Stochastik 1

Leif Döring

Universität Mannheim

Inhaltsverzeichnis

Was machen wir, was nicht?

Vorlesung 1

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

Modellierung und theoretische Untersuchung zufälliger Experimente

$\Updownarrow$

Wahrscheinlichkeitstheorie ($\rightsquigarrow$ Döring)

•

Anpassung der Modelle auf „echte“ zufällige Experimente

$\Updownarrow$

Mathematische Statistik ($\rightsquigarrow$ Schlather)

•

Ausführung der Modelle („Zufall erzeugen“)

$\Updownarrow$

Stochastische Numerik ($\rightsquigarrow$ Neuenkirch)

•

Anwendung auf Finanzmärkte

$\Updownarrow$

Finanzmathematik ($\rightsquigarrow$ Prömel)

•

Anwendung auf Wirtschaftsdaten

$\Updownarrow$

Ökonometrie ($\rightsquigarrow$ Trenkler, Rothe)

•

Zählen von Möglichkeiten $\rightsquigarrow$ Gleichverteilung (z. B. Lotto; Ziehen aus Urnen)

$\Updownarrow$

Kombinatorik

Teil 1: Maß- und Integrationstheorie

Kapitel 1

Maßtheorie

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 Standdardbeispiel 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. $A^C = \{w \in \Omega \,|\, w \notin A\}$. $\mathcal{P}(\Omega)$ 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) $A_1, A_2, ... \in \mathcal{A} \Rightarrow \bigcup\limits_{k=1}^{\infty} A_k \in \mathcal{A}$, 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.

- $\mathcal{A} = \{\emptyset, \Omega\}$

- $\mathcal{A} = \{\emptyset, \Omega, A, A^C\}$ für $A \subseteq \Omega$ beliebig
- $\mathcal{A} = \{A \subseteq \Omega \mid A$ oder A^C 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 $A_1, A_2, ...$ 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 $\mathcal{A}$. Ist eine der Mengen nicht abzählbar, sagen wir A_j, so ist das Komplement A_j^C abzählbar. Doch dann ist wegen

$$\Big(\bigcup_{n=1}^{\infty} A_i\Big)^C = \bigcap_{n=1}^{\infty} A_i^C \subseteq A_j^C$$

das Komplement der Vereinigung nach Analysis 1 abzählbar.

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

(i) $\emptyset \in \mathcal{A}$

(ii) $A_1, A_2, ... \in \mathcal{A} \Rightarrow \bigcap\limits_{n=1}^{\infty} A_n \in \mathcal{A}$

(iii) Aus $A, B \in \mathcal{A}$ folgt A $\setminus B := A \cap B^C \in A$ 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 <u>nicht</u> 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 $+\infty$.

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

(i) $\mu(\emptyset) = 0$

(ii) Sind $A_1, A_2, ... \in \mathcal{A}$ paarweise disjunkte Mengen, so gilt $\mu(\dot\bigcup\limits_{n=1}^{\infty} A_n) = \sum\limits_{n=1}^{\infty} \mu(A_n)$. 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

$$\mu\Big(\dot{\bigcup_{n=1}^{k}} A_n\Big) = \sum_{n=1}^{k} \mu(A_n).$$

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

Bemerkung 1.1.5. Oft werden Wahrscheinlichkeitsmaße mit $\mathbb{P}$ anstelle von μ geschrieben und **Verteilungen** genannt.

Definition 1.1.6.

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

Bemerkung. 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 1.1.7. (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 = „Gegenereigniss"
- $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$ bedeutet „Eintreten von Ereigniss bekannt, impliziert Eintreten von Gegenereigniss bekannt."
- $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$ bedeutet „Gegenereigniss 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 Addititivität 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 $\Omega = \{1, ..., 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ und $\mathbb{P}(\{1\}) = ... = \mathbb{P}(\{6\}) = \frac{1}{6}$. Ein Ereigniss $A \in \mathcal{A}$ bedeutet also „Eine der Zahlen in A ist gewürfelt worden". Die Wahrscheinlichkeiten aller weiteren Ereignisse sind festgelegt, indem das Ereigniss in die disjunkten Elementarereignisse zerlegt wird, z. B. die Wahrscheinlichkeit eine gerade Zahl zu würfeln:

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

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, ist folgende kleine Rechnung wichtig:

Lemma 1.1.8. (Monotonie)
Seien $A, B \in \mathcal{A}, B \subseteq A$ und μ ein Maß auf A. Dann ist $\mu(B) \leq \mu(A)$.

Beweis. $\mu(B) \leq \mu(B) + \mu(A \backslash B) = \mu(A)$, wobei wir beide definierenden Eigenschaften des Maßes genutzt haben. □

Kommen wir nun zu ein paar Beispielen:

Beispiel 1.1.9. (endliche Gleichverteilung)
Sei $\#\Omega < \infty$ und $\mathcal{A} = \mathcal{P}(\Omega)$. Dann heißt $\mu(A) = \frac{\#A}{\#\Omega}$ 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 1.1.10. (abzählbare Verteilungen, Zählmaß)
Sei Ω abzählbar, also ohne Einschränkung $\Omega = \mathbb{N}$. Wir wählen $\mathcal{A} = \mathcal{P}(\Omega)$ und eine Folge $(p_k)_{k\in\mathbb{N}}$ reeller Zahlen. Definieren wir

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

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 $\sum_{k=0}^{\infty} p_k = \mu(\mathbb{N}) = 1$ gelten. In dem Fall würden wir wieder $\mathbb{P}$ statt μ schreiben.

- Ist $p_k = 1$ für alle $k \in \mathbb{N}$ heißt μ **Zählmaß** weil $\mu(A) = \#A$ die Anzahl der Elemente von A zählt.

Beispiel 1.1.11. (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 $k \in \mathbb{N}$,

- $\sum\limits_{k=0}^{\infty} p_k = \sum\limits_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum\limits_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = e^{-\lambda+\lambda} = 1.$

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 1.1.12. (Diracmaß)
Sei $\mathcal{A}$ eine σ-Algebra auf Ω und $x \in \Omega$, so heißt

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

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

(i) Aufgrund der Definition gilt natürlich $\delta_x(\emptyset) = 0$.

(ii) Für disjunkte Mengen $A_1, A_2, ...$ gilt

$$\delta_x\Big(\dot\bigcup_{n=1}^{\infty} A_n\Big) = \begin{cases} 1, & x \in \dot\bigcup\limits_{n=1}^{\infty} A_n \\ 0, & x \notin \dot\bigcup\limits_{n=1}^{\infty} A_n \end{cases} = \sum_{n=1}^{\infty}\delta_x(A_n),$$

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

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

(i) Aus $A_n \uparrow A$ (d. h. $A_1 \subseteq A_2 \subseteq A_3 \subseteq ...$, $\bigcup_{n=1}^{\infty} A_n = A$) folgt $\lim\limits_{n\to\infty} \mu(A_n) = \mu(A)$.

(ii) Aus μ endlich und $A_n \downarrow A$ (d. h. $A_1 \supseteq A_2 \supseteq ...$, $\bigcap\limits_{n=1}^{\infty} A_n = A$) folgt $\lim\limits_{n\to\infty} \mu(A_n) = \mu(A)$.

Beweis.

