsstellen t von F gilt, so gilt $\lim_{n \rightarrow \infty} F_n^{-1}(y) = F^{-1}(y)$ für alle Stetigkeitsstellen von F^{-1} .

Beweis. Der Beweis basiert hauptsächlich auf einer Eigenschaft der Pseudoinversen, die wir aus Lemma 4.3.2 (iii) schon kennen. Wir nutzen die Eigenschaft aber auch als strikte Abschätzung in die andere Richtung:

$$a < F^{-1}(u) \leq b \Leftrightarrow F(a) < u \leq F(b) \quad \text{für alle } a, b \in \mathbb{R} \text{ und } u \in (0, 1). \quad (4.3)$$

Die Aussage folgt genau wie in Lemma 4.3.2 direkt aus der Definition von F^{-1} . Wir wählen jetzt Stetigkeitsstellen a und b von F mit $a < F^{-1}(y) < b$ und ein $v \in (y, 1)$ mit $F^{-1}(v) < b$. Damit es so ein v gibt, nutzen wir die Stetigkeit von F^{-1} an der Stelle y (das ist die $\epsilon - \delta$ Definition für $\epsilon = b - F^{-1}(y)$). Jetzt formen wir mit (4.3) um und nutzen die Definition der Folgenkonvergenz:

$$\begin{aligned} a < F^{-1}(y) &\leq F^{-1}(v) < b \\ \stackrel{(4.3)}{\implies} & F(a) < y < v \leq F(b) \\ \implies & \exists N \in \mathbb{N} : F_n(a) < y < F_n(b) \text{ für alle } n \geq N \\ \stackrel{(4.3)}{\implies} & \exists N \in \mathbb{N} : a < F_n^{-1}(y) \leq b \text{ für alle } n \geq N \end{aligned}$$

Wir wissen noch nicht, dass die Folge $(F_n^{-1}(y))_{n \in \mathbb{N}}$ konvergiert, nutzen also \liminf und \limsup , um sauber zu formulieren: Da steht also

$$a \leq \liminf_{n \rightarrow \infty} F_n^{-1}(y) \leq \limsup_{n \rightarrow \infty} F_n^{-1}(y) \leq b.$$

Weil a und b beliebig nah an $F^{-1}(y)$ gewählt werden können (es gibt nur abzählbar viele Unstetigkeitsstellen von F^{-1}), gilt also $\liminf_{n \rightarrow \infty} F_n^{-1}(y) = \limsup_{n \rightarrow \infty} F_n^{-1}(y) = F^{-1}(y)$. Damit existiert der Grenzwert und es gilt $\lim_{n \rightarrow \infty} F_n^{-1}(y) = F^{-1}(y)$. \square

Das Lemma ist etwas technisch, daraus folgt aber folgende äußerst wichtige Aussage. Wir hätten die Konvergenz in Verteilung auch durch Konvergenz der Verteilungsfunktionen definieren können:

Satz 4.5.9. ▶ Seien X, X_1, X_2, \dots Zufallsvariablen mit $X \sim F$ und $X_n \sim F_n$, $n \in \mathbb{N}$, so gilt:

$$X_n \xrightarrow{(d)} X, \quad n \rightarrow \infty \quad \iff \quad \lim_{n \rightarrow \infty} F_n(t) = F(t) \quad \text{für alle Stetigkeitsstellen von } F.$$

Beweis.

„ \Rightarrow “: Weil die Konvergenz in Verteilung gerade $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ bedeutet, sind wir schon fast fertig. Wir brauchen nur $f = \mathbf{1}_{(-\infty, t]}$ einsetzen und dann haben wir die Konvergenz der Verteilungsfunktionen an der Stelle t . Leider geht das nicht, f ist zwar beschränkt, jedoch nicht stetig. Um das zu umschiffen, wählen wir für beliebiges $\delta > 0$ folgende Modifikation: Wir nehmen die Funktionen f_+ und f_- , die wie folgt definiert sind: f_+ ist der Indikator auf $(-\infty, t]$ und verbindet dann linear 1 und 0 auf $[t, t + \delta]$, rechts von δ ist f_+ null. f_- ist ähnlich, verbindet aber auf $[t - \delta, t]$ linear die 1 und die 0. Es gilt also

$$\mathbf{1}_{(-\infty, t - \delta]} \leq f_- \leq f \leq f_+ \leq \mathbf{1}_{(-\infty, t + \delta]}$$

und f_-, f_+ sind beschränkt und stetig weil es einfach Indikatorfunktionen sind, die die Unstetigkeitsstelle linear stetig machen. Wir können also die angenommene Konvergenz für f_+ und f_- anwenden.

Mit Monotonie des Erwartungswertes gilt

$$F_n(t) = \mathbb{E}[\mathbf{1}_{(-\infty, t]}(X_n)] \leq \mathbb{E}[f_+(X_n)] \xrightarrow{f_+ \text{ stet.}} \mathbb{E}[f_+(X)] \leq \mathbb{E}[\mathbf{1}_{(-\infty, t + \delta]}(X)] = F(t + \delta).$$

Analog mit f_- schätzen wir nach unten ab:

$$F_n(t) = \mathbb{E}[\mathbf{1}_{(-\infty, t]}(X_n)] \geq \mathbb{E}[f_-(X_n)] \xrightarrow{f_- \text{ stet.}} \mathbb{E}[f_-(X)] \geq \mathbb{E}[\mathbf{1}_{(-\infty, t-\delta]}(X)] = F(t - \delta).$$

Weil $\delta > 0$ beliebig war, gilt also $F_n(t) = F(t)$. An dieser letzten Stelle haben wir die Annahme benutzt, dass t eine Stetigkeitsstelle von F ist!

„ \Leftarrow “: Sei $U \sim \mathcal{U}((0, 1))$ durch die kanonische Konstruktion konstruiert. Der Wahrscheinlichkeitsraum ist also $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda_{[0,1]})$ und die Zufallsvariable die Identitätsabbildung. Definiere

$$Y = F^{-1}(U) \quad \text{und} \quad Y_n = F_n^{-1}(U), \quad n \in \mathbb{N}.$$

Wegen der inversen Transformations Methode gilt

$$Y \sim F \quad \text{und} \quad Y_n \sim F_n, \quad n \in \mathbb{N}.$$

Aufgrund der Voraussetzung und Lemma 4.5.8 konvergiert $F_n^{-1}(t)$ für alle Stetigkeitsstellen von F^{-1} gegen $F^{-1}(t)$. Die Unstetigkeitsstellen der nicht-fallenden Funktion F^{-1} sind abzählbar, also eine Nullmenge in dem gewählten Wahrscheinlichkeitsraum (abzählbare Mengen sind Nullmengen für das Lebesguemaß). Weil U die Identitätsabbildung ist, gilt also

$$Y_n = F_n^{-1}(U) \xrightarrow{\text{f.s.}} F^{-1}(U) = Y, \quad n \rightarrow \infty.$$

Wir können in diesem Fall sogar das Ereignis von Maß 1 benennen, auf dem die Konvergenz gilt: Das sind gerade die Stetigkeitsstellen von F^{-1} in $(0, 1)$. Weil Erwartungswerte von identisch verteilten Zufallsvariablen gleich sind (Lemma 4.1.10 und die Bemerkung danach), gilt also für f stetig und beschränkt

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] &\stackrel{X_n \sim Y_n}{=} \lim_{n \rightarrow \infty} \mathbb{E}[f(Y_n)] \stackrel{3.2.5}{=} \mathbb{E}[\lim_{n \rightarrow \infty} f(Y_n)] \\ &\stackrel{Y_n \rightarrow Y \text{ f.s.}}{=} \mathbb{E}[f(Y)] \stackrel{X \sim Y}{=} \mathbb{E}[f(X)]. \end{aligned}$$

Damit ist die Konvergenz in Verteilung gewiesen. □

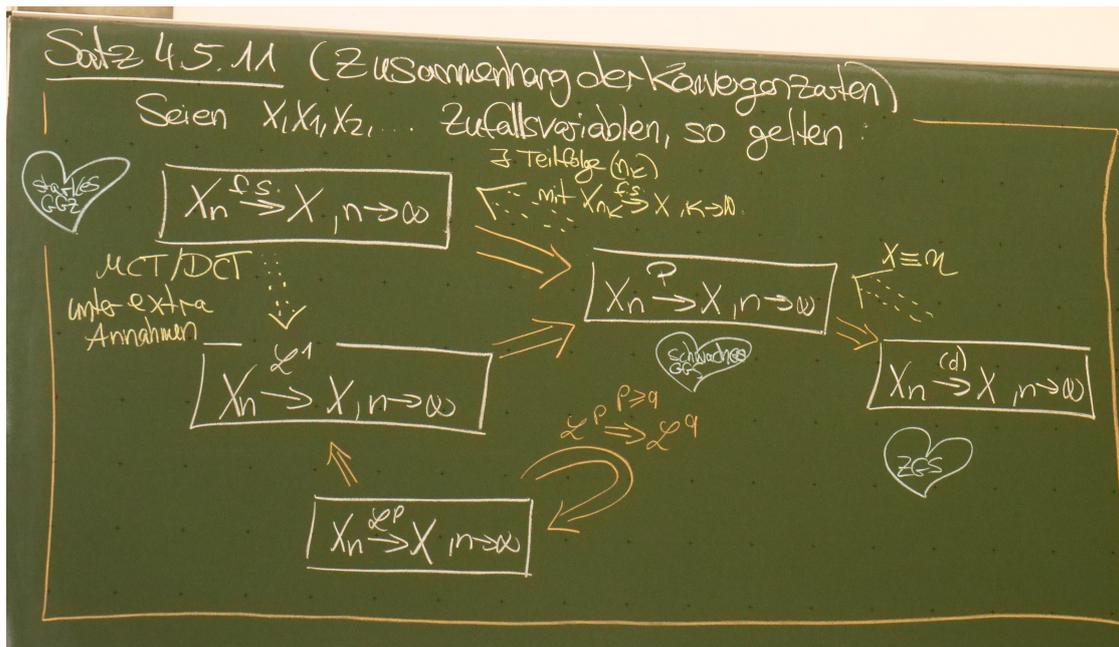
Im Prinzip sagt die Aussage, dass schwache Konvergenz (fast) das gleiche ist, wie punktweise Konvergenz der Verteilungsfunktionen. Die Stetigkeitsstellen sind nur für die Grenzverteilungsfunktion gefordert. In vielen Fällen, z. B. wenn die Grenzverteilung eine Normalverteilung ist, ist F stetig. In dem Fall ist schwache Konvergenz nichts anderes als die punktweise Konvergenz der Verteilungsfunktionen. Das taucht gleich zum Beispiel beim zentralen Grenzwertsatz noch auf.

Bemerkung 4.5.10. ▶ Die Konstruktion im Beweis heißt Skorokhod-Kopplung: Für jede Folge X_1, \dots von Zufallsvariablen (auf beliebigen Wahrscheinlichkeitsräumen), die schwach gegen eine Zufallsvariable X konvergiert, gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit Zufallsvariablen Y, Y_1, \dots , so dass

- $X \sim Y, X_n \sim Y_n$ für alle $n \in \mathbb{N}$,
- $Y_n \xrightarrow{\text{f.s.}} Y, n \rightarrow \infty$.

Die Skorokhod-Kopplung ist besonders spektakulär, wenn man mit dem nächsten Satz vergleicht, in dem der Zusammenhang der Konvergenzarten beleuchtet wird:

Satz 4.5.11. ▶ [Zusammenhang Konvergenzarten]



Satz 4.5.11 (Zusammenhang Konvergenzarten)

Beweis. Konvergenz in $\mathcal{L}^p \Rightarrow$ Konvergenz in \mathcal{L}^q für $q < p$:

Wir wissen schon, wie aus Hölder $\mathbb{E}[|X|^q] \leq \mathbb{E}[|X|^p]$ folgt, man fügt einfach eine 1 hinzu. Als Wiederholung machen wir das nochmal: Definiere dazu $r = \frac{p}{q}$ und $r' = \frac{r}{r-1}$, also gilt

$$\mathbb{E}[|X_n - X|^q] = \mathbb{E}[|X_n - X|^q \cdot 1] \leq (\mathbb{E}[|X_n - X|^{qr}]^{1/r} (\mathbb{E}[1^{r'}])^{1/r'}) = \mathbb{E}[|X_n - X|^p]^{q/p} \cdot 1.$$

Wenn die rechte Seite gegen 0 konvergiert, konvergiert also auch die linke Seite gegen 0. Das ist genau das, was wir zeigen wollten.

Konvergenz in $\mathcal{L}^1 \Rightarrow$ Stochastische Konvergenz:

Mit der Markov Ungleichung gilt:

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|]}{\varepsilon} \xrightarrow{\text{Vor.}} 0, \quad n \rightarrow \infty.$$

Das war es schon.