(i) Definiere

$$\begin{aligned} A_1' &:= A_1 \\ A_2' &:= A_2 \backslash A_1 \\ A_n' &:= A_n \backslash A_{n-1}, \quad n \geq 3. \end{aligned}$$

Weil die A_n' paarweise disjunkt sind und $A_n = \dot{\bigcup\limits_{i=1}^{n}} A_i'$ gilt, folgt

$$\begin{aligned} \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \mu\Big(\dot{\bigcup_{i=1}^{n}} A_i'\Big) &\overset{\substack{\text{Def.}\\ \text{Maß}}}{=} \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A_i') \overset{\text{Def.}}{=} \sum_{i=1}^{\infty} \mu(A_i') \\ &\overset{\substack{\text{Def. Maß}\\ \text{und disj.}}}{=} \mu\Big(\dot{\bigcup_{i=1}^{\infty}} A_i'\Big) = \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \mu(A). \end{aligned}$$

(ii) Folgt sofort aus (i) weil

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

Weil μ endlich ist gilt für alle messbaren Mengen

$$\mu(\Omega) = \mu(A \,\dot{\cup}\, A^C) \overset{\text{Maß}}{=} \mu(A) + \mu(A^C)$$

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

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

□

Beispiel 1.1.14. (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 $A_n = \{n, n+1, ...\}$ 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 1.1.13 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 $\mathcal{A} := \bigcap_{i\in I} A_i$. Wir checken die drei Eigenschaften einer σ-Algebra für $\mathcal{A}$:

(i) $\emptyset \in \mathcal{A}$ ist klar weil $\emptyset \in \mathcal{A}_i$ für alle $i \in I$ und damit ist $\emptyset$ 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

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

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

□

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.

(i) Existenz:

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

tut's.

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

□

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

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

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 $\mathcal{E} = \{\{x\} \colon x \in \Omega\}$. Dann ist

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

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. 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.

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

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

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

gelten. Warum reicht das? Dazu nutzen wir zwei Eigenschaften, die direkt aus der Definition der erzeugen 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 erzeuge σ-Algebra enthält auch all die Mengen selbst. Umgekehrt existieren für jedes Element x einer offenen Menge O irgendwelche $a_1, ...a_d, b_1, ...b_d \in \mathbb{Q}$ mit $x \in (a_1, b_1) \times ... \times (a_d, b_d) \subseteq O$. Damit gilt:

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

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) $A_1, A_2, ... \in \mathcal{D}$ **paarweise disjunkt** $\Rightarrow \dot{\bigcup}_{k=1}^{\infty} A_k \in \mathcal{D}$

Im Gegensatz zu einer σ-Algebra ist ein Dynkin-System als nur abgeschlossen bezüglich paarweise disjunkter Vereinigungen.

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 $(\Omega, \mathcal{A})$ 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 $A_1, A_2, ... \in \mathcal{M}$ 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

  $$\mu_1\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big) = \sum_{n=1}^{\infty} \mu_1(A_n) = \sum_{n=1}^{\infty} \mu_2(A_n) = \mu_2\Big(\dot{\bigcup_{n\in\mathbb{N}}} A_n\Big),$$

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

Satz 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 $A, B \in \mathcal{D},\ B \subseteq \mathcal{A} \Rightarrow A \backslash B \in \mathcal{D}$ weil

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

(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

$$A \cup B = A \,\dot{\cup}\, (B \backslash (A \cap B)) \in \mathcal{D}.$$

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

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

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

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

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

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

$\square$

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 $\sigma(\mathcal{E})$ das kleinste Dynkin-System ist das $\mathcal{E}$ enhält, 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 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) in $d(\mathcal{E})$ sein. Wir müssen also nur noch zeigen, dass $d(\mathcal{E})$ $\cap$-stabil, wenn $\mathcal{E}$ $\cap$-stabil ist:

(a) Definiere dazu zunächst

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

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

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

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

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

(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 $d(\mathcal{E}) \subseteq d(\mathcal{D}_D) \overset{(a)}{=} \mathcal{D}_D$ 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 $d(\mathcal{E})$.

$\square$

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.

Korollar 1.2.12. Sei $(\Omega, \mathcal{A})$ 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 haben schon gezeigt, dass

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

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. $\square$

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

Korollar 1.2.13. (Von endlich nach σ-endlich)
Es sei $(\Omega, \mathcal{A})$ 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 $E_n \uparrow \Omega$, $n \to \infty$, 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 $E_n \uparrow \Omega$, $n \to \infty$, 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 ausdehen. Beispiele folgen noch.

Beweis von 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 $\mu_i^n(\Omega) = \mu_i(\Omega \cap E_n) = \mu_i(E_n) \overset{\text{Ann.}}{<} \infty$. Nach 1.2.11 gilt $\mu_1^n = \mu_2^n$ für alle $n \in \mathbb{N}$. Nun gilt wegen Stetigkeit von Maßen

$$\begin{aligned}\mu_1(A) = \mu_1(A \cap \Omega) = \mu_1\Big(A \cap \bigcup_{n=1}^{\infty} E_n\Big) = \mu_1\Big(\bigcup_{n=1}^{\infty}(A \cap E_n)\Big) \overset{1.1.13}{=} \lim_{n\to\infty} \mu_1(A \cap E_n)\\ \overset{\text{Def.}}{=} \lim_{n\to\infty} \mu_1^n(A) = \lim_{n\to\infty} \mu_2^n(A) \overset{\text{Def.}}{=} \lim_{n\to\infty} \mu_2(A \cap E_n) \overset{1.1.13}{=} \mu_2\Big(A \cap \bigcup_{n=1}^{\infty} E_n\Big) = \mu_2(A).\end{aligned}$$

□

So, endlich ein 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:

- $0 \le F_{\mathbb{P}} \le 1$
- $F_{\mathbb{P}}$ ist nicht fallend
- $\lim_{t\to+\infty} F_{\mathbb{P}}(t) = 1$
- $\lim_{t\to-\infty} F_{\mathbb{P}}(t) = 0$

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

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

Das folgt aus 1.2.12 mit $\mathcal{E} = \{(-\infty, t] \colon t \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R})$. 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 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 $\mu_1 = \mu_2$.

1.3 Konstruktion von Maßen

Gerade haben wir gesehen, dass zwei endliche Maße auf $\mathcal{B}(\mathbb{R})$ schon auf den Intervallen eindeutlich 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“

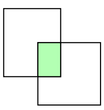
(iii) $A, B \in \mathcal{S}$, $B \subseteq A \Rightarrow$ es gibt paarweise disjunkte Mengen $C_1, ..., C_n \in \mathcal{S}$ mit

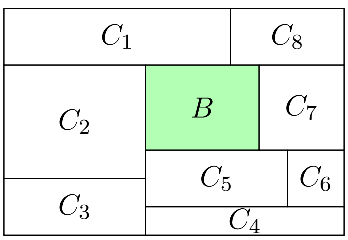
$$A \backslash B = \dot{\bigcup_{k=1}^{\infty}} C_k.$$

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 $\emptyset = (a_1, a_1) \times ... \times (a_d, a_d)$ für beliebige $a_1, ..., a_d \in \mathbb{R}$.

(ii) 

(iii) 

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 $B_1, B_2, ..., B_r \subseteq A$ paarweise disjunkt, so existieren $C_1, ..., C_n \in \mathcal{S}$ paarweise disjunkt mit

$$A \backslash (B_1 \mathbin{\dot\cup} ... \mathbin{\dot\cup} B_r) = \dot{\bigcup_{k=1}^{n}} C_k.$$

Beweis. Vollständige Induktion bezüglich r.

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

IV: Gelte die Behauptung für ein beliebiges $r \in \mathbb{N}$.

IS: Es folgt für $A, B_1, ..., B_{r+1} \in \mathcal{S}$, $B_1, ..., B_{r+1} \subseteq A$, $B_i \cap B_j = \emptyset$, $i \neq j$. Nach Induktionsvoraussetzung und Rechenregeln mit Schnitten von Mengen gilt

$$A \backslash \dot{\bigcup_{i=1}^{r+1}} B_i = A \cap \bigcap_{i=1}^{r+1} B_i^C = A \backslash \dot{\bigcup_{i=1}^{r}} B_i \cap B_{r+1}^C \overset{\text{IA}}{=} \dot{\bigcup_{k=1}^{n}} [C_k \backslash B_{r+1}] = \dot{\bigcup_{k=1}^{n}} [C_k \backslash (C_k \cap B_{r+1})].$$

Nun existieren für alle $1 \leq k \leq n$ wegen $C_k \cap B_{r+1} \in \mathcal{S}$ paarweise disjunkte Mengen $C_{k,m} \in \mathcal{S}$, $m \leq n_k$, mit

$$C_k \backslash (C_k \cap B_{r+1}) = \dot{\bigcup_{m=1}^{n_k}} C_{k,m}.$$

Also gilt

$$A \backslash \bigcup_{i=1}^{r+1} B_i = \dot{\bigcup_{k=1}^{n}} \dot{\bigcup_{m=1}^{n_k}} C_{k,m}.$$

Da die Mengen $C_{k,m}$, $(1 \leq k \leq n, 1 \leq m \leq n_k)$, paarweise disjunkt sind, haben wir die gewünschte Darstellung für $(r+1)$ gefunden.

□

Weiter geht es mit einer Verschärfung von Semiringen:

Definition 1.3.3. $\mathcal{R} \in \mathcal{P}(\Omega)$ heißt Ring, falls

(i) $\emptyset \in \mathcal{R}$

(ii) $A, B \in \mathcal{R} \Rightarrow A \backslash B \in \mathcal{R}$

(iii) $A, B \in \mathcal{R} \Rightarrow A \cup B \in \mathcal{R}$

Bemerkung.

(i) A^C ist nicht unbedingt in $\mathcal{R}$, denn $\Omega \in \mathcal{R}$ muss nicht gelten. Das sieht harmlos aus, macht uns das Leben aber deutlich schwieriger.

(ii) Mit vollständiger Induktion sind auch endliche Vereinigungen wieder in $\mathcal{R}$.

Vorlesung 4

Bemerkung 1.3.4. $\mathcal{R}$ Ring $\Rightarrow$ $\mathcal{R}$ Semiring.

Beweis.

(i) ✓

(ii) Folgt sofort aus der Beobachtung $A \cap B = A \backslash (A \backslash B)$.

(iii) Folgt sofort mit $n = 1$, denn $A \backslash B \in \mathcal{R}$.

□

Um ein besseres Gefühl für die Definitionen zu bekommen, zeigen wir, dass die Menge der altbekannten disjunkten Vereinigungen von Quadern ein Ring bildet. Allgemeiner geht das für beliebige Semigringe:

Bemerkung 1.3.5. Ist $\mathcal{S}$ ein Semiring, so ist die Menge aller endlicher disjunkter Vereinigungen

$$\mathcal{R} := \left\{ \dot{\bigcup_{i=1}^{n}} A_i : n \in \mathbb{N},\ A_1, ..., A_n \in \mathcal{S} \right\}$$

der kleinste Ring, der $\mathcal{S}$ enthält.

Beweis. Dass $\mathcal{R}$ der *kleinste* Ring ist, ist klar (es gilt $\mathcal{S} \subseteq \mathcal{R}$ mit $n = 1$ und damit müssen aufgrund der Eigenschaft eines Rings alle endlichen Vereinigungen wieder enthalten sein). Wir müssen also noch zeigen, dass $\mathcal{R}$ tatsächlich ein Ring ist. Checken wir also die Eigenschaften:

(i) ✓

(ii) Seien dazu $A, B \in \mathcal{R}$, also

$$A = \dot{\bigcup_{i=1}^{n}} A_i \quad \text{und} \quad B = \dot{\bigcup_{j=1}^{m}} B_j.$$

Dann gilt

$$A \backslash B = \dot{\bigcup_{i=1}^{n}} \underbrace{\Big(A_i \backslash \dot{\bigcup_{j=1}^{m}} \underbrace{(A_i \cap B_i)}_{\substack{\in \mathcal{S},\text{ weil} \\ \mathcal{S}\ \cap\text{-stabil}}} \Big)}_{\overset{1.3.2}{=} \dot{\bigcup}_{k=1}^{r} C_{i,k}}$$
$$= \dot{\bigcup_{i=1}^{n}} \dot{\bigcup_{k=1}^{r_i}} C_{i,k}.$$

(iii) Seien A, B wie in (ii), so gilt $A \cup B = \underbrace{(A \backslash B)}_{\in \mathcal{R}} \,\dot{\cup}\, \underbrace{B}_{\in \mathcal{R}} \overset{\text{Def.}}{\in} \mathcal{R}$ weil eine disjunkte Vereinigungen disjunkter Mengen in $\mathcal{S}$ wieder eine disjunkte Vereinigung von Mengen in $\mathcal{S}$ ist.

□

Definition 1.3.6. $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt Algebra, falls $\mathcal{A}$ ein Ring ist und $\Omega \in \mathcal{A}$.

Die Definition einer Algebra ist äquivalent zur Forderung

(i) $\emptyset \in \mathcal{A}$,

(ii) $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$,

(iii) $A_1, ..., A_n \in \mathcal{A} \Rightarrow \bigcup\limits_{i=1}^{n} A_i \in \mathcal{A}$,

weil $\Omega \backslash A = A^C$.

Definition 1.3.7. $\mathcal{R} \subseteq \mathcal{P}(\Omega)$ heißt σ-Ring, falls (iii) in der Definition des Rings durch

$$A_1, A_2, ... \in \mathcal{R} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{R}$$

ersetzt wird. Das σ steht also wieder für „abzählbar viele“.

Ein σ-Ring $\mathcal{R}$ mit $\Omega \in \mathcal{R}$ ist also nichts anderes als eine σ-Algebra.

Lemma 1.3.8. (Rechenregeln ähnlich zu Maßen auf Semiringen)
Sei $\mathcal{S}$ ein Semiring und $\mu \colon \mathcal{S} \to [0, \infty]$ eine Mengenfunktion mit

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

Dann gilt

(i) Monotonie: $\mu(A) \leq \mu(B)$ für alle $A, B \in \mathcal{S}$ mit $A \subseteq B$.

(ii) „**Subadditivität**“: Sind $A, A_1, A_2, ... \in \mathcal{S}$ und $A \subseteq \bigcup\limits_{n=1}^{\infty} A_n$, so gilt

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

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 in den Eigenschaften wieder in $\mathcal{S}$ liegen. Sonst wäre $\mu(A)$ gar nicht definiert. Auch zu beachten ist, dass Komplemente nicht automatisch in $\mathcal{S}$ liegen. Daher sind leichte Eigenschaften für Maße auf σ-Algebren, nicht so klar für Mengenfunktionen auf Semiringen.

Beweis.