Fast sichere Konvergenz \Rightarrow Stochastische Konvergenz:

Der Trick ist es, die Definition der Konvergenz aus Analysis 1 als Schnitte und Vereinigungen von Mengen zu schreiben:

$$\begin{aligned} 1 &= \mathbb{P}(X_n \rightarrow X) \\ &\stackrel{\text{Notation}}{=} \mathbb{P}(\{\omega: X_n(\omega) \rightarrow X(\omega), n \rightarrow \infty\}) \\ &\stackrel{\text{Def.}}{=} \mathbb{P}(\{\omega: \forall \varepsilon > 0 \exists N \in \mathbb{N}: |X_n(\omega) - X(\omega)| < \varepsilon \forall n \geq N\}) \\ &= \mathbb{P}\left(\bigcap_{\varepsilon > 0} \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} \{\omega: |X_n(\omega) - X(\omega)| < \varepsilon\}\right). \end{aligned}$$

Mit Komplementbildung und de Morgan'schen Regeln folgt daraus

$$\begin{aligned}
0 &= \mathbb{P}\left(\left(\bigcap_{\varepsilon>0} \bigcup_{N \in \mathbb{N}} \bigcap_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| < \varepsilon\}\right)^C\right) \\
&= \mathbb{P}\left(\bigcup_{\varepsilon>0} \bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}\right) \\
&\stackrel{\text{Mon. } \varepsilon \text{ fest}}{\geq} \mathbb{P}\left(\bigcap_{N \in \mathbb{N}} \bigcup_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}\right) \\
&\stackrel{\text{Mon. Ma\ss e}}{=} \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcup_{n \geq N} \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}\right) \\
&\stackrel{\text{Mon.}}{\geq} \lim_{N \rightarrow \infty} \mathbb{P}(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \geq 0.
\end{aligned}$$

Also gilt $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$ für alle $\varepsilon > 0$, das ist die stochastische Konvergenz.

Stochastische Konvergenz \Rightarrow Konvergenz in Verteilung:

Wir zeigen die Aussage in zwei Schritten, erst für gleichmäßig stetiges f , dann für stetiges f .

(i) Sei f gleichmäßig stetig, d. h.

$$\forall \varepsilon > 0 \exists \delta > 0: |x - x'| < \delta \Rightarrow |f(x) - f(x')| < \varepsilon. \quad (4.4)$$

Sei nun $\varepsilon > 0$ beliebig und δ dazugehörig aus (4.4). Definiere $A_n = \{\omega : |X_n(\omega) - X(\omega)| \geq \delta\}$. Damit gilt, mit $\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$,

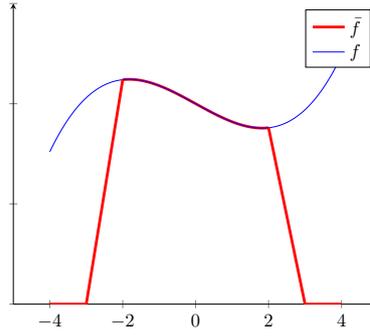
$$\begin{aligned}
0 &\leq |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \\
&\stackrel{\Delta}{\leq} \mathbb{E}[|f(X_n) - f(X)|] \\
&= \mathbb{E}[|f(X_n) - f(X)| \cdot \underbrace{1}_{\mathbf{1}_{A_n} + \mathbf{1}_{A_n^C}}] \\
&= \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{A_n}] + \mathbb{E}[|f(X_n) - f(X)| \cdot \mathbf{1}_{A_n^C}] \\
&\stackrel{(4.4), \text{Def. } A_n}{\leq} \mathbb{E}[2\|f\|_\infty \mathbf{1}_{A_n}] + \mathbb{E}[\varepsilon \mathbf{1}_{A_n^C}] \\
&= 2\|f\|_\infty \mathbb{P}(A_n) + \varepsilon \mathbb{P}(A_n^C) \\
&\leq 2\|f\|_\infty \mathbb{P}(A_n) + \varepsilon \\
&= 2\|f\|_\infty \mathbb{P}(|X_n - X| > \delta) + \varepsilon \\
&\stackrel{\text{Vor.}}{\xrightarrow{}} \varepsilon, \quad n \rightarrow \infty.
\end{aligned}$$

Weil $\varepsilon > 0$ beliebig, gilt damit $\lim_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| = 0$ und somit

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

Damit ist die Definition der Konvergenz in Verteilung für gleichmäßig stetige beschränkte Funktionen gezeigt.

(ii) Sei jetzt f eine beliebige stetige beschränkte Funktion und $\varepsilon > 0$ fest. Für Intervalle $[-k, k]$ gilt wegen der Stetigkeit von Maßen $\mathbb{P}(X \notin [-k, k]) \rightarrow 0$, $k \rightarrow \infty$. Daher können wir ein $k \in \mathbb{N}$ mit $\mathbb{P}(X \notin [-k+1, k-1]) < \varepsilon$ wählen. Wir definieren jetzt die stetige Funktion \bar{f} wie in dem Bildchen, in dem $k = 2$ gewählt ist.



In Worten: \bar{f} ist gleich f in $[-k, k]$, null außerhalb von $[-k - 1, k + 1]$ und verbindet dazwischen linear 0 und $f(k)$ bzw. $f(-k)$. Die wichtige dazu gewonnene Information ist, dass \bar{f} gleichmäßig stetig ist. Das gilt, weil stetige Funktionen auf kompakten Mengen gleichmäßig stetig sind (siehe Analysis 1). Jetzt kommt ein mehrfach genutzter Trick aus der Analysis. Wir addieren zwei Mal 0 und nutzen die Dreiecksungleichung

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \stackrel{\Delta}{\leq} \underbrace{\mathbb{E}[|f(X_n) - \bar{f}(X_n)|]}_{:= I_n} + \underbrace{\mathbb{E}[|\bar{f}(X_n) - \bar{f}(X)|]}_{:= II_n} + \underbrace{\mathbb{E}[|\bar{f}(X) - f(X)|]}_{:= III_n} \quad (4.5)$$

und betrachten einzeln die Grenzwerte der drei Summanden. Aus (i), und weil \bar{f} gleichmäßig stetig, wissen wir, dass $II_n \rightarrow 0$ für $n \rightarrow \infty$. Für den dritten Summanden gilt

$$\begin{aligned} III_n &= \mathbb{E}[|f(X) - \bar{f}(X)| \cdot (\mathbf{1}_{[-k, k]}(X) + \mathbf{1}_{[-k, k]^c}(X))] \\ &\leq 0 + \mathbb{E}[2\|f\|_{\infty} \mathbf{1}_{[-k, k]^c}(X)] \\ &= 2\|f\|_{\infty} \mathbb{P}(X \in [-k, k]^c) \\ &= 2\|f\|_{\infty} \mathbb{P}(X \notin [-k, k]) \\ &\stackrel{\text{Mon.}}{\leq} 2\|f\|_{\infty} \mathbb{P}(X \notin [-k + 1, k - 1]) < 2\|f\|_{\infty} \epsilon, \end{aligned}$$

wegen der Wahl von k und weil $f(x) - \bar{f}(x) = 0$ für $x \in [-k, k]$. Schließlich noch der erste Summand. Wegen der angenommenen stochastischen Konvergenz gilt

$$0 \leq \mathbb{P}(X_n \notin [-k, k], X \in [-k + 1, k - 1]) \leq \mathbb{P}(|X_n - X| > 1) \rightarrow 0, \quad n \rightarrow \infty.$$

Damit zerlegen wir wie folgt:

$$\begin{aligned} I_n &= \mathbb{E}[|f(X_n) - \bar{f}(X_n)| \cdot (\mathbf{1}_{[-k, k]}(X_n) + \mathbf{1}_{[-k, k]^c}(X_n))] \\ &\leq 0 + \mathbb{E}[2\|f\|_{\infty} \mathbf{1}_{[-k, k]^c}(X_n)] \\ &= 2\|f\|_{\infty} \mathbb{P}(X_n \notin [-k, k]) \\ &= 2\|f\|_{\infty} (\mathbb{P}(X_n \notin [-k, k], X \in [-k + 1, k - 1]) + \mathbb{P}(X_n \notin [-k, k], X \notin [-k + 1, k - 1])) \\ &\leq 2\|f\|_{\infty} (\mathbb{P}(|X_n - X| \geq 1) + \mathbb{P}(X \notin [-k + 1, k - 1])) \\ &\leq 2\|f\|_{\infty} (\mathbb{P}(|X_n - X| \geq 1) + \epsilon). \end{aligned}$$

Die rechte Seite konvergiert dann gegen $2\|f\|_{\infty} \epsilon$. In der Rechnung haben wir viele kleine Eigenschaften benutzt, checkt es mal selber Zeile für Zeile: Monotonie und Linearität von Erwartungswerten, dass Erwartungswerte von Indikatoren Wahrscheinlichkeiten sind (siehe Proposition 4.1.14), sowie die σ -Additivität und Monotonie von Maßen.

Die drei einzelnen Betrachtungen zusammen ergeben wegen (4.5)

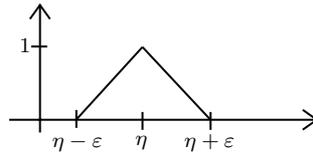
$$0 \leq \limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leq 2\|f\|_{\infty} \epsilon + 0 + 2\|f\|_{\infty} \epsilon.$$

Weil ε beliebig war, folgt daraus $\limsup_{n \rightarrow \infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| = 0$ und damit $\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$. Das ist die Konvergenz in Verteilung.

Die Hauptrichtungen sind nun vollständig bewiesen. Wir zeigen jetzt noch, dass eine Umkehrung gilt, wenn der Grenzwert eine konstante Zufallsvariable ist:

Konvergenz in Verteilung gegen eine konstante Zufallsvariable \Rightarrow Stochastische Konvergenz:

Sei nun (X_n) eine Folge von Zufallsvariablen, die gegen eine fast sicher konstante Zufallsvariable X in Verteilung konvergiert. Sei $\eta \in \mathbb{R}$ der Wert, den X \mathbb{P} -fast sicher annimmt und sei $\varepsilon > 0$ beliebig. Sei nun $I = [\eta - \varepsilon, \eta + \varepsilon]$ und f eine stetige Funktion mit $f \leq \mathbf{1}_I$ sowie $f(\eta) = 1$. Natürlich gibt es so ein f , zum Beispiel das aus folgendem Bildchen:



Wegen der angenommenen Konvergenz in Verteilung gilt damit

$$\begin{aligned} 1 &= f(\eta) \\ &\stackrel{4.1.14 \text{ (iii)}}{=} \mathbb{E}[f(X)] \\ &= \lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] \\ &\stackrel{\text{Monotonie}}{\leq} \lim_{n \rightarrow \infty} \mathbb{E}[\mathbf{1}_I(X_n)] \\ &\stackrel{4.1.14 \text{ (iv)}}{=} \lim_{n \rightarrow \infty} \mathbb{P}(X_n \in I) \leq 1, \end{aligned}$$

weil Wahrscheinlichkeiten immer durch 1 dominiert sind. Weil obere und untere Schranke gleich sind, bekommen wir $\lim_{n \rightarrow \infty} \mathbb{P}(X_n \in I) = 1$ und damit

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - \eta| > \varepsilon) = \mathbb{P}(X_n \notin I) = 1 - \mathbb{P}(X_n \in I) \rightarrow 0, \quad n \rightarrow \infty.$$

Stochastische Konvergenz \Rightarrow Fast sichere Konvergenz einer Teilfolge:

Den Teil liefern wir nach dem Borel-Cantelli Lemma nach, siehe die Bemerkung 4.6.6. \square

Aus dem Konvergenzdiagramm folgt natürlich, dass fast sichere Konvergenz auch Konvergenz in Verteilung impliziert. Überlegt doch mal, warum diese Aussage sehr viel schneller auch ohne den Umweg über die stochastische Konvergenz folgt (Stichwort dominierte Konvergenz).

Das Ziel ist immer, die „starken“ Konvergenzen (links im Bildchen) zu zeigen. Das geht leider nicht immer! Um die Begriffe mit Leben zu füllen, zeigen wir drei berühmte Sätze, je einen für stochastische Konvergenz, fast sichere Konvergenz und Konvergenz in Verteilung:

- schwaches Gesetz der großen Zahlen (stochastische Konvergenz)
- starkes Gesetz der großen Zahlen (fast sichere Konvergenz)
- Zentraler Grenzwertsatz (Konvergenz in Verteilung)

Beim schwachen Gesetz der großen Zahlen werden wir merken, dass Konvergenz in \mathcal{L}^p gerade für $p = 1$ oder $p = 2$ ein extrem nützliches Werkzeug ist. Der Grund ist, dass man mit Momenten gut rumrechnen kann (Ausmultiplizieren, Linearität, ...).

Satz 4.5.12. ▶ [Schwaches Gesetz der großen Zahlen]

(i) Klassische Variante: Sind X_1, X_2, \dots u.i.v. mit $\mathbb{E}[X_1^2] < \infty$, so gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mathbb{E}[X_1], \quad n \rightarrow \infty.$$

- (ii) Variante mit schwächeren Annahmen: Sind X_1, X_2, \dots quadratintegrierbar, paarweise unkorreliert (z. B. paarweise unabhängig) mit identischen Erwartungswert und

$$\frac{1}{n^2} \sum_{k=1}^n \mathbb{V}(X_k) \rightarrow 0, \quad n \rightarrow \infty, \quad (4.6)$$

so gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mathbb{E}[X_1], \quad n \rightarrow \infty.$$

Fürs bessere Verständniss kann man überlegen, wie die Annahme „identischer Erwartungswert“ in (ii) ersetzt werden kann, da gibt es verschiedene Möglichkeiten.