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

$$B \backslash A = \bigcup_{i=1}^{n} C_i.$$

Damit gilt wegen der Additivität von μ

$$\begin{aligned}\mu(B) &= \mu(A \mathbin{\dot\cup} (B \backslash A)) = \mu(A \mathbin{\dot\cup} C_1 \mathbin{\dot\cup} ... \mathbin{\dot\cup} C_n) = \mu(A) + \mu(C_1) + ... + \mu(C_n)\\ &\geq \mu(A).\end{aligned}$$

(ii) Erst machen wir die A_n disjunkt:

$$\begin{aligned}A_1' &:= A_1,\\ A_2' &:= A_2 \backslash A_1',\\ &\dots\\ A_n' &:= A_n \backslash (A_1' \mathbin{\dot\cup} ... \mathbin{\dot\cup} A_{n-1}'), \quad n \geq 4.\end{aligned}$$

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

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

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

$$A_n \backslash A_n' = \dot{\bigcup_{n=1}^{m_n}} D_{n,k}. \tag{1.3}$$

Damit gelten

- $A \cap C_{n,j} \in \mathcal{S}$ weil $\mathcal{S}$ $\cap$-stabil ist,
- $$A = \dot{\bigcup_{n=1}^{\infty}} A_n' \cap A = \dot{\bigcup_{n=1}^{\infty}} \dot{\bigcup_{j=1}^{m}} A \cap C_{n,j}$$

  weil $A \subseteq \bigcup_{n=1}^{\infty} A_n = \dot{\bigcup_{n=1}^{\infty}} A_n'$ angenommen wurde,
- $$A_n = A_n' \mathbin{\dot\cup} (A_n \backslash A_n') \overset{(1.2)}{\underset{(1.3)}{=}} \dot{\bigcup_{j=1}^{l_n}} C_{n,j} \mathbin{\dot\cup} \dot{\bigcup_{n=1}^{m_n}} D_{n,k}.$$

Alles zusammen ergibt die Subadditivität:

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

Folgende Animation macht das Argument an einem Beispiel mit Quadern bzw. Rechtecken besser verständlich[1]:

$\square$

Definition 1.3.9. $\mu^* : \mathcal{P}(\Omega) \to [0, \infty]$ heißt **äußeres Maß**, falls

(i) $\mu^*(\emptyset) = 0$

(ii) $A \subseteq B \subseteq \Omega \Rightarrow \mu^*(A) \leq \mu^*(B)$

[1]Zum Abspielen wird der Adobe Acrobat Reader empfohlen. Eventuell muss 3D-content erlaubt und anschließend das Dokument neu geöffnet werden.

(iii)

$$A_1, A_2, \ldots \in \Omega \Rightarrow \mu^*\Big(\bigcup_{n=1}^{\infty} A_n\Big) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$$

Definition 1.3.10. Sei μ^* ein äußeres Maß auf Ω. Dann heißt $A \in \Omega$ **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 $\mathcal{A}_{\mu^*}$.

Satz 1.3.11.

(i) $\mathcal{A}_{\mu^*}$ ist eine σ-Algebra.

(ii) μ^* eingeschränkt auf $\mathcal{A}_{\mu^*}$ ist ein Maß.

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

(a)

(i) $\emptyset \in \mathcal{A}_{\mu^*}$. Sei dazu $Z \subseteq \Omega$. Dann gilt

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

(ii) Sei $A \in \mathcal{A}_{\mu^*}$, dann ist (+ drehen) auch $A^C \in \mathcal{A}_{\mu^*}$.

(iii) Seien $A_1, A_2 \in \mathcal{A}_{\mu^*}$. Sei $Z \subseteq \Omega$ beliebig, dann folgt mit $Z' := Z \cap A_2$

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

Folglich gilt auch

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

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

Damit ist $\mathcal{A}_{\mu^*}$ eine Algebra. Vorlesung 5

(b) Seien also $A_1, A_2, \ldots \in \mathcal{A}_{\mu^*}$, zu zeigen ist $\bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}_{\mu^*}$. 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 disjunkten Mengen

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

und beachten, dass damit $\bigcup_{n=1}^{\infty} A_n = \dot{\bigcup}_{n=1}^{\infty} A'_n$ gilt. Wenn also die Vereinigung der disjunkten A'_n wieder in $\mathcal{A}_{\mu^*}$ ist, ist auch die Vereinigung über die A_n in $\mathcal{A}_{\mu^*}$. 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 $A_1, A_2, ...$ paarweise disjunkt sind, so sparen wir uns die $'$ in den Gleichungen. Aufgrund der Definition von $\mathcal{A}_{\mu^*}$ wählen wir ein $Z \subset \Omega$ beliebig. Wir zeigen erstmal induktiv

$$\mu^*(Z) = \sum_{i=1}^{n} \mu^*(Z \cap A) + \mu^*\Big(Z \cap \bigcap_{i=1}^{n} A_i^C\Big), \quad \forall n \in \mathbb{N}. \tag{1.4}$$

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

$$\begin{aligned}\mu^*\Big(Z \cap \bigcap_{i=1}^{n} A_i^C\Big) &= \mu^*\Big(Z \cap \bigcap_{i=1}^{n} A_i^C \cap A_{n+1}\Big) + \mu^*\Big(Z \cap \bigcap_{i=1}^{n} A_i^C \cap A_{n+1}^C\Big) \\ &= \mu^*\Big(Z \cap \bigcap_{i=1}^{n+1} A_i^C\Big) + \mu^*(Z \cap A_{n+1}).\end{aligned}$$

Einsetzen der (aufgelösten) IV in die rechte Seite gibt

$$\mu^*(Z) - \sum_{i=1}^{n} \mu^*(Z \cap A_i) = \mu^*\Big(Z \cap \bigcap_{i=1}^{n+1} A_i^C\Big) + \mu^*(Z \cap A_{n+1}),$$

also

$$\mu^*(Z) = \sum_{i=1}^{n+1} \mu^*(Z \cap A_i) + \mu^*\Big(Z \cap \bigcap_{i=1}^{n+1} A_i^C\Big).$$

Damit ist die Induktion gezeigt.

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

$$\mu^*(Z) \geq \sum_{i=1}^{n} \mu^*(Z \cap A_i) + \mu^*\Big(Z \cap \bigcap_{i=1}^{\infty} A_i^C\Big), \quad \forall n \in \mathbb{N}.$$

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

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

Für die letzten beiden Ungleichungen haben wir 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

$$\mu^*(Z) = \mu^*\Big(Z \cap \bigcup_{i=1}^{\infty} A_i\Big) + \mu^*\Big(Z \cap \Big(\bigcup_{i=1}^{\infty} A_i\Big)^C\Big).$$

Weil Z beliebig war, ist damit

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}_{\mu^*}$$

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

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

$$\mu^*(Z) = \sum_{i=1}^{\infty} \mu^*(Z \cap A_i) + \mu^*\Big(Z \cap \bigcap_{i=1}^{\infty} A_i^C\Big).$$

Wählen wir $Z = \bigcup\limits_{i=1}^{\infty} A_i$, so gilt wegen $\mu(\emptyset) = 0$

$$\mu^*\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu^*(A_i) + 0$$

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

$\square$

Kommen wir endlich zum Höhepunkt der ersten Wochen:

Satz 1.3.12. (Fortsetzungssatz von Carathéodory)
Sei $\mathcal{S}$ ein Semiring, $\mu \colon \mathcal{S} \to [0, \infty]$ mit

- $\mu(\emptyset) = 0$,
- μ ist σ-additiv.

Dann existiert ein Maß $\bar{\mu}$ auf $\sigma(S)$ 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

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

Wir zeigen nacheinander (a) $\mu*$ 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 $\mathcal{A}_{\mu^*}$ eine σ-Algebra ist und μ^* ein Maß auf $\mathcal{A}_{\mu^*}$ ist. Wegen (b) gilt $\sigma(S) \subseteq \mathcal{A}_{\mu^*}$, weil die kleinste σ-Algebra die $\mathcal{S}$ enthält, auch Teilmenge von allen σ-Algebras ist, die $\mathcal{S}$ enthalten. Damit ist auch die Einschränkung von μ^* auf $\sigma(S)$ ein Maß (das nennen wir dann $\bar{\mu}$) und wegen (c) ist $\bar{\mu}$ eine Fortsetzung von μ.

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

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

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

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

$$A \overset{\text{Def.}}{=} \bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A_{n,k},$$

gilt

$$\mu^*(A) \underset{\text{als inf}}{\overset{\text{Def. } \mu^*}{\leq}} \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mu(A_{n,k}) \leq \sum_{n=1}^{\infty} \left(\mu^*(A_n) + \frac{\varepsilon}{2^n} \right) = \sum_{n=1}^{\infty} \mu^*(A_n) + \underbrace{\sum_{n=1}^{\infty} \frac{\varepsilon}{2^n}}_{\epsilon}.$$

Beachte: Weil das Infimum einer Menge kleiner oder gleich jedem Element der Menge ist, gilt nach Definition $\mu^*(A) \leq \sum \mu(A_k)$ für beliebige Überdeckungen $A \subseteq \bigcup A_k$ durch Mengen $A_k \in \mathcal{S}$. Weil ε beliebig gewählt wurde, gilt damit

$$\mu^* \Big(\bigcup_{n=1}^{\infty} A_n \Big) = \mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n).$$

Daraus folgt Subadditivität und damit ist μ^* ein äußeres Maß.

(b) Nächster Schritt: Wir zeigen $\mathcal{S} \subseteq \mathcal{A}_{\mu^*}$. Dazu brauchen wir $\mu^*(z) = \mu^*(z \cap S) + \mu^*(z \cap S^C)$ für alle $z \in \Omega$, $S \in \mathcal{S}$.

„$\leq$“: $\mu^*(z) = \mu^*(z \cap S \,\dot\cup\, z \cap S^C) \leq \mu^*(z \cap S) + \mu^*(z \cap S^C)$

„$\geq$“: Sei $(B_n) \in \mathcal{S}$ mit

$$z \subseteq \bigcup_{n=1}^{\infty} B_n$$

Weil $\mathcal{S}$ ein Semiring ist, existieren $C_{n,1}, ..., C_{n,m_n} \in \mathcal{S}$ mit

$$B_n \cap S^C = B_n \setminus \underbrace{B_n \cap S}_{\substack{\in \mathcal{S}, \text{ denn} \\ \mathcal{S}\ \cap\text{-stabil}}} = \bigcup_{k=1}^{m_n} C_{n,k}.$$

Es gilt:

- $$z \cap S \subseteq \Big(\bigcup_{n=1}^{\infty} B_n\Big) \cap S = \Big(\bigcup_{n=1}^{\infty} B_n \cap S\Big)$$
- $$z \cap S^C \subseteq \Big(\bigcup_{n=1}^{\infty} B_n \cap S^C\Big).$$

Mit den Definitionen folgt

$$\begin{aligned}
\mu^*(z \cap S) + \mu^*(z \cap S^C) &\overset{\text{subadd.}}{\leq} \sum_{n=1}^{\infty} \mu^*(B_n \cap S) \sum_{n=1}^{\infty} \mu^*(B_n \cap S^C) \\
&\overset{\text{subadd.}}{\leq} \sum_{n=1}^{\infty} \Big(\mu^* \underbrace{(B_n \cap S)}_{\in \mathcal{S}} \sum_{k=1}^{m_n} \mu^* \underbrace{(C_{n,k})}_{\in \mathcal{S}}\Big) \\
&= \sum_{n=1}^{\infty} \Big(\mu^*(B_n \cap S) \sum_{k=1}^{m_n} \mu^*(C_{n,k})\Big) \\
&\overset{\sigma\text{-add.}}{=} \sum_{n=1}^{\infty} \mu\Big(B_n \cap S \,\dot\cup\, \dot{\bigcup_{k=1}^{m_n}} C_{n,k}\Big) = \sum_{n=1}^{\infty} \mu(B_n \cap S \,\dot\cup\, B_n \cap S^C).
\end{aligned}$$

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

$$\mu^*(z) = \inf\left\{\sum_{i=1}^{\infty} \mu(B_i) \colon B_1, B_2, ... \in \mathcal{S} \text{ mit } z \subseteq \bigcup_{i=1}^{\infty} B_i\right\}.$$

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

(c) Fehlt noch $\mu^*(A) = \mu(A)$ für alle $A \in \mathcal{S}$. Dann ist $\bar{\mu} := \mu^*|_{\mu(\mathcal{S})}$ das gewünschte Maß auf $\mu(\mathcal{S})$. Es gilt

„$\leq$“:

$$\mu^*(A) = \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) \colon A_1, A_2, ... \in \mathcal{S} \text{ mit } A \subseteq \bigcup_{i=1}^{\infty} A_i\right\} \leq \mu(A)\ \forall A \in \mathcal{S}$$

„$\geq$“:

$$\text{Ist } A \subseteq \bigcup_{n=1}^{\infty} A_n,\ A_1, A_2, ... \in \mathcal{S},$$

so gilt

$$\mu(A) \overset{1.3.8}{\leq} \sum_{n=1}^{\infty} \mu(A_n).$$

Folglich gilt

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

□

Satz 1.3.13. (Existenz und Eindeutigkeit von Maßen)
Ist $(\Omega, \mathcal{A})$ ein messbarer Raum, $\mathcal{E}$ ein Semiring mit $\sigma(\mathcal{E}) = \mathcal{A}$. Sei $\mu\colon \mathcal{E} \to [0, \infty]$ mit

- $\mu(\emptyset) = 0$
- μ ist σ-additiv
- es gibt Folge eine $E_1, E_2, ... \in \mathcal{E}$ 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: 1.3.12
Eindeutigkeit: 1.2.13 □

1.4 Das Beispiel – Maße aus Verteilungsfunktionen

Definition 1.4.1. $F\colon \mathbb{R} \to \mathbb{R}$ 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) $\lim_{t\to\infty} F(t) = 1$ und $\lim_{t\to-\infty} F(t) = 0$.