Beweis. (i) Wenn man den Beweis gesehen hat, ist alles ziemlich simple: **Erst Tschebycheff, dann Bienaymé.** Mit Tschebycheff und Bienaymé gilt für beliebiges $\varepsilon > 0$ aufgrund der u.i.v. Annahme

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - \mathbb{E}[X_1]\right| \geq \varepsilon\right) &= \mathbb{P}\left(\left|\sum_{k=1}^n X_k - n\mathbb{E}[X_1]\right| \geq n\varepsilon\right) \\ &\stackrel{4.1.22}{\leq} \frac{\mathbb{V}\left[\sum_{k=1}^n X_k\right]}{n^2\varepsilon^2} \\ &\stackrel{4.2.31}{=} \frac{\sum_{k=1}^n \mathbb{V}[X_k]}{n^2\varepsilon^2} \\ &= \frac{n \cdot \mathbb{V}[X_1]}{n^2\varepsilon^2} \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Also gilt $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mathbb{E}[X_1]$ für $n \rightarrow \infty$. Beachtet dabei, dass wir für Tschebycheff $\mathbb{E}[\sum_{k=1}^n X_k] = n\mathbb{E}[X_1]$ genutzt haben. Um ganz genau zu sein (und wir wollen natürlich genau sein!), merken wir noch an, dass „ $>$ “ oder „ \geq “ in der Definition der schwachen Konvergenz natürlich keine Rolle spielt, ε ist schließlich beliebig.

Man kann das Argument auch anders hinschreiben. Benutzt zum Üben doch mal statt Tschebycheff die Markov Ungleichung mit $h(x) = x^2$ plus die Verschiebungsformel für die Varianz.

(ii) Um die schwächeren Annahmen von (ii) zu verstehen, brauchen wir nur in den vier Zeilen des Arguments zu schauen, wie wir die Unabhängigkeit abschwächen können, so dass die Konvergenz immer noch folgt. Wir sehen dann sofort, dass für Bienaymé nur paarweise unkorreliert gebraucht wird, dann aber für die Konvergenz noch (4.6) gefordert werden muss (es wird nicht mehr identisch verteilt angenommen, es gilt also nicht $\mathbb{V}[X_k] = \mathbb{V}[X_1]$). Schaut euch das in Ruhe an, um die Idee „erst Tschebyscheff, dann Bienaymé“ einzubrennen. \square

Zum besseren Verständnis kann man mal ausprobieren, ob man das schwache Gesetz genauso mit der Annahme $\mathbb{E}[|X_1|] < \infty$ beweisen könnte. Warum funktioniert der Beweis nicht, wenn man die Markovungleichung für das erste Moment statt für das zweite Moment ausprobiert? Der Trick an dem Beweis mit zweiten Momenten ist, dass die Varianz der Summe, die eigentlich aus n^2 vielen Summanden besteht, sich aufgrund der Annahme (unabhängig oder unkorreliert) zu einer Summe aus nur n vielen Summanden reduziert (im Beweis von Bienaymé kürzen sich die meisten Terme raus). Damit dominiert der Nenner mit n^2 und die obere Schranke konvergiert gegen 0. Der Effekt passiert beim ersten Moment nicht, weil keine „gemischten Terme“ $\mathbb{E}[X_i X_j] = 0$ auftauchen. Deshalb stünde sowohl in Zähler als auch in Nenner etwas mit n , die obere Schranke würde also nicht gegen 0 konvergieren, in Formeln (für $\mathbb{E}[X_1] = 0$):

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{k=1}^n X_k\right| \geq \varepsilon\right) \stackrel{4.1.22, h(x)=|x|}{\leq} \frac{\mathbb{E}\left[\left|\sum_{k=1}^n X_k\right|\right]}{n\varepsilon} \leq \frac{\sum_{k=1}^n \mathbb{E}[|X_k|]}{n\varepsilon} = \frac{n\mathbb{E}[|X_1|]}{n\varepsilon} \not\rightarrow 0,$$

für $n \rightarrow \infty$. Genau den selben Effekt werden wir beim Beweis des starken Gesetzes der großen Zahlen sehen, bei dem wir endliche 4.te Momente annehmen und bei der Markovungleichung mit $h(x) = x^4$ genug Summanden verschwinden, dass der Nenner dominiert.

4.6 Starkes Gesetz der großen Zahlen

Vorlesung 26

Was bedeutet die stochastische Konvergenz, bzw. warum ist sie schwach? Seien als Beispiel X_1, X_2, \dots u.i.v. Würfel, also diskret gleichverteilt auf $\{1, \dots, 6\}$. Wegen $\mathbb{E}[X_1] = 3,5$ bedeutet das schwache Gesetz der großen Zahlen (wähle zum Beispiel $\epsilon = 0,01$), dass

$$\mathbb{P}\left(3,49 < \frac{1}{n} \sum_{i=1}^n X_i < 3,51\right) = \mathbb{P}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X_1]\right| < 0,01\right) \rightarrow 1, \quad n \rightarrow \infty.$$

In Worten steht hier: „Der Mittelwert wird mit hoher Wahrscheinlichkeit nah bei 3,5 liegen, wenn n groß ist.“ Es wird damit nicht ausgeschlossen, dass der Mittelwert mit kleiner Wahrscheinlichkeit weit vom Erwartungswert entfernt liegt. Genau hier liegt der Unterschied zum starken Gesetz der großen Zahlen, das wir als nächstes beweisen wollen. Hier wird die stochastische Konvergenz durch fast sichere Konvergenz ersetzt. Weil in der Definition der fast sicheren Konvergenz alle n gemeinsam *innerhalb* der Wahrscheinlichkeit auftauchen, $\mathbb{P}(X_n \rightarrow X) = 1$, kann der Effekt nicht auftreten.

Als Hilfsmittel für den Beweis diskutieren wir zunächst das Borel-Cantelli Lemma. Dazu zunächst ein paar Definitionen:

Definition 4.6.1. ▶ Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \dots \in \mathcal{A}$ beliebige Ereignisse.

(i)

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &:= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \\ &= \{\omega \in \Omega: \omega \in A_n \text{ für unendlich viele } n\} \\ &\stackrel{\text{Notation}}{=} \{A_n \text{ unendlich oft}\} \end{aligned}$$

heißt **Limes superior** der Folge (A_n) von Ereignissen.

(ii)

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &:= \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \\ &= \{\omega \in \Omega: \omega \in A_n \text{ schließlich immer}\} \\ &\stackrel{\text{Notation}}{=} \{A_n \text{ schließlich immer}\} \end{aligned}$$

Limes inferior der Folge (A_n) von Ereignissen.

Weil aufgrund einer σ -Algebra abzählbare Schnitte und Vereinigungen wieder in \mathcal{A} sind, sind auch $\limsup_{n \rightarrow \infty} A_n$ und $\liminf_{n \rightarrow \infty} A_n$ in \mathcal{A} . Wem der Begriff „schließlich immer“ suspekt ist, der oder die schaue einfach die formelle Definition an. Diese besagt, dass es eine natürliche Zahl n gibt, so dass das Ereignis danach *immer* eintritt (der Durchschnitt von Mengen enthält alle Elemente, die in allen Mengen enthalten sind).

Tatsächlich haben die neuen Begriffe \liminf und \limsup für Mengen auch etwas mit den uns bekannten Begriffen \liminf und \limsup für Folgen zu tun:

Lemma 4.6.2. ▶ Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \dots \in \mathcal{A}$ beliebige Ereignisse. Dann gelten:

- (i) $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n$,
- (ii) $(\liminf_{n \rightarrow \infty} A_n)^C = \limsup_{n \rightarrow \infty} A_n^C$,

$$(iii) \limsup_{n \rightarrow \infty} \mathbf{1}_{A_n}(\omega) = \mathbf{1}_{\limsup_{n \rightarrow \infty} A_n}(\omega), \quad \forall \omega \in \Omega,$$

$$(iv) \liminf_{n \rightarrow \infty} \mathbf{1}_{A_n}(\omega) = \mathbf{1}_{\liminf_{n \rightarrow \infty} A_n}(\omega), \quad \forall \omega \in \Omega.$$

Beweis. Denkt einfach mal kurz darüber nach, was $\liminf_{n \rightarrow \infty} a_n = 1$ oder $\liminf_{n \rightarrow \infty} a_n = 0$ für eine reelle Folge (a_n) bedeutet, wenn diese nur die Werte 0 und 1 annimmt. Übung. \square

Das Borel-Cantelli Lemma gibt uns gleich ein Kriterium, ob die Wahrscheinlichkeit des $\limsup A_n$ null ist oder (mit einer stärkeren Annahme) 1 ist. Dafür basteln wir mit Zufallsvariablen rum, alles basiert auf folgender Bemerkung:

Bemerkung 4.6.3. ▶ Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \dots \in \mathcal{A}$ beliebige Ereignisse. Dann gilt

$$A_1, A_2, \dots \text{ sind unabhängig} \Leftrightarrow \mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots \text{ sind unabhängig.}$$

Beachte: A_1, A_2, \dots ist ein Folge von Ereignissen, wohingegen $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots$ eine Folge von Zufallsvariablen ist. Das ist also ein guter Moment, in Kapitel 4.4 die Definitionen von Unabhängigkeit von Ereignissen und Zufallsvariablen nochmal zu vergleichen! Warum gilt die Äquivalenz? Checken wir die Definitionen:

$$\begin{aligned} \mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots \text{ unabhängig} &\stackrel{\text{Def. 4.4.10}}{\Leftrightarrow} \sigma(\mathbf{1}_{A_1}), \sigma(\mathbf{1}_{A_2}), \dots \text{ unabhängig} \\ &\stackrel{2.1.9}{\Leftrightarrow} \{\emptyset, \Omega, A_1, A_1^C\}, \{\emptyset, \Omega, A_2, A_2^C\}, \dots \text{ unabhängig} \\ &\Leftrightarrow A_1, A_2, \dots \text{ unabhängig.} \end{aligned}$$

Für die dritte Äquivalenz haben wir Definition 4.4.8 genutzt, sowie die Eigenschaft, dass Unabhängigkeit von Ereignissen sich auch auf die Komplemente überträgt (Übung).

Satz 4.6.4. ▶ [Borel-Cantelli-Lemma]

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \dots \in \mathcal{A}$ beliebige Ereignisse, so gelten:

(i)

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty \Rightarrow \mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0.$$

(ii) Sind die A_1, A_2, \dots *zusätzlich* paarweise unabhängig, so gilt

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \Rightarrow \mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1.$$

Damit kennt ihr nun euer erstes „0-1-Gesetz“: Sind die Ereignisse A_1, A_2, \dots paarweise unabhängig, so ist $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) \in \{0, 1\}$ und es gilt $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 1$ genau dann, wenn $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. Das nützliche an 0-1-Gesetzen ist, dass man nur zeigen muss, dass etwas strikt positive Wahrscheinlichkeit hat, um sogar Wahrscheinlichkeit 1 zu schließen.

Beweis.

(i) „triviale Rechnung“: Die einfache Richtung folgt aus einfachen Manipulationen mit Mengen

und Eigenschaften des Maßes:

$$\begin{aligned} \mathbb{P}(\limsup_{n \rightarrow \infty} A_n) &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) \\ &\stackrel{\text{Stet. Maße}}{=} \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcap_{n=1}^N \bigcup_{k=n}^{\infty} A_k\right) \\ &\stackrel{\text{Monotonie}}{\leq} \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcup_{k=N}^{\infty} A_k\right) \\ &\stackrel{\text{subadd.}}{\leq} \lim_{N \rightarrow \infty} \sum_{k=N}^{\infty} \mathbb{P}(A_k) = 0. \end{aligned}$$

Die letzte Gleichheit gilt nach Annahme und Analysis 1 (Eigenschaft konvergenter Reihen).

- (ii) Die Rückrichtung ist deutlich schwieriger, wir nutzen die sogenannte „zweite-Momente-Methode“. Dafür werden erstes und zweites Moment einer geeigneten Zufallsvariable miteinander verglichen. Wir betrachten im Folgenden die Zufallsvariablen $\mathbf{1}_{A_n}$, die aufgrund von Bemerkung 4.6.3 unabhängig sind. Weil die Zufallsvariablen nur die Werte 0 und 1 annehmen, sind sie Bernoulli-verteilt. Genauer, es gilt $\mathbf{1}_{A_n} \sim \text{Ber}(p_n)$, wobei $p_n = \mathbb{P}(A_n)$ die Wahrscheinlichkeit für den Wert 1 ist. Kleine Erinnerung: Für $\text{Ber}(p)$ -verteilte Zufallsvariablen ist der Erwartungswert p und die Varianz $p(1 - p)$. Das werden wir im Folgenden ausnutzen.

Wir betrachten die Folge

$$Z_k = \sum_{n=1}^k \mathbf{1}_{A_n}, \quad k \in \mathbb{N},$$

weil mit dieser Folge $\limsup_{n \rightarrow \infty} A_n$ beschrieben werden kann:

$$\omega \in \limsup_{n \rightarrow \infty} A_n \iff +\infty = \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = \lim_{k \rightarrow \infty} Z_k(\omega). \quad (4.7)$$

Das gilt natürlich weil die Reihe nur aus Summanden 0 oder 1 besteht und daher unendlich ist genau dann, wenn unendlich viele Summanden 1 sind, also wenn ω in unendlich vielen A_n ist. Für Erwartungswert und Varianz gelten

$$\mathbb{E}[Z_k] = \sum_{n=1}^k \mathbb{E}[\mathbf{1}_{A_n}] = \sum_{n=1}^k \mathbb{P}(A_n) \xrightarrow{\text{Ann.}} +\infty, \quad k \rightarrow \infty,$$

sowie

$$\mathbb{V}[Z_k] \stackrel[4.6.3]{4.2.31} \sum_{n=1}^k \mathbb{V}[\mathbf{1}_{A_n}] = \sum_{n=1}^k \mathbb{P}(A_n)(1 - \mathbb{P}(A_n)) \leq \sum_{n=1}^k \mathbb{P}(A_n) = \mathbb{E}[Z_k],$$

weil unabhängige Zufallsvariablen auch unkorreliert sind. Jetzt benutzen wir Tschebyscheff mit den Formeln für Erwartungswert und Varianz:

$$\mathbb{P}\left(|Z_k - \mathbb{E}[Z_k]| \geq \frac{\mathbb{E}[Z_k]}{2}\right) \stackrel{4.1.22}{\leq} \frac{\mathbb{V}[Z_k]}{\frac{1}{4}\mathbb{E}[Z_k]^2} \leq \frac{4\mathbb{E}[Z_k]}{\mathbb{E}[Z_k]^2} = \frac{4}{\mathbb{E}[Z_k]} \rightarrow 0, \quad (4.8)$$

für $k \rightarrow \infty$. Sei nun $\lambda > 0$ beliebig. Dann existiert wegen der Divergenz der Erwartungswerte gegen unendlich ein $k_0 \in \mathbb{N}$ mit $\frac{\lambda}{\mathbb{E}[Z_k]} < \frac{1}{2}$ für alle $k \geq k_0$. Weil aufgrund der Definition von

Z_k auch $Z_k \leq \sum_{n=1}^{\infty} \mathbf{1}_{A_n}$ gilt, bekommen wir für beliebiges $k \geq k_0$

$$\begin{aligned} \mathbb{P}\left(\sum_{n=1}^{\infty} \mathbf{1}_{A_n} < \lambda\right) &\leq \mathbb{P}(Z_k < \lambda) \\ &= \mathbb{P}\left(\frac{Z_k}{\mathbb{E}[Z_k]} < \frac{\lambda}{\mathbb{E}[Z_k]}\right) \\ &\leq \mathbb{P}\left(\frac{Z_k}{\mathbb{E}[Z_k]} < \frac{1}{2}\right) \\ &= \mathbb{P}\left(\frac{Z_k}{\mathbb{E}[Z_k]} - 1 < -\frac{1}{2}\right) \\ &\leq \mathbb{P}\left(\left|\frac{Z_k}{\mathbb{E}[Z_k]} - 1\right| \geq \frac{1}{2}\right) \\ &= \mathbb{P}\left(|Z_k - \mathbb{E}[Z_k]| \geq \frac{\mathbb{E}[Z_k]}{2}\right) \stackrel{(4.8)}{\rightarrow} 0, \quad k \rightarrow \infty. \end{aligned}$$

Also gilt $\mathbb{P}\left(\sum_{n=1}^{\infty} \mathbf{1}_{A_n} < \lambda\right) = 0$. Weil λ beliebig gewählt war, folgt aufgrund der Stetigkeit von Maßen auch $\mathbb{P}\left(\sum_{n=1}^{\infty} \mathbf{1}_{A_n} < \infty\right) = \lim_{N \rightarrow \infty} \mathbb{P}\left(\sum_{n=1}^{\infty} \mathbf{1}_{A_n} < N\right) = 0$. Wegen (4.7) gilt nun

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = \mathbb{P}\left(\sum_{n=1}^{\infty} \mathbf{1}_{A_n} = +\infty\right) = 1.$$

□

Nach diesem Highlight der Maßtheorie, nun zurück zur fast sicheren Konvergenz und dem starken Gesetz der großen Zahlen.

Korollar 4.6.5.  Seien X, X_1, X_2, \dots Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, dann gelten:

(i)

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty \text{ für alle } \epsilon > 0 \implies X_n \xrightarrow{\text{f.s.}} X, \quad n \rightarrow \infty.$$

(ii) Unter der zusätzlich Annahme $X_1 - X, X_2 - X, \dots$ sind paarweise unabhängig, gilt:

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) = +\infty \text{ für ein } \epsilon > 0 \implies X_n \not\xrightarrow{\text{f.s.}} X, \quad n \rightarrow \infty.$$

Beweis. Beide Aussagen folgen sofort aus Borel-Cantelli:

(i) Mit $\epsilon = \frac{1}{k}$ impliziert das Borel-Cantelli-Lemma $\mathbb{P}(|X_n - X| > \frac{1}{k} \text{ unendlich oft}) = 0$ bzw. $\mathbb{P}(|X_n - X| > \frac{1}{k} \text{ nur endlich oft}) = 1$ für alle $k \in \mathbb{N}$. Daraus folgt

$$\begin{aligned} \mathbb{P}(X_n \rightarrow X) &= \mathbb{P}\left(\left\{\omega : \forall k \in \mathbb{N} \exists N \in \mathbb{N} : |X_n(\omega) - X(\omega)| \leq \frac{1}{k} \forall n \geq N\right\}\right) \\ &= \mathbb{P}\left(\bigcap_{k=1}^{\infty} \left\{|X_n - X| > \frac{1}{k} \text{ endlich oft}\right\}\right) = 1, \end{aligned}$$

weil der Durchschnitt abzählbar vieler Mengen von Maß 1 auch Maß 1 hat (wegen Komplementbildung, abzählbare Vereinigungen von Nullmengen sind Nullmengen).

(ii) Borel-Cantelli impliziert $\mathbb{P}(|X_n - X| > \epsilon \text{ unendlich oft}) = 1$, also ist die Wahrscheinlichkeit, dass X_n gegen X konvergiert, sogar 0.

□

Wegen Borel-Cantelli können wir den Unterschied von stochastischer und fast sicherer Konvergenz nun besser verstehen:

Bemerkung 4.6.6.  [Stochastische Konvergenz vs. fast sichere Konvergenz]

- Um stochastische Konvergenz zu zeigen, muss $a_n := \mathbb{P}(|X_n - X| > \epsilon)$ für beliebiges $\epsilon > 0$ eine Nullfolge sein. Konvergiert die Nullfolge so schnell gegen 0, dass auch der Reihengrenzwert $\sum_{n=1}^{\infty} a_n$ endlich ist, so konvergiert nach Korollar 4.6.5 die Folge (X_n) fast sicher gegen X . Wenn wir uns an Analysis 1 erinnern, ist es ein großer Unterschied, ob die Reihe über eine Folge konvergiert oder die Folge eine Nullfolge ist. Beispielsweise reicht für stochastische Konvergenz die Abschätzung $\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{1}{n}$ für fast sichere Konvergenz nicht!
- Wir müssen noch einen Teil des Beweises von Satz 4.5.11 nachliefern. Das folgt aus dem ersten Teil des Korollars mit der Wahl der Teilfolge n_k , die

$$\mathbb{P}\left(|X_{n_k} - X| > \frac{1}{k}\right) < \frac{1}{2^k}$$

erfüllt.

Beispiel 4.6.7.  Schauen wir uns das Beispiel 4.5.7 nochmal etwas allgemeiner an, um ein konkretes Beispiel für Korollar 4.6.5 zu haben: Seien X_1, X_2, \dots unabhängige Zufallsvariablen mit

$$\mathbb{P}(X_n = 1) = \frac{1}{n^p}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n^p},$$

also $X_n \sim \text{Ber}(\frac{1}{n^p})$, $n \in \mathbb{N}$. Man stelle sich wieder unabhängige Versuche vor (1 bedeutet „Erfolg“, 0 bedeutet „Misserfolg“), bei denen die Wahrscheinlichkeit für „Erfolg“ immer kleiner wird. Fragen wir uns wieder: Konvergiert die Folge fast sicher gegen die Zufallsvariable $X = 0$? Wegen der angenommenen Unabhängigkeit gilt mit Korollar 4.6.5

$$X_n \xrightarrow{\text{f.s.}} 0, \quad n \rightarrow \infty \quad \Leftrightarrow \quad \sum_{n=1}^{\infty} \mathbb{P}(X_n = 1) = \sum_{n=1}^{\infty} \frac{1}{n^p} < +\infty \quad \Leftrightarrow \quad p > 1.$$

Ausformuliert ist die Aussage noch spektakulärer weil die Konvergenz gegen 0 einer Folge mit Werten 0 oder 1 bedeutet, dass die Folge irgendwann nur noch den Wert 0 annimmt. Ist $p \leq 1$, so ist der Versuch also immer mal wieder erfolgreich, wohingegen für $p > 1$ der Versuch nur endlich oft erfolgreich ist. Wenn man bedenkt, dass der Versuch unendlich oft ausgeführt wird und jedes Mal positive Wahrscheinlichkeit für Erfolg besteht, ist der Fall $p > 1$ schon überraschend! Hier ist ein ganz konkretes Anwendungsbeispiel. Stellt euch vor, ihr schreibt jedes Jahr die Stochastik 1 Klausur, ohne euch mit dem Inhalt zu beschäftigen. Weil ihr den Inhalt mit der Zeit vergesst, wird die Wahrscheinlichkeit des Bestehens immer kleiner. Die Frage ist nun: Wenn ihr immer wieder probiert, werdet ihr irgendwann bestehen? Das Beispiel zeigt, dass das davon abhängt, wie schnell die Bestehenswahrscheinlichkeit fällt, oder anders formuliert, wie schnell ihr vergesst.

Jetzt zum starken Gesetz der großen Zahlen:

Satz 4.6.8.  [Starkes Gesetz der großen Zahlen]

Sind X_1, X_2, \dots u.i.v. mit $\mathbb{E}[|X_1|] < \infty$, so gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{f.s.}} \mathbb{E}[X_1], \quad n \rightarrow \infty.$$

Beweis. Wir beweisen den Satz nur unter der zusätzlichen Annahme $\mathbb{E}[X_1^4] < \infty$. Die Aussage gilt auch unter der schwachen Annahme $\mathbb{E}[|X_1|] < \infty$, den Beweis werden wir aber erst in einer

weiterführenden Vorlesung diskutieren. Wir nehmen ohne Einschränkung $\mathbb{E}[X_1] = 0$ an (sonst mit $\bar{X}_k := X_k - \mathbb{E}[X_k]$ verschieben, man nennt das Zentrieren). Wir wenden jetzt Korollar 4.6.5 an:

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n X_k - 0\right| \geq \epsilon\right) &\stackrel{4.1.22, h(x)=x^4}{\leq} \frac{\mathbb{E}\left[\left|\frac{1}{n}\sum_{k=1}^n X_k\right|^4\right]}{\epsilon^4} \\ &= \frac{1}{\epsilon^4 n^4} \mathbb{E}\left[\sum_{k_1, k_2, k_3, k_4=1}^n X_{k_1} X_{k_2} X_{k_3} X_{k_4}\right] \\ &\stackrel{\text{unabh., zentr.}}{=} \frac{1}{\epsilon^4 n^4} \left(\sum_{k=1}^n \mathbb{E}[X_k^4] + \sum_{i \neq l=1}^n \mathbb{E}[X_i^2 X_l^2]\right) \\ &\stackrel{\text{ident. vert.}}{=} \frac{1}{\epsilon^4 n^4} \left(n \mathbb{E}[X_1^4] + \frac{n(n-1)}{2} \mathbb{E}[X_1^2] \mathbb{E}[X_1^2]\right) \\ &\leq \frac{C}{n^2}, \end{aligned}$$

mit $C = \frac{\mathbb{E}[X_1^4]}{\epsilon^4} + \frac{\mathbb{E}[X_1^2]^2}{2\epsilon^4}$. Damit haben wir die Voraussetzung von Korollar 4.6.5 (i) gecheckt und die Aussage folgt. Wie beim schwachen Gesetz gilt auch hier wieder: Weil ϵ beliebig ist, spielt „ $>$ “ oder „ \geq “ keine Rolle.