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

Vorlesung 6

Beweis. Um den Fortsetzungssatz zu nutzen, müssen wir zunächst einen Semiring wählen, der die Borel-σ-Algebra erzeugt. Wir wissen bereits, dass alle möglichen Varianten von Intervallen $\mathcal{B}(\mathbb{R})$ erzeugt, die meisten sind aber keine Semiringe. Wir nehmen

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

und stellen sofort fest (Eigenschaften checken), dass $\mathcal{E}$ ein Semiring mit $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$ ist. Desweiteren wählen wir $E_n := (-n,n]$ für $n \in \mathbb{N}$, so dass offenbar $E_n \uparrow \mathbb{R}$ gilt. Als Mengenfunktion auf $\mathcal{E}$ 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) $\mu(E_n) = \mu((-n,n]) = F(n) - F(-n) \leq 1$

(iii) Für die σ-Additivität seien $(a_n,b_n] \in \mathcal{E}$ mit $(a_n,b_n] \cap (a_k,b_k] = \emptyset$, $n \neq k$, und sei

$$\dot{\bigcup_{n=1}^{\infty}} (a_n,b_n] \in \mathcal{E}, \quad \text{also} \quad \dot{\bigcup_{n=1}^{\infty}} (a_n,b_n] =: (a,b],$$

für geeignete $a,b \in \mathbb{R}$. Als Anschauungsbeispiel halte man $(0,1] = \dot{\bigcup}_{n=1}^{\infty} (\frac{1}{n+1}, \frac{1}{n}]$ im Kopf. Um den Beweis besser zu verstehen, schauen wir uns erstmal den endlichen Fall an, d. h.

$$(a,b] = \dot{\bigcup_{n=1}^{N}} (a_n,b_n],$$

für ein $N \in \mathbb{N}$. Dann bekommen wir die σ-Additivität sofort:

$$\mu((a,b]) \overset{\text{Def.}}{=} F(b) - F(a) \overset{\text{Teleskop}}{=} \sum_{n=1}^{N} (F(b_n) - F(a_n)) = \sum_{n=1}^{N} \mu((a_n,b_n]),$$

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

$$F(b) - F(a) = \sum_{n=1}^{\infty} (F(b_n) - F(a_n), \tag{1.6}$$

denn das ist gerade

$$\mu\Big(\dot{\bigcup_{n=1}^{\infty}} (a_n,b_n] \Big) = \sum_{n=1}^{\infty} \mu((a_n,b_n])$$

für $(a,b] = \dot{\bigcup}_{n=1}^{\infty} (a_n,b_n]$.

Für die Gleichheit (1.6) zeigen wir beide Ungleichungen:

„$\geq$“: Weil F monoton ist, folgt

$$F(b) - F(a) \geq \sum_{n=1}^{N} (F(b_n) - F(a_n))$$

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

$$F(b) - F(a) \geq \lim_{N\to\infty} \sum_{n=1}^{N} (F(b) - F(a)) = \sum_{n=1}^{\infty} (F(b) - F(a)).$$

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

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

Die $\tilde{b}_n$ existieren, weil F rechtsstetig ist. Weil

$$(a, b] = \bigcup_{n=1}^{\infty} (a_n, \tilde{b}_n)$$

gilt

$$[a + \varepsilon, b] \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n].$$

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

$$[a + \varepsilon, b] \subseteq \bigcup_{n=1}^{N} (a_n, \tilde{b}_n).$$

Daraus folgt dann

$$\begin{aligned} F(b) - F(a + \varepsilon) &\overset{\text{F mon.}}{\leq} \sum_{n=1}^{N} (F(\tilde{b}_n) - F(a + \varepsilon)) \\ &\leq \sum_{n=1}^{\infty} (F(\tilde{b}_n) - F(a + \varepsilon)) \\ &\overset{(1.7)}{\leq} \sum_{n=1}^{\infty} (F(b) - F(a) + \frac{\varepsilon}{2^n}) = \sum_{n=1}^{\infty} (F(b) - F(a)) + \varepsilon, \end{aligned}$$

wobei wir im letzten Schritt die geometrische Reihe genutzt haben. Wegen der Rechtsstetigkeit von F folgt damit

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

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

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

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:

$$\begin{aligned}\mathbb{P}_F(\mathbb{R}) &= \mathbb{P}_F\Big(\bigcup_{n=1}^{\infty}(-n,n]\Big)\\ &\overset{\text{Stetigkeit}}{\underset{\text{von Maßen}}{=}} \lim_{n\to\infty}\mathbb{P}_F((-n,n])\\ &\overset{\text{Def.}}{=} \lim_{n\to\infty}(F(n)-F(-n))\\ &= \lim_{n\to\infty}F(n)-\lim_{n\to\infty}F(-n)=1.\end{aligned}$$

Achtung, das Argument werden wir jetzt immer wieder nutzen! Ganz ähnlich zeigen den Zusammenhang von F und $\mathbb{P}_F$ auf unendlichen Intervallen, wie im Satz behauptet:

$$\mathbb{P}_F((-\infty,t]) = \mathbb{P}_F\Big(\bigcup_{n=\lceil |t|\rceil}^{\infty}(-n,t]\Big) = \lim_{n\to\infty}\mathbb{P}_F((-n,t]) = F(t)-\lim_{n\to\infty}F(-n)=F(t),$$

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

$\square$

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

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

Satz 1.4.4. Es gibt ein eindeutiges Maß λ auf $(\mathbb{R}^d,\mathcal{B}(\mathbb{R}^d))$ mit $\lambda(Q)=$ Volumen(Q), siehe Analysis 2, 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

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

$\mathcal{S}$ ist ein Semiring. $\mu(Q):=$ Volumen(Q) $=\prod_{i=1}^{n}(b_i-a_i)$ ist eine σ-additive Mengenfunktion auf $\mathcal{S}$ (die σ-Additivität ist der einzige komplizierte Schritt). $E_n := (-n,n]\times\dots(-n,N]$ ist die benötigte Folge in $\mathcal{S}$ mit endlichem Volumen, die gegen $\mathbb{R}^d$ wächst. λ sei nun das eindeutige Maß auf $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ das μ fortsetzt. Es fehlt noch, dass λ ein unendliches Maß ist. Aber auch das geht mit den Argumenten des vorherigen Beweises:

$$\begin{aligned}\lambda(\mathbb{R}^d) &\overset{\text{Stetigkeit}}{\underset{\text{von Maßen}}{=}} \lim_{n\to\infty}\lambda((n,-n]\times ...\times(n,-n])\\ &= \lim_{n\to\infty}2n\cdot ...\cdot 2n\\ &= \lim_{n\to\infty}2^d n^d=\infty.\end{aligned}$$

$\square$

Jetzt kommen wir zu konkreten Beispielen von Verteilungsfunktionen, die uns erneut in der Stochastik begegnen werden.

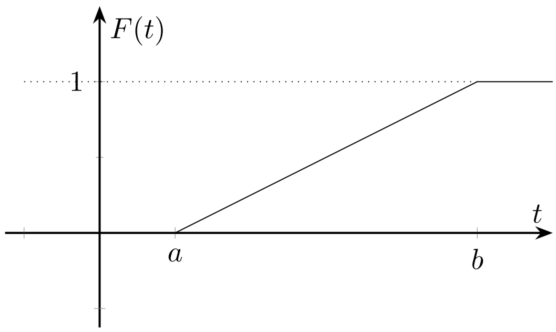
Beispiel 1.4.5. Für $a < b$ sei

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

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 $[a,b]$.



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

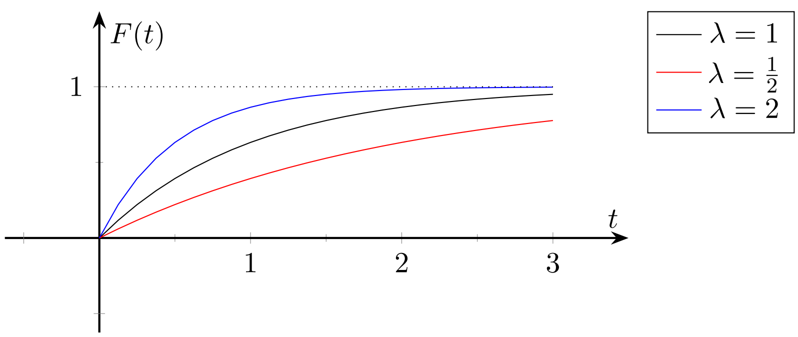
Beispiel 1.4.6. Für $\lambda > 0$ sei

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

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 $\lambda > 0$.