Der Trick ist die dritte Gleichheit. Eigentlich sollten bei Tschebyscheff mit vierter Potenz n^4 Summanden im Zähler auftauchen. Wegen der Unabhängigkeit fallen von den Summanden $\mathbb{E}[X_{i_1} X_{i_2} X_{i_3} X_{i_4}]$ jedoch alle als 0 weg, bei denen eine der Zufallsvariablen nur einmal auftaucht (es gilt wegen der Unabhängigkeit zum Beispiel $\mathbb{E}[X_1 X_1 X_4 X_5] = \mathbb{E}[X_1^2] \mathbb{E}[X_3] \mathbb{E}[X_4] = \mathbb{E}[X_1^2] 0 = 0$). Es bleiben also nicht n^4 viele Summanden stehen, sondern nur die, bei denen entweder immer die gleiche oder nur zwei verschiedene Zufallsvariablen auftauchen. Das sind aber nur etwa n^2 viele und damit bringt der Nenner n^4 die Reihe zum konvergieren. \square

Bemerkung 4.6.6 zeigt uns ganz genau den Unterschied zwischen unseren Beweisen der schwachen und dem starken Gesetze der großen Zahlen: Im Beweis des schwachen Gesetzes haben wir mit zweiten Momenten die obere Schranke $\frac{\mathbb{V}[X_1]}{\epsilon n}$ hergeleitet. Das reichte für stochastische Konvergenz, aber nicht für fast sichere Konvergenz weil die harmonische Reihe divergiert. Das Tschebyscheff Argument mit vierten Momenten hingegen gibt die summierbare obere Schranke $\frac{C}{n^2}$. Dritte Momente funktionieren übrigens auch nicht, da stört der Betrag. Der „richtige“ Beweis (nur unter der Voraussetzung $\mathbb{E}[|X_1|] < \infty$ funktioniert anders, Tschebyscheff ist einfach keine gute Abschätzung).

Bemerkung 4.6.9.  Unabhängig von den Annahmen an die Folge X_1, X_2, \dots spricht man immer von einem „starken Gesetz“, wenn fast sichere Konvergenz vorliegt. Man spricht von einem „schwachen Gesetz“, wenn stochastische Konvergenz vorliegt. Weil fast sichere Konvergenz die stochastische Konvergenz impliziert, impliziert ein starkes Gesetz immer ein schwaches Gesetz. Wir haben also das schwache Gesetz der großen Zahlen für u.i.v. Folgen mit endlichen zweiten Momenten gezeigt, das starke Gesetz der großen Zahlen für u.i.v. Folgen mit endlichen ersten Momenten. Weil $\mathbb{E}[X_1^2] < \infty$ auch $\mathbb{E}[|X_1|] < \infty$ impliziert (Hölder mit 1!), ist die Annahme $\mathbb{E}[X_1^2] < \infty$ im schwachen Gesetz natürlich viel zu stark, es gilt schließlich mit der schwächeren Annahme $\mathbb{E}[|X_1|] < \infty$ die stärkere Aussage der fast sicheren Konvergenz! Teil (i) in Satz 4.5.12 ist also im Prinzip überflüssig und wurde nur aus didaktischen Gründen behandelt. Interessanter ist eigentlich Teil (ii), denn unter diesen schwächeren Annahmen muss das starke Gesetz der großen Zahlen nicht gelten. Ein Beispiel ist folgendes: Ist X_1, X_2, \dots eine Folge von unabhängigen

(nicht identisch verteilten) Zufallsvariablen mit

$$\begin{aligned}\mathbb{P}(X_n = n) &= \frac{1}{2n \log(n+1)}, \\ \mathbb{P}(X_n = -n) &= \frac{1}{2n \log(n+1)}, \\ \mathbb{P}(X_n = 0) &= 1 - \frac{1}{n \log(n+1)},\end{aligned}$$

so gilt das schwache Gesetz, aber nicht das starke Gesetz. Es gibt also durchaus einen Grund für Teil (ii) von Satz 4.5.12.

Am Ende des Kapitels noch eine kleine Bonus-Anwendung von Borel-Cantelli. Der Satz kommt in der Vorlesung nach dem Abspann weil die Aussage zwar nützlich, aber nicht notwendig für euer Verständniss ist (also nicht Klausurrelevant ist). Wir zeigen, dass Mittelwerte $\frac{1}{n} \sum_{k=1}^n X_k$ von u.i.v. Folgen von Zufallsvariablen nur für $\mathbb{E}[|X_1|] < \infty$ fast sicher konvergieren können und (dann gilt das starke Gesetz der großen Zahlen) daher nur gegen den Erwartungswert konvergieren können!

Satz 4.6.10. ▶ Es seien X_1, X_2, \dots unabhängig und identisch verteilt, so dass $\frac{1}{n} \sum_{k=1}^n X_k$ fast sicher für $n \rightarrow \infty$ gegen eine Zufallsvariable X konvergiert. Dann gilt $\mathbb{E}[|X_1|] < \infty$ und

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{f.s.}} \mathbb{E}[X_1], \quad n \rightarrow \infty.$$

Beweis. Elementare Umformungen geben

$$\frac{X_n}{n} = \frac{\sum_{k=1}^n X_k}{n} - \frac{\sum_{k=1}^{n-1} X_k}{n} = \frac{\sum_{k=1}^n X_k}{n} - \frac{n-1}{n} \frac{\sum_{k=1}^{n-1} X_k}{n-1} \xrightarrow{\text{f.s.}} 0, \quad n \rightarrow \infty,$$

weil beide Summanden der rechten Seite nach Annahme gegen X konvergieren. Damit gilt dann

$$\mathbb{P}\left(\left|\frac{X_n}{n}\right| > 1 \text{ unendlich oft}\right) = 0$$

und mit $A_n := \{|X_n| > n\}$ folgt mit der zweiten Aussage von Borel-Cantelli $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Der Vergleich von Integralen mit Reihen aus Analysis 1 (für die monoton fallende Funktion $f(t) = \mathbb{P}(|X_1| > t)$ mit $f(n) = \mathbb{P}(|X_1| > n) \stackrel{\text{u.i.v.}}{=} \mathbb{P}(|X_n| > n) = \mathbb{P}(A_n)$) impliziert damit

$$\mathbb{E}[|X_1|] = \int_0^{\infty} \mathbb{P}(|X_1| > t) dt < \infty.$$

Zur Erinnerung: Der Satz der Analysis besagt

$$\sum_{n=1}^{\infty} f(n) < \infty \quad \Leftrightarrow \quad \int_1^{\infty} f(x) dx < \infty.$$

Dabei haben wir eine Übungsaufgabe benutzt: Es gilt nämlich für jede nicht-negative Zufallsvariable Y die Identität $\mathbb{E}[Y] = \int_0^{\infty} \mathbb{P}(Y > t) dt$. Das folgt aus Fubini, wenn man die Wahrscheinlichkeit als Erwartungswert $\mathbb{E}[\mathbf{1}_{(t, +\infty)}(Y)]$ schreibt.

Damit ist die erste Aussage gezeigt. Weil nun $\mathbb{E}[|X_1|] < \infty$ gilt, ist die Voraussetzung von Satz 4.6.8 erfüllt und die zweite Aussage folgt. □

Vorlesung 27

Anwendung 4.6.11. ▶ **[Empirisches Gesetz der großen Zahlen]**

Seien X_1, X_2, \dots unabhängig und identisch verteilt, so gilt

$$\frac{1}{n} \#\{k \leq n: X_k \in A\} \xrightarrow{\text{f.s.}} \mathbb{P}(X_1 \in A), \quad n \rightarrow \infty,$$

für $A \in \mathcal{B}(\mathbb{R})$. Insbesondere gilt mit der Wahl $A = (-\infty, t]$

$$\frac{1}{n} \#\{k \leq n: X_k \leq t\} \xrightarrow{\text{f.s.}} F(t), \quad n \rightarrow \infty,$$

und mit der Wahl $A = \{a_l\}$ für diskrete Zufallsvariablen mit Werten $\{a_1, \dots, a_N\}$ und Wahrscheinlichkeiten p_1, \dots, p_N

$$\frac{1}{n} \#\{k \leq n: X_k = a_l\} \xrightarrow{\text{f.s.}} p_l, \quad n \rightarrow \infty.$$

Beweis. Wir definieren $Y_k := \mathbf{1}_A(X_k)$. Die Y_i sind u.i.v. (siehe Korollar 4.2.28) mit endlichem Erwartungswert (sie sind durch 1 beschränkt). Berechnen wir noch den Erwartungswert: $\mathbb{E}[Y_1] = \mathbb{E}[\mathbf{1}_A(X_1)] = \mathbb{P}(X_1 \in A)$. Also kann das Gesetz der großen Zahlen angewandt werden und es gilt, weil Y_k nur die Werte 0 und 1 annehmen,

$$\frac{1}{n} \#\{k \leq n: X_k \in A\} = \frac{1}{n} \sum_{k=1}^n Y_k \xrightarrow{\text{f.s.}} \mathbb{E}[Y_1] = \mathbb{P}(X_1 \in A), \quad n \rightarrow \infty$$

□

In der Statistik nennt man $F_n(t) := \frac{1}{n} \#\{k \leq n: X_k \leq t\}$, $t \in \mathbb{R}$, empirische Verteilungsfunktion der Stichprobe X_1, \dots, X_n . Das Korollar besagt, dass die empirische Verteilungsfunktion punktweise gegen die Verteilungsfunktion konvergiert, wenn die Beobachtungsgröße wächst. Wozu ist das nützlich? Stellen wir uns vor, wir könnten ein Experiment beobachten, kennen aber nicht die Verteilungsfunktion der beschreibenden Zufallsvariablen. Wenn wir die Verteilungsfunktion $F(t)$ „schätzen“ wollen, beobachten wir also möglichst viele unabhängige Ausführungen des Experiments und nehmen $F_n(t)$ als Schätzwert für $F(t)$. Fragen dieser Art werden in Vorlesungen der Statistik und Ökonometrie thematisiert.

Anwendung 4.6.12. [Monte-Carlo-Methode]

Numerische Methoden die darauf basieren, eine „gesuchte Größe“ μ als Erwartungswert irgendeiner Zufallsvariablen zu schreiben und diese mit dem Gesetz der Großen Zahlen als $\frac{1}{N} \sum_{k=1}^N X_k$ zu approximieren, heißen **Monte-Carlo Methoden**. Als Beispiel schauen wir uns die Berechnung von Integralen $\int_0^1 f(x) dx$ mit einer Monte-Carlo Methode an. Sei f integrierbar und $U \sim \mathcal{U}([0, 1])$.

Dann ist die Zufallsvariable $f(U)$ integrierbar mit

$$\mathbb{E}[f(U)] = \int_{\mathbb{R}} f(x) \cdot \mathbf{1}_{[0,1]}(x) dx = \int_0^1 f(x) dx.$$

Der Monte-Carlo Ansatz zur Berechnung funktioniert nun wie folgt. Sind U_1, U_2, \dots u.i.v. mit $U_1 \sim \mathcal{U}([0, 1])$, so gilt

$$\frac{1}{n} \sum_{k=1}^n f(U_k) \xrightarrow{\text{f.s.}} \int_0^1 f(x) dx, \quad n \rightarrow \infty.$$

Weil die Konvergenz fast sicher ist, müssen wir also „nur“ auf dem Computer eine möglichst lange Realisierung $U_1(\omega), \dots, U_N(\omega)$ uniformer Zufallsvariablen erzeugen und $\frac{1}{N} \sum_{k=1}^N f(U_k(\omega))$ als Approximation von $\int_0^1 f(x) dx$ nehmen. Wie man an solch eine Realisierung $U_1(\omega), \dots, U_N(\omega)$ im Computer rankommt, lernt ihr am Anfang der Vorlesung Monte Carlo Methoden.

Anwendung 4.6.13. [Momentenschätzer]

Eine Grundfrage der Statistik ist folgende: Gegeben sei eine Verteilung einer parametrischen Klasse von Verteilungen $\{F_\theta : \theta \in \Theta\}$. Wir kennen den Parameter nicht, können aber unabhängige Zufallsvariablen unserer Verteilung beobachten. Stellt euch vor, ihr habt ein unfaire Münze, kennt

aber die Wahrscheinlichkeit für Kopf nicht. Ihr habt also eine Verteilung aus der parametrischen Klasse $\{\text{Ber}(p) : p \in (0, 1)\}$ und wüsstet gerne den Parameter p . In manchen Fällen hilft das GGZ, und zwar dann, wenn der unbekannte Parameter θ mit dem Erwartungswert zusammenhängt. In dem Beispiel der Münze gilt beispielsweise $p = \mathbb{E}[X_1]$. Wenn ihr also eine u.i.v. Folge der Verteilung beobachten könnt (in dem Beispiel werft ihr immer wieder die Münze), so gibt das starke GGZ euch den Parameter der Verteilung: $p = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k$. Für ein festes N nennt man deshalb $\hat{p}_N := \frac{1}{N} \sum_{k=1}^N X_k$ daher einen Schätzer (Schätzwert) von p . Ganz analog geht man zum Beispiel vor, wenn man von vornherein weiß, was die unbekannte Verteilung $\mathcal{N}(\mu, \sigma^2)$ ist, für einen unbekanntem Parameter μ . Dann wäre $\hat{\mu}_N = \frac{1}{N} \sum_{k=1}^N X_k$ ein Schätzer von μ . Fragen dieser Art lernt ihr in der Stochastik 2 Vorlesung viel genauer kennen!

Die letzten zwei Anwendungen kommen aus zwei verschiedenen Gebieten, der Numerik und der Statistik. Ihr habt vielleicht gemerkt, dass die Fragestellungen in gewisser Art sehr ähnlich sind und in diesen einfachen Beispielen aus dem GGZ motiviert sind. In der stochastischen Numerik spricht man von Realisierungen, in der Statistik eher von Beobachtungen, gemeint ist immer $X(\omega)$. Der strukturelle Unterschied von stochastischer Numerik und Statistik ist eigentlich nur folgender: In der Statistik wird angenommen, dass man aus dem echten Leben eine Beobachtung von Zufallsvariablen hat (z. B. ökonomische Kenngrößen), in der stochastischen Numerik wird eine Realisierung der Zufallsvariablen selber erzeugen. Mit beiden Sichtweisen kann man mittels GGZ etwas über die gleiche Kenngröße aussagen (den Erwartungswert der Zufallsvariablen), die Anwendungen haben jedoch völlig unterschiedliche Motivationen.

4.7 Zentraler Grenzwertsatz

In diesem letzten Abschnitt besprechen wir den zentralen Grenzwertsatz (ZGS). Als Motivation stellen wir uns folgende Frage: Wir wollen wie im letzten Beispiel das Integral $\int_0^1 f(x) dx$ numerisch approximieren, indem wir auf dem Computer viele uniforme Zufallsvariablen erzeugen. In der Realität können wir n nicht gegen unendlich schicken, sagen wir also N ist eine große feste Zahl, z. B. $N = 10000$. Wir fragen nun:

$$\text{Wie gut ist die Näherung } \frac{1}{10000} \sum_{k=1}^{10000} f(U_k) \quad \text{von} \quad \int_0^1 f(x) dx \quad ?$$

Die Antwort ist leider: nicht so gut. Schauen wir uns dazu zunächst den zentralen Grenzwertsatz an:

Satz 4.7.1. [Zentraler Grenzwertsatz]

Sind X_1, X_2, \dots u.i.v. Zufallsvariablen mit $\mathbb{E}[X_1] = \mu$ und endlicher Varianz $\sigma^2 := \mathbb{V}[X_1] > 0$. Dann gilt

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n} \cdot \sigma} \xrightarrow{(d)} Y, \quad n \rightarrow \infty,$$

mit $Y \sim \mathcal{N}(0, 1)$.

Oft sieht man den zentralen Grenzwertsatz auch mit Wahrscheinlichkeiten geschrieben als

$$\mathbb{P}\left(\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n} \cdot \sigma} \leq t\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx =: \Phi(t), \quad n \rightarrow \infty,$$

oder

$$\mathbb{P}\left(a \leq \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n} \cdot \sigma} \leq b\right) \rightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a), \quad n \rightarrow \infty.$$

Diese Umformulierungen folgen aus Satz 4.5.9 und der Stetigkeit der Verteilungsfunktion Φ der Normalverteilung.

Beweis. Wir geben den Beweis nur unter der stärkeren Annahme $\mathbb{E}[|X_1^3|] < \infty$. Das Argument funktioniert auch für $\mathbb{E}[X_1^2] < \infty$ (was gerade $\mathbb{V}[X_1] < \infty$ entspricht), wäre aber länger und komplizierter. Um das Hauptargument kompakt zu halten, besprechen wir zunächst vier Zutaten:

- (i) Standardisierung: Ähnlich wie im Beweis des starken GGZ (dort haben wir zentriert) können wir ohne Einschränkung $\mu = 0$ und $\sigma = 1$ annehmen, um die Notation zu vereinfachen. Dazu definieren wir $Z_k = \frac{X_k - \mu}{\sigma}$ und bemerken, dass Z standardisiert ist (Linearität des Erwartungswertes und Verschiebungsformel der Varianz) sowie $\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\cdot\sigma} = \frac{\sum_{k=1}^n Z_k}{\sqrt{n}}$ gilt.
- (ii) Wir erinnern an die Skalierungs- und Faltungseigenschaften der Normalverteilung: Sind Y_1, \dots, Y_n u.i.v. $\mathcal{N}(0, 1)$, so gilt $\frac{\sum_{k=1}^n Y_k}{\sqrt{n}} \sim \mathcal{N}(0, 1)$, siehe 4.3.16 und 4.3.18.
- (iii) Statt die Konvergenz in Verteilung für alle beschränkten stetigen Funktionen zu zeigen, zeigen wir sie nur für glatte f weil wir dafür Taylor nutzen können. Es reicht, die Konvergenz für $f \in C^3(\mathbb{R})$ mit beschränkten f', f'', f''' zu zeigen. Warum? Das Argument haben wir im Beweis von Satz 4.5.9 schon gesehen: Wir approximieren die Indikatorfunktionen $\mathbf{1}_{(-\infty, t]}$ durch Funktionen f_+ und f_- , dieses mal aber nicht stückweise linear sondern glatt. Genau wie im Beweis der Hinrichtung von Satz 4.5.9 bekommen wir also die punktweise Konvergenz der Verteilungsfunktionen. Weil die Grenzverteilungsfunktion Φ stetig ist, ist das nach der Rückrichtung von Satz 4.5.9 gerade die Konvergenz in Verteilung.
- (iv) Wir wenden auf die Funktionen in (iii) Taylor mit der Restglieddarstellung aus dem Mittelwert an. Für $f \in C^3(\mathbb{R})$ gilt, für $x, x_0 \in \mathbb{R}$,

$$f(x + x_0) = f(x_0) + f'(x_0)x + \frac{f''(x_0)}{2}x^2 + \frac{f'''(\tilde{x})}{6}x^3$$

für ein \tilde{x} zwischen x und x_0 .

Auf geht's: Es sei nun X_1 zentriert, f wie in (iii) und Y_1, Y_2, \dots eine u.i.V. Folge mit $X_1 \sim \mathcal{N}(0, 1)$, die unabhängig von der Folge X_1, X_2, \dots ist. Dann gilt mit einem coolen Teleskop Trick:

$$\begin{aligned} & \left| \mathbb{E} \left[f \left(\frac{\sum_{k=1}^n X_k}{\sqrt{n}} \right) \right] - \mathbb{E} [f(Y)] \right| \\ & \stackrel{(ii)}{=} \left| \mathbb{E} \left[f \left(\frac{\sum_{k=1}^n X_k}{\sqrt{n}} \right) - f \left(\frac{\sum_{k=1}^n Y_k}{\sqrt{n}} \right) \right] \right| \\ & \stackrel{\text{Teleskop}}{=} \left| \sum_{i=1}^n \left(\mathbb{E} \left[f \left(\frac{\sum_{k=1}^{i-1} X_k + X_i + \sum_{k=i+1}^n Y_k}{\sqrt{n}} \right) \right] - \mathbb{E} \left[f \left(\frac{\sum_{k=1}^{i-1} X_k + Y_i + \sum_{k=i+1}^n Y_k}{\sqrt{n}} \right) \right] \right) \right|. \end{aligned}$$

Auf alle $2n$ auftretenden Funktionen f wenden wir jetzt Taylor (ω -weise) wie in (iii) an, wobei wir jedes Mal $X_0 := \sum_{k=1}^{i-1} X_k + \sum_{k=i+1}^n Y_k$ wählen. Exemplarisch stehen dort ω -weise

$$f(X_0(\omega)) + f'(X_0(\omega))X_i(\omega) + f''(X_0(\omega))\frac{X_i(\omega)^2}{2} + f'''(\tilde{X}_i(\omega))\frac{X_i(\omega)^3}{6}$$

mit einem $\tilde{X}_i(\omega) \in \mathbb{R}$ aus dem Taylor-Restglied. Als Erwartungswert geschrieben, bekommen wir lauter Summanden der Form

$$\mathbb{E} \left[f(X_0) + f'(X_0)X_i + f''(X_0)\frac{X_i^2}{2} + f'''(\tilde{X}_i)\frac{X_i^3}{6} \right]$$

bzw.

$$\mathbb{E} \left[f(X_0) + f'(X_0)Y_i + f''(X_0)\frac{Y_i^2}{2} + f'''(\tilde{Y}_i)\frac{Y_i^3}{6} \right].$$

In der Differenz fallen die $f(X_0)$ -Terme weg. Die f' -Terme fallen weg weil alle Zufallsvariablen zentriert sind und daher aufgrund der angenommenen Unabhängigkeit $\mathbb{E}[f'(X_0)X_k] = \mathbb{E}[f'(X_0)]\mathbb{E}[X_k] = 0$ gilt (genauso für die Y_k). Auch die $f''(X_0)X_0^2/2$ -Terme fallen in der Differenz weg, weil wieder aufgrund der Zentrierung und Unabhängigkeit $\mathbb{E}[f''(X_0)X_k^2] = \mathbb{E}[f''(X_0)]\mathbb{E}[X_k^2] = \mathbb{E}[f''(X_0)]$ und $\mathbb{E}[f''(X_0)Y_k^2] = \mathbb{E}[f''(X_0)]\mathbb{E}[Y_k^2] = \mathbb{E}[f''(X_0)]$ gelten. Wenn wir nun alles einsetzen und zurecht kürzen, bekommen wir

$$\begin{aligned} & \left| \mathbb{E} \left[f \left(\frac{\sum_{k=1}^n X_k}{\sqrt{n}} \right) \right] - \mathbb{E} [f(Y)] \right| \\ &= \left| \mathbb{E} \left[\sum_{i=1}^n \left(\frac{f'''(\tilde{X}_i)}{6} \frac{X_i^3}{n^{3/2}} - \frac{f'''(\tilde{Y}_i)}{6} \frac{Y_i^3}{n^{3/2}} \right) \right] \right| \\ &\stackrel{\Delta}{\leq} \mathbb{E} \left[\sum_{i=1}^n \left(\left| \frac{f'''(\tilde{X}_i)}{6} \frac{X_i^3}{n^{3/2}} \right| + \left| \frac{f'''(\tilde{Y}_i)}{6} \frac{Y_i^3}{n^{3/2}} \right| \right) \right] \\ &\leq \frac{\sup_{x \in \mathbb{R}} |f'''(x)|}{6} \sum_{i=1}^n \frac{\mathbb{E}[|X_i|^3] + \mathbb{E}[|Y_i|^3]}{n^{3/2}} \\ &\stackrel{\text{u.i.v.}}{\leq} \frac{\sup_{x \in \mathbb{R}} |f'''(x)|}{6} \frac{\mathbb{E}[|X_1|^3] + \mathbb{E}[|Y_1|^3]}{n^{1/2}} \\ &\rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Aufgrund von (iii) ist der Beweis damit beendet. \square

Der Beweis war nicht sehr lang, aber äußerst komplex weil viel Verständniss verschiedener Objekte der Vorlesung nötig sind. Weil viele wichtige Themen der Vorlesung benutzt werden, könnt ihr viele beim Verstehen des Beweises viel wiederholen.

Bemerkung 4.7.2.  Wegen Satz 4.7.1 bezeichnet man $\mathcal{N}(0, 1)$ als **universelle Verteilung**. Wir haben nur $\mathbb{V}[X_1] < \infty$ angenommen – nichts weiter über die Verteilung. Dennoch kommt immer die Normalverteilung als Grenzwert raus!

In den Anwendungen des GGZ approximieren wir jeweils einen festen Wert durch eine Folge von Zufallsvariablen. Typischerweise wird für großes N zwar $\frac{1}{N} \sum_{k=1}^N X_k \approx \mathbb{E}[X_1]$ gelten, jedoch nicht $\frac{1}{N} \sum_{k=1}^N X_k = \mathbb{E}[X_1]$, die Approximation $\frac{1}{N} \sum_{k=1}^N X_k$ ist schließlich eine Zufallsvariable. Es fragt sich also, wie sehr diese Zufallsvariable um $\mathbb{E}[X_1]$ konzentriert ist. Dabei hilft uns der zentrale Grenzwertsatz in der anders geklammerten Form

$$\frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{k=1}^n X_k - \mathbb{E}[X_1] \right) \stackrel{(d)}{\rightarrow} Y, \quad n \rightarrow \infty.$$

Es gilt mit der Skalierungseigenschaft der Normalverteilung also für große N ungefähr

$$\frac{1}{N} \sum_{k=1}^N X_k \approx \frac{\sigma}{\sqrt{N}} Y + \mathbb{E}[X_1] \sim \mathcal{N} \left(\mathbb{E}[X_1], \frac{\sigma^2}{N} \right). \quad (4.9)$$

Wenn wir jetzt an unsere Konzentrationsungleichungen zurückdenken, so gilt in etwa: Mit Wahrscheinlichkeit 0,997 ist die Abweichung $\left| \frac{1}{N} \sum_{k=1}^N X_k - \mathbb{E}[X_1] \right|$ kleiner als $\frac{3\sigma}{\sqrt{N}}$.