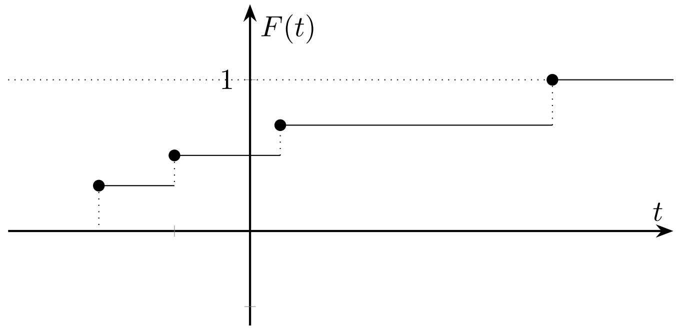


Man nennt das Maß auch Exp(λ). In der Graphik ist Exp(λ) für drei verschiedene λ geplotted.

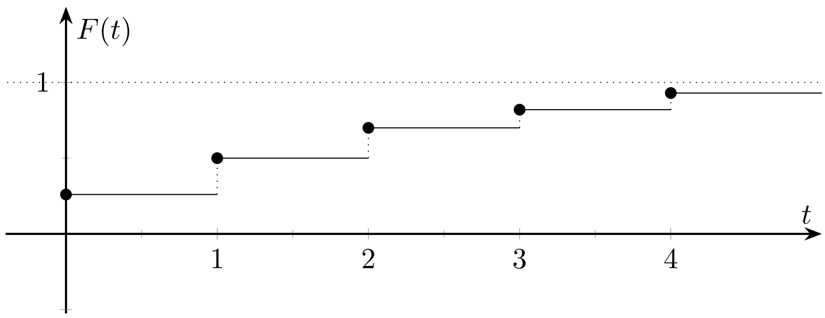
Beispiel 1.4.7. Für $a_1, ..., a_N \in \mathbb{R}$ mit $p_1, ..., p_N \geq 0$ und $\sum\limits_{k=1}^{N} p_k = 1$ ist

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

eine Verteilungsfunktion. Die zugehörigen Maße $\mathbb{P}_F$ werden **diskrete Maße** genannt weil die Menge $\{a_1, ..., a_N\}$ eine diskrete Menge ist (ohne Häufungspunkte).



Die diskrete Verteilungsfunktion funktioniert unverändert auch für $N = \infty$. 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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Beachte: Weil wir die Poissonverteilung bereits auf $\mathbb{N}$ definiert haben gibt es eine gewisse Doppeldeutigkeit, mit der Diskussion der nächsten Vorlesung wird aber klar, dass beide Maße das gleiche beschreiben. Die Poissonverteilung mit Parameter λ wird auch als **Poi**(λ) genannt.



Definition 1.4.8. Ist $F : \mathbb{R} \to [0, \infty)$, $f \in L^1(\mathbb{R})$ und $\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$, dann heißt f **Dichtefunktion** der Verteilungsfunktion

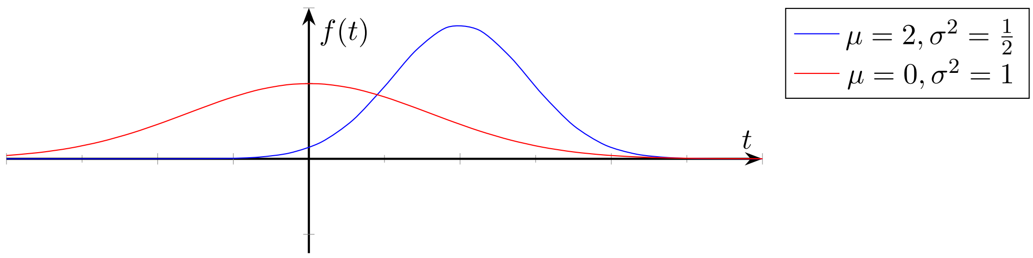
$$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x, \quad t \in \mathbb{R}. \tag{1.8}$$

Ist umgekehrt F von der Form (1.8), so heißt f **Dichte** von F. In der großen Übung wurde diskutiert, warum solch ein F die vier Eigenschaften einer Verteilungsfunktion erfüllt.

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

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

Verteilungsfunktion der **(standard) Normalverteilung**. Das Maß $\mathbb{P}_F$ nennt man dann auch (standard) normalverteilt und schreibt $\mathcal{N}(0,1)$. 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 $\int_{-\infty}^{t} e^{-x^2/2} dx$!

Vorlesung 7

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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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 1.4.10. Wir starten mit einer nicht sehr rigorosen aber dennoch hilfreichen Interpretation:

„F beschreibt, wie 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: Zunächst berechnen wir die Masse einer Einpunktmenge $\{t\}$ aus den bekannten Eigenschaften. 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:

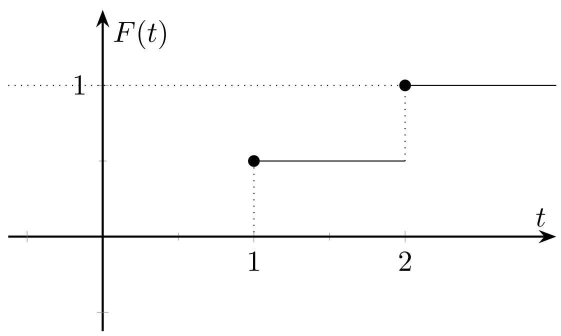
$$\begin{aligned}\mathbb{P}_F(\{t\}) &= \mathbb{P}_F\Big(\bigcap_{n=1}^{\infty}(t-\frac{1}{n},t]\Big)\\ &\overset{\text{Stetigkeit}}{\underset{\text{von Maßen}}{=}} \lim_{n\to\infty}\mathbb{P}_F\Big(\Big(t-\frac{1}{n},t\Big]\Big)\\ &\overset{\text{Def. }\mathbb{P}_F}{=} \lim_{n\to\infty}\Big(F(t)-F\Big(t-\frac{1}{n}\Big)\Big)\\ &= F(t)-\lim_{n\to\infty}F\Big(t-\frac{1}{n}\Big)=F(t)-F(t-),\end{aligned}$$

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. Insbesondere haben alle einpunktigen Mengen keine Masse, wenn F eine stetige Funktion ist (z. B. $U[a,b]$, $\operatorname{Exp}(\lambda)$, $\mathcal{N}(\mu,\sigma^2)$). Klingt komisch, oder? Ist es aber nicht. Hier sehen wir, warum Maße erst auf überabählbaren Mengen wirklich spannend werden:

$$\mathbb{P}_F((a,b]) = \mathbb{P}_F\Big(\dot{\bigcup_{x\in(a,b]}}\{x\}\Big) \neq \sum_{x\in(a,b]}\mathbb{P}_F(\{x\}),$$

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

F **konstant:** Überlegen wir nun, was es für $\mathbb{P}_F$ bedeutet, wenn F 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.

Wenn wir die Beobachtung auf $\operatorname{Poi}(\lambda)$ aus Beispiel 1.4.7 anwenden, so sehen wir, dass das zugehögige Maß $\mathbb{P}_F$ nur Masse auf $\mathbb{N}$ hat. Damit kann man das $\operatorname{Poi}(\lambda)$-verteilte Maß auf $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ mit der Definition aus Beispiel 1.1.11 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, wo F am stark 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 a nah an b, so bedeutet das natürlich, dass F dort stark wächst. Schauen wir uns wieder ein passendes Beispiel an, die Exponentialverteilung $\operatorname{Exp}(\lambda)$ für verschiedene $\lambda > 0$. Am Bildchen in Beispiel 1.4.6 ist zu erkennen, dass viel Masse nah bei der 0 liegt, wenn λ groß ist, die Verteilungsfunktion bei 0 möglichst 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.