Anwendung 4.7.3.  **[Monte Carlo Methoden]**

Für die stochastische Numerik besagt die obige Diskussion, dass die Konvergenzordnung von Monte Carlo Verfahren (siehe Anwendung 4.6.12) $\frac{1}{2}$ ist, der Approximationsfehler also in der Wurzel der Simulationsgröße gegen 0 konvergiert. Das ist extrem langsam, $\frac{1}{\sqrt{10000}} = 0,01$ ist nicht sehr klein! Monte Carlo Verfahren sind daher grundsätzlich nicht sehr gut, funktionieren allerdings auch in schwierigen Situation immer (!) mit der Konvergenzordnung $\frac{1}{2}$. Wir sehen aus (4.9) auch, dass die Varianz der gewählten Approximationsfolge eine Rolle spielt, kleines σ

gibt mehr Konzentration um den gesuchten Wert $\mu = \mathbb{E}[X_1]$. Tricks, die aus einer gegebenen u.i.v. Folge X_1, X_2, \dots mit $\mathbb{E}[X_1] = \mu$ eine neue u.i.v. Folge $\hat{X}_1, \hat{X}_2, \dots$ mit $\mathbb{E}[\hat{X}_1] = \mu$ und kleinerer Varianz machen, nennt man Varianzreduktionsmethoden. Solche Tricks könnt ihr in der Monte Carlo Methoden Vorlesung kennenlernen.

Beispiel 4.7.4. ▶ Wir schauen uns mal den Zusammenhang zur Binomialverteilung an, sagen wir $n = 20$ und $p = \frac{1}{2}$. Was ist $\mathbb{P}(X \leq 12)$ für $X \sim \text{Bin}(20, \frac{1}{2})$? Die Wahrscheinlichkeit können wir natürlich als $\sum_{k=0}^{12} p_k$ mit $p_k = \binom{20}{k} \frac{1}{2^{20}}$ durch Einsetzen mühsam von Hand ausrechnen, das gibt $\frac{910596}{2^{20}} \approx 0,8684$. Alternativ machen wir das mit dem ZGS weil $\text{Bin}(20, \frac{1}{2})$ sich auch als Summe von 20 unabhängigen $\text{Ber}(\frac{1}{2})$ Zufallsvariablen X_1, \dots, X_{20} schreiben lässt. Diese haben Erwartungswert $\mu = \frac{1}{2}$ und Varianz $\sigma^2 = \frac{1}{4}$, also bekommen wir durch Erweitern

$$\mathbb{P}(X \leq 12) = \mathbb{P}\left(\sum_{k=1}^{20} X_k \leq 12\right) = \mathbb{P}\left(\frac{\sum_{k=1}^{20} X_k - 20 \cdot \frac{1}{2}}{\sqrt{20} \cdot \frac{1}{2}} \leq \frac{12 - 20 \cdot \frac{1}{2}}{\sqrt{20} \cdot \frac{1}{2}}\right) \stackrel{\text{ZGS}}{\approx} \Phi(0,8944).$$

Jetzt sucht ihr euch eine Tabelle für die Verteilungsfunktion Φ der Standardnormalverteilung (oder eine App) und findet etwa 0,8133. So richtig gut ist das nicht, aber 20 ist auch nicht sonderlich groß und wie oben besprochen ist die Konvergenz im ZGS langsam.

Ganz zum Schluss noch ein Beispiel, bei dem der zentrale Grenzwertsatz (also die Stochastik) mit der Zahlentheorie/Kombinatorik zusammentrifft. Mega!

Beispiel 4.7.5. ▶ [Stirlingformel für Fakultäten]

Habt ihr euch schon mal gefragt, wie groß $n!$ oder konkret $1000!$ ist? Tatsächlich riesig, aber wie riesig? $n!$ ist für großes n ziemlich genau $n^n e^{-n} \sqrt{2\pi n}$. Genauer, es gilt

$$\frac{n!}{n^n e^{-n} \sqrt{2\pi n}} \rightarrow 1, \quad n \rightarrow \infty. \quad (4.10)$$

Warum, was hat das mit dem ZGS zu tun? Man nimmt dazu X_1, X_2, \dots u.i.v. mit $X_1 \sim \text{Poi}(1)$, es gilt mit der diskreten Faltungformel also $\sum_{k=1}^n X_k \sim \text{Poi}(n)$. Mit $f(x) = x^+$ benutzt man nun den ZGS:

$$\mathbb{E}\left[f\left(\frac{\sum_{k=1}^n X_k - n}{\sqrt{n}}\right)\right] \rightarrow \mathbb{E}[f(Y)] = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x e^{-\frac{x^2}{2}} dx, \quad n \rightarrow \infty.$$

Das uneigentliche Integral auf der rechten Seite ist 1 (Substitution) und die Summe auf der linken Seite mit der Berechnungsformel aus Satz 4.1.11 gerade

$$\begin{aligned} \mathbb{E}\left[f\left(\frac{\sum_{k=1}^n X_k - n}{\sqrt{n}}\right)\right] &\stackrel{\text{Poi}(n)}{=} \sum_{k=1}^{\infty} f\left(\frac{k-n}{\sqrt{n}}\right) e^{-n} \frac{n^k}{k!} \\ &= \sum_{k=n+1}^{\infty} \left(\frac{k-n}{\sqrt{n}}\right) e^{-n} \frac{n^k}{k!} \\ &= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n+1}^{\infty} k \frac{n^k}{k!} - \sum_{k=n+1}^{\infty} n \frac{n^k}{k!} \right) \\ &= \frac{e^{-n}}{\sqrt{n}} \left(\sum_{k=n+1}^{\infty} \frac{n^{k-1} n}{(k-1)!} - \sum_{k=n+1}^{\infty} n \frac{n^k}{k!} \right) \\ &= \frac{e^{-n} n}{\sqrt{n}} \left(\sum_{k=n+1}^{\infty} \frac{n^k}{k!} - \sum_{k=n}^{\infty} \frac{n^k}{k!} \right) = \frac{e^{-n} \sqrt{n} n^n}{n!}. \end{aligned}$$

Das war es schon, wenn man $\sqrt{2\pi}$ vom Grenzwert rüberzieht. Eine kleine Warnung: Die Funktion $f(x) = x^+$ ist zwar stetig, aber nicht beschränkt, passt also nicht zur Definition der Konvergenz in Verteilung. Wenn wir aber in unseren Beweis des ZGS schauen, bekommen wir die Konvergenz auch für x^+ hin, indem wir mit einer glatten Funktion f den Knick approximieren.

ENDE

Kapitel 5

Sammlung von Fakten und Beispielen zu Zufallsvariablen

In diesem Teil des Appendix sammeln wir wesentliche Fakten und Beispiele. Am Ende des Tages ist Stochastik nicht nur abstrakte Maßtheorie, ihr müsst auch mit Beispielen rumrechnen können und ein Gefühl für Eigenschaften bekommen.

Allgemeine Definitionen für beliebige Zufallsvariablen:

- Für eine Zufallsvariable X auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ schreibt man $X \sim F$, falls

$$\mathbb{P}(X \in (a, b]) = \mathbb{P}_X((a, b]) = F(b) - F(a).$$

Wozu haben wir Maßtheorie betrieben? Um die Existenz von stochastischen Modellen (W-Raum und Zufallsvariable) zu beweisen! Wegen Carathéodory und der Borel- σ -Algebra gibt es für alle Verteilungsfunktionen Zufallsvariablen mit $X \sim F$.

- Wenn das Integral existiert, definiert man für $g: \mathbb{R} \rightarrow \overline{\mathbb{R}}$ messbar $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(X(\omega)) d\mathbb{P}(\omega)$ und es gilt aufgrund des Transformationssatzes $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) d\mathbb{P}_F(x)$. Einige Spezialfälle bekommen eigene Namen:
 - Erwartungswert für $g(x) = x$
 - k.tes Moment für $g(x) = x^k$
 - exponentielles Moment für $g(x) = e^{\lambda x}$

Wozu haben wir Integrationstheorie betrieben? Um Erwartungswerte und ähnliches für beliebige Zufallsvariablen zu definieren! Ohne Integrationstheorie ginge das nur für diskrete Zufallsvariablen oder Zufallsvariablen mit Dichten.

Absolutstetige Zufallsvariablen

Definitionen und Fakten:

- $X \sim F$ heißt absolutstetig, falls F eine Dichte f hat. Es gilt dann

$$\mathbb{P}(X \in (a, b]) = \int_a^b f(x) dx.$$

Weil $\{a\} = [a, a]$, gilt für absolutstetige Zufallsvariablen

$$\mathbb{P}(X = a) = \int_a^a f(x) dx = 0$$

und daher auch

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in (a, b)) = \mathbb{P}(X \in [a, b)) = \mathbb{P}(X \in (a, b]),$$

es gibt also keine Masse auf einpunktigen Mengen.

- Momente lassen sich einfach berechnen. Für messbare $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ ist

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x) dx,$$

wenn eine der Seiten (und damit die andere) definiert ist.

Wichtige Bemerkung für das Verständnis: Wenn wir irgendeine nichtnegative Funktion f mit $\int_{\mathbb{R}} f(x) dx < \infty$ kennen, so bekommen wir mit $h(x) := Cf(x)$ eine Dichte, wobei $\frac{1}{C} = \int_{\mathbb{R}} f(x) dx$ ist. Unsere Beispiele bestehen also im Prinzip aus den integrierbaren nichtnegativen Funktionen, die wir gut integrieren können. Das sind die Exponentialfunktion (gibt die Exponentialverteilung), Exponentialfunktion mit einem Quadrat (gibt die Normalverteilung), Varianten der Funktion $1/x^2$ (gibt die Cauchy Verteilung) oder konstante Funktionen auf kompakten Mengen (gibt die uniformen Verteilungen).

Normalverteilung - $\mathcal{N}(\mu, \sigma^2)$

Die wichtigste Verteilung der Stochastik ist die Normalverteilung. Das liegt am zentralen Grenzwertsatz.

Vorteile:

- Alle Momente existieren und können berechnet werden.
- Die Verteilungsfunktion ist zwar nicht explizit, dennoch kann vieles ausgerechnet werden.
- Zwei Parameter μ und σ^2 geben viel Freiheit, die Verteilung an echte Daten anzupassen.
- Die Verteilung von Summen unabhängiger Normalverteilungen kann einfach mit der momenterzeugenden Funktion berechnet werden.
- Normalverteilungen können einfach verschoben und skaliert werden: Ist $X \sim \mathcal{N}(0, 1)$, so ist $\sigma X + \mu =: Z \sim \mathcal{N}(\mu, \sigma^2)$.

Nachteil:

- Verteilungsfunktion ist nicht explizit gegeben, Wahrscheinlichkeiten der Form $\mathbb{P}(X \in [a, b])$ müssen daher abgeschätzt werden (Konzentrationsungleichungen).

Wichtige Charakteristiken auf einen Blick:

Parameter	$\mu \in \mathbb{R}$ (Verschiebung), $\sigma^2 > 0$ (Stauchung)
Wertebereich	\mathbb{R}
Verteilungsfunktion	$F_{\mu, \sigma^2}(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$
Dichte	$f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Grobe Verteilung der Masse	Viel Masse nah bei μ , extrem wenig bei $+\infty$ und $-\infty$.
Erwartungswert	$\mathbb{E}[X] = \mu$
Varianz	$\mathbb{V}[X] = \sigma^2$
Momenterzeugende Funktion	definiert für alle $t \in \mathbb{R}$, $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

Zum rumspielen hier eine interaktive Graphik. Zu beachten ist: Viel Masse liegt dort, wo die Dichte groß ist. Das spiegelt sich dadurch wieder, dass dort die Verteilungsfunktion stark wächst. Andersrum: Wenig Masse liegt dort, wo die Verteilungsfunktion flach ist bzw. die Dichte klein ist.

Exponentialverteilung - $\text{Exp}(\lambda)$

Wichtige Verteilung, besonders in der Anwendungen bei Markovprozessen.

Vorteile:

- Alle Momente existierten und können berechnet werden.
- Verteilungsfunktion und Dichte sind explizit und sehr einfach. Vieles kann berechnet werden.

Nachteil:

- Nur ein Parameter um Verteilung auf Daten anzupassen

Wichtige Charakteristiken auf einen Blick:

Parameter	$\lambda > 0$ (Stauchung)
Wertebereich	$[0, \infty)$
Verteilungsfunktion	$F_\lambda(t) = (1 - e^{-\lambda t}) \mathbf{1}_{[0, \infty)}(t)$
Dichte	$f_\lambda(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, \infty)}(x)$
Grobe Verteilung der Masse	Viel Masse nah bei 0, wenig bei $+\infty$
Erwartungswert	$\mathbb{E}[X] = \frac{1}{\lambda}$
Varianz	$\mathbb{V}[X] = \frac{1}{\lambda^2}$
Momenterzeugende Funktion	definiert für alle $t < \lambda$, $M_X(t) = \frac{\lambda}{\lambda - t}$

Uniforme Verteilung - $\mathcal{U}([a, b])$

Modell für das einfachste Experiment, Ziehen aus einem Intervall ohne Bevorzugung verschiedener Bereiche.

Vorteile:

- Alle Momente existierten und können berechnet werden.
- Verteilungsfunktion und Dichte sind explizit und sehr einfach. Vieles kann berechnet werden.
- Man kann alle anderen Verteilungen aus $\mathcal{U}([0, 1])$ gewinnen, theoretisch reicht also die uniforme Verteilung für alles aus!

Nachteile:

- Gar kein Parameter um Verteilung auf Daten anzupassen
- Eher langweilig.

Wichtige Charakteristiken auf einen Blick:

Parameter	keine Parameter
Wertebereich	$[a, b]$
Verteilungsfunktion	$F(t) = \frac{t-a}{b-a} \mathbf{1}_{[a, b]}(t) + \mathbf{1}_{[b, \infty)}(t)$
Dichte	$f(x) = \frac{1}{b-a} \mathbf{1}_{[a, b]}(x)$
Grobe Verteilung der Masse	uniform auf $[a, b]$ (daher der Name)
Erwartungswert	$\mathbb{E}[X] = \frac{a+b}{2}$
Varianz	$\mathbb{V}[X] = \frac{(b-a)^2}{12}$
Momenterzeugende Funktion	definiert für alle $t \in \mathbb{R}$, $M_X(t) = \begin{cases} \frac{e^{bt} - e^{at}}{(b-a)t} & : t \neq 0 \\ 1 & : t = 0 \end{cases}$

Gammaverteilung - $\Gamma(\alpha, \beta)$

Vorteile:

- Sehr flexibel durch zwei Parameter, erlaubt Modellierung sehr unterschiedlicher Masseverteilungen mit einem Modell. Ein Parameter beschreibt Verhalten bei 0, einer bei $+\infty$
- Exponentialverteilung ist Spezialfall: $\text{Exp}(\lambda) = \Gamma(1, \lambda)$
- Alle Momente existieren und können berechnet werden, Momenterzeugende Funktion kann problemlos abgeleitet werden.
- Viel kann berechnet werden.

Nachteil:

- Verteilungsfunktion nicht explizit.

Wichtige Charakteristiken auf einen Blick:

Parameter	$\alpha > 0$, (Masse bei 0) $\beta > 0$ (Masse bei $+\infty$)
Wertebereich	$(0, \infty)$
Verteilungsfunktion	nicht explizit
Dichte	$f_{\alpha, \beta}(x) = 1_{(0, \infty)}(x) \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
Grobe Verteilung der Masse	wenig Masse bei $+\infty$ (Dichte fällt exp.) viel Masse bei 0 für $p < 1$, wenig für $p > 0$
Erwartungswert	$\mathbb{E}[X] = \frac{\alpha}{\beta}$
Varianz	$\mathbb{V}[X] = \frac{\alpha}{\beta^2}$
Momenterzeugende Funktion	definiert für $t < \beta$, $M_X(t) = (\beta/(\beta - t))^\alpha$

Cauchyverteilung - Cauchy(s, t)

Die Cauchyverteilung ist eine sogenannte „heavy-tailed“ Verteilung, d. h. viel Masse liegt bei unendlich. Für uns hat die Verteilung eigentlich nur Nachteile, es geht so ziemlich alles schief.

Vorteile:

- Zwei Parameter s und t geben viel Freiheit, die Verteilung an echte Daten anzupassen.
- Verteilungsfunktion und Dichte sind explizit.

Nachteile

- Klassisches Beispiel für eine Verteilung, deren Erwartungswert nicht wohldefiniert ist.

Wichtige Charakteristiken auf einen Blick:

Parameter	$t \in \mathbb{R}$ (Verschiebung), $s > 0$ (Stauchung)
Wertebereich	\mathbb{R}
Verteilungsfunktion	$F_{s,t}(r) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{r-t}{s}\right)$
Dichte	$f_{s,t}(x) = \frac{1}{\pi} \frac{s}{s^2 + (x-t)^2}$
Grobe Verteilung der Masse	viel Masse bei 0, aber auch viel bei $+\infty$ und $-\infty$
Erwartungswert	nicht wohldefiniert!
Varianz	existiert nicht!
Momenterzeugende Funktion	für kein $t \in \mathbb{R}$ definiert

Paretoverteilung - Pareto(k, a)

Wie die Cauchyverteilung ist die Paretoverteilung eine sogenannte „heavy-tailed“ Verteilung, d. h. viel Masse liegt bei unendlich. Im Gegensatz zur Cauchy-Verteilung hat die Paretoverteilung nur Masse auf den positiven reellen Zahlen. Etwas genauer, sogar nur auf den positiven reellen Zahlen, die größer als ein vorgegebener Wert a liegen. In der Statistik wird die Verteilung sehr wichtig sein, sie ist eine sogenannte Extremwertverteilung.

Vorteile:

- Die einzige Verteilung dieser Vorlesung mit Werten in $[a, \infty)$ für ein beliebig wählbares $a \geq 0$.
- Verteilungsfunktion und Dichte sind ganz besonders einfach, man kann gut rechnen.

Nachteile

- Es gibt keinen endlichen höheren Momente, die Momenterzeugende Funktion ist nicht berechenbar.

Wichtige Charakteristiken auf einen Blick:

Parameter	$k > 0$ (kontrolliert Masse bei unendlich), $a > 0$ (kleinster Wert)
Wertebereich	$[a, \infty)$
Verteilungsfunktion	$F_{k,a}(t) = (1 - \frac{a^k}{t^k}) \mathbf{1}_{[a,\infty)}(t)$
Dichte	$f_{k,a}(x) = k \frac{a^k}{x^{k+1}} \mathbf{1}_{[a,\infty)}(x)$
Grobe Verteilung der Masse	viel Masse bei $+\infty$, keine Masse in $(-\infty, a)$
Erwartungswert	$+\infty$ für $k \leq 1$, $a \frac{k}{k-1}$ für $k < 1$
Varianz	existiert nicht für $k \leq 2$, $a^2 \frac{k}{(k-2)(k-1)^2}$ für $k > 2$
Momenterzeugende Funktion	gibt keine Formel

Diskrete Zufallsvariablen

Definitionen und Fakten:

- X heißt diskret, wenn X endlich oder abzählbar viele Werte a_1, \dots, a_N mit Wahrscheinlichkeiten p_1, \dots, p_N annimmt. Der einfachste Weg eine diskrete Zufallsvariable zu beschreiben, ist den Wertebereich $\{a_1, \dots, a_N\}$ der angenommenen Werte und die dazugehörigen Wahrscheinlichkeiten (Wahrscheinlichkeitsgewichte) anzugeben. Mit der Verteilungsfunktion formuliert ist X diskret, falls die Verteilungsfunktion F diskret ist:

$$F(t) = \sum_{k=1}^N p_k \mathbf{1}_{[a_k, \infty)}(t), \quad t \in \mathbb{R},$$

wobei $N \in \mathbb{N} \cup \{+\infty\}$, $a_i \in \mathbb{R}$ für $i = 1, \dots, N$ und $\sum_{k=1}^N p_k = 1$. Für Berechnungen nutzen wir immer die Formel

$$\mathbb{P}(X \in A) = \sum_{a_k \in A} \mathbb{P}(X = a_k) = \sum_{a_k \in A} p_k$$

für alle $A \in \mathcal{B}(\mathbb{R})$. Für diskrete Verteilungen kennen wir das Maß \mathbb{P}_F expliziert, es gilt

$$\mathbb{P}_F = \sum_{k=1}^N p_k \delta_{a_k}.$$

Im Gegensatz zu absolutstetigen Zufallsvariablen hat die Verteilung diskrete Zufallsvariablen also durchaus Masse auf einpunktigen Mengen, nämlich gerade auf den Mengen $\{a_1\}, \dots$

- Momente diskreter Zufallsvariablen lassen sich durch Summen berechnen. Für messbare $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ ist

$$\mathbb{E}[g(X)] = \sum_{k=1}^N p_k g(a_k).$$

Man summiert also alle möglichen Werte von $g(X)$ multipliziert mit den Wahrscheinlichkeiten.

Wie bei absolutstetigen Verteilungen können wir uns aus den Analysis 1 Kenntnissen im Prinzip alle klassischen diskreten Verteilungen erraten. Es gibt zwei wesentliche Fälle:

- $N < \infty$: Man wähle irgendwelche $a_1, \dots, a_N \in \mathbb{R}$ und denke sich irgendwelche $p_1, \dots, p_N \geq 0$ aus, die sich zu 1 addieren.
- $N = +\infty$: Man wähle irgendeine Folge $a_1, a_2, \dots \in \mathbb{R}$ und benutze eine bekannte konvergente Reihe $\sum_{k=1}^{\infty} b_k$ aus Analysis 1. Ist $C = \sum_{k=1}^{\infty} b_k$, so setzt man $p_k = \frac{1}{C} b_k$. Schon haben wir eine diskrete Verteilung! Wenn wir jetzt an Analysis 1 denken, fallen uns nur wenige konvergente Reihen mit positiven Summanden ein. Das ist die geometrische Reihe (gibt die geometrische Verteilung) und die die Exponentialreihe (gibt die Poisson Verteilung). Das war es eigentlich schon.

Bernoulli Verteilung - $\text{Ber}(p)$

Vorteile:

- Total einfach, Modell für den Münzwurf einer unfairen Münze.
- Alles, wirklich alles, lässt sich sofort ausrechnen.

Nachteil:

- Komplette langweilig!

Wichtige Charakteristiken auf einen Blick:

Parameter	$p \in [0, 1]$
Wertebereich	0 oder 1
Wahrscheinlichkeitsgewichte	$p_1 = p, p_0 = q$ wobei $q = 1 - p$
Grobe Verteilung der Masse	Anteil p der Masse auf der 1, Rest auf der 0
Erwartungswert	$\mathbb{E}[X] = p$
Varianz	$\mathbb{V}[X] = pq$
Momenterzeugende Funktion	definiert für alle $t \in \mathbb{R}$, $M_X(t) = 1 - p + pe^t$

Binomial Verteilung - $\text{Bin}(n, p)$

Vorteil:

- Modell für Ziehen mit Zurücklegen, klare Interpretation (p_k ist die Wahrscheinlichkeit, bei n Versuchen, k Treffer zu landen).
- Viele Formeln berechenbar. Es gibt einen einfachen Beweis für den zentralen Grenzwertsatz.

Nachteil:

- Teilweise fuzzielig zum Rechnen. Muss immer irgendwie die Binomialformel ins Spiel bringen.

Wichtige Charakteristiken auf einen Blick:

Parameter	$n \in \mathbb{N}, p \in [0, 1]$
Wertebereich	$\{0, \dots, n\}$
Wahrscheinlichkeitsgewichte	$p_k = \binom{n}{k} p^k (1-p)^{n-k}$
Grobe Verteilung der Masse	viel Masse in der Mitte von $\{0, \dots, n\}$, wenig bei 0 und n
Erwartungswert	$\mathbb{E}[X] = np$
Varianz	$\mathbb{V}[X] = npq$
Momenterzeugende Funktion	definiert für alle $t \in \mathbb{R}$, $M_X(t) = (pe^t + 1 - p)^n$

Poisson Verteilung $\text{Poi}(\lambda)$

Vorteile:

- Viele Formeln berechenbar, ist auch eine gute Reihe!

Wichtige Charakteristiken auf einen Blick:

Parameter	$\lambda > 0$
Wertebereich	\mathbb{N}_0
Wahrscheinlichkeitsgewichte	$p_k = e^{-\lambda} \frac{\lambda^k}{k!}$
Grobe Verteilung der Masse	viel Masse auf kleinen Zahlen, wenig bei $+\infty$.
Erwartungswert	$\mathbb{E}[X] = \lambda$
Varianz	$\mathbb{V}[X] = \lambda$
Momenterzeugende Funktion	definiert für alle $t \in \mathbb{R}$, $M_X(t) = \exp(\lambda(e^t - 1))$

Geometrische Verteilung $\text{Geo}(p)$

Vorteile:

- Viele Formeln berechenbar, ist auch eine gute Reihe!

Nachteil:

- Teilweise fuzzielig zum Rechnen. Viel Indexverschiebung als analog zur Substitution im Fall der Dichten.

Wichtige Charakteristiken auf einen Blick:

Parameter	$p > 0$
Wertebereich	$\mathbb{N} \setminus \{0\}$
Wahrscheinlichkeitsgewichte	$p_k = p(1-p)^{k-1}$
Grobe Verteilung der Masse	viel Masse auf kleinen Zahlen, wenig bei $+\infty$.
Erwartungswert	$\mathbb{E}[X] = \frac{1}{p}$
Varianz	$\mathbb{V}[X] = \frac{1-p}{p^2}$
Momenterzeugende Funktion	definiert für alle $t \in \mathbb{R}$, $M_X(t) = \frac{pe^t}{1-(1-p)e^t}$