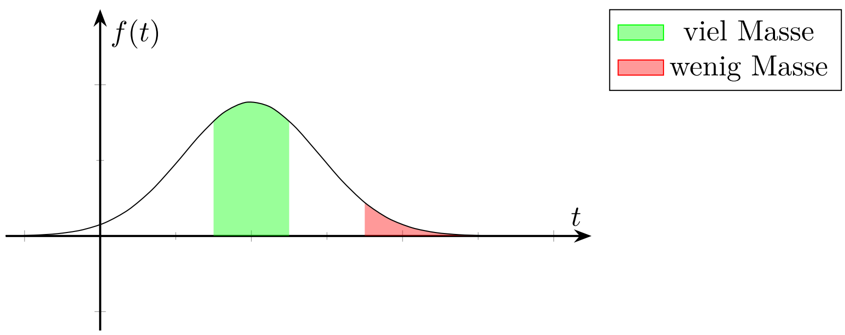
Der Fall mit Dichten: Die obige Diskussion können wir für Verteilungsfunktionen mit Dichten noch konkretisieren. Sei dazu F eine Verteilungsfunktion mit Dichte f, also $F(t) = \int\limits_{-\infty}^{t} f(x)\mathrm{d}x$. Weil F stetig ist, haben alle einpunktigen Mengen keine Masse. Aber wie können wir an f direkt sehen, wo die Masse verteilt ist? Ist f stetig, so folgt aus dem Hauptsatz der Analysis, dass $F'(t) = f(t)$ 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."

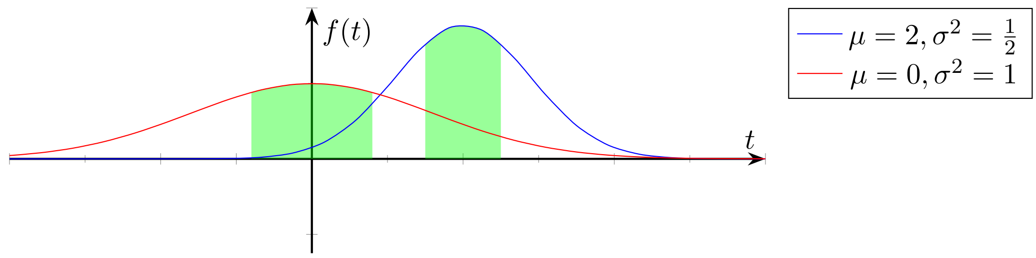
Die nützlichste Interpretation ist durch den Flächeninhalt zwischen Graphen von f und der x-Achse. Weil

$$\mathbb{P}_F((a,b]) = F(b) - F(a) = \int\limits_{-\infty}^{b} f(x)\mathrm{d}x - \int\limits_{-\infty}^{a} f(x)\mathrm{d}x = \int\limits_{a}^{b} f(x)\mathrm{d}x,$$

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 $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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.

Abschließend noch ein paar Worte zur stochastischen Modellierung von Zufall in überabzählbaren Mengen. Warum haben wir so exzessiv Maße auf der Borel-σ-Algebra diskutiert?

Diskussion 1.4.11. (Stochastische Modellierung, Nr. 2)
Das Modellieren von endlich vielen Möglichkeiten ist leicht, siehe Diskussion 1.1.7. 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 dann 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 Überabzählbarkeit automatisch zu $\mathbb{P}(\{t\}) = 0$ für alle $t \in [0,1]$. 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 $\mathbb{P}((a,b]) = b - a = F(b) - F(a)$, wobei F die Verteilungsfunktion aus Beispiel 1.4.5 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, dass den Intervallen die geforderten Wahrscheinlichkeiten gibt. Aufgrund des Fortsetzungssatzes gibt es so ein Maß, das ist gerade $\mathrm{U}([0,1])$.
Diese Motivation zur Modellierung mit der Borel-σ-Algebra und dem Fortsetzungssatz setzt nicht zwingend gleichverteilten Zufall voraus. Wenn wir also ein zufälliges Experiment mit reellen Beobachtungen modellieren wollen, sollten wir die Wahrscheinlichkeiten von „Ausgang des Experiments ist kleiner oder gleich t“ kennen. Damit definieren wir die Verteilungsfunktion $F(t)$ und wenn diese rechtsstetig ist, gibt uns die Maßtheorie das mathematische Modell $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_F)$ für das zufällige Experiment.

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) $\mathcal{P}(\mathbb{R})$ 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 $\mathcal{B}(\mathbb{R}) \subsetneq \mathcal{P}(\mathbb{R})$, elementare Beispiele für nicht Borel-messbare gibt es leider nicht.

(ii) $\mathcal{B}(\mathbb{R})$ funktioniert wunderbar! Insbesondere weil wir sehr handliche Erzeuger haben (z. B. verschiedene Arten von Intervallen) und deshalb aufgrund der bewiesenen Theoreme nur mit Intervallen arbeiten müssen.

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!

Kapitel 2

Abbildungen zwischen messbaren Räumen

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:

Objekte	Funktoren
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 $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ messbare Räume und $f : \Omega \to \Omega'$. 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

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

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:

Satz 2.1.4. Ist $\mathcal{E}'$ ein Erzeuger von $\mathcal{A}'$ und $f\colon \Omega \to \Omega'$. 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.

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

„$\Leftarrow$“: Sei

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

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 auch $\mathcal{A}' = \mathcal{F}'$ gilt. Wir überprüfen die definierenden Eigenschaften und können auf elementare Eigenschaften des Urbildes von Abbildungen in Analysis 1 zurückgreifen:

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

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

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

(iii) Sind $A_1, A_2, ... \in \mathcal{A}$, so gilt

$$f^{-1}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \bigcup_{n=1}^{\infty} f^{-1}(A_n) \in \mathcal{A}.$$

□

Vorlesung 8

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

Beispiel 2.1.6.

- Jede stetige Abbildung $f \colon \mathbb{R} \to \mathbb{R}$ ist auch Borel-messbar. Warum?
  - (i) Urbilder offener Mengen sind offen
  - (ii) $\sigma(\{O \colon O \in \mathbb{R} \text{ offen}\}) = (\mathbb{R})$
  - (iii) Satz 2.1.4
- Indikatorfunktionen

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

sind messbar genau dann wenn A messbar ist weil wir alle Urbilder direkt hinschreiben können:

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

Das selbe funktioniert auch für Indikatorfunktionen $\mathbf{1}_A : \Omega \to \mathbb{R}$, das ist eine kleine Übungensaufgabe.

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

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

Wegen Satz 2.1.4 ist deshalb $f \colon \mathbb{R} \to \mathbb{R}$ Borel-messbar genau dann, wenn

$$f^{-1}((-\infty, t]) = \{x \colon f(x) \leq t\} =: \{f \leq t\} \in \mathcal{B}(\mathbb{R})$$

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

$$f^{-1}((-\infty, t)) = \{x \colon f(x) < t\} =: \{f < t\} \in \mathcal{B}(\mathbb{R})$$

für alle $t \in \mathbb{R}$ oder

$$f^{-1}((a, b)) = \{x \colon f(x) \in (a, b)\} =: \{f \in (a, b)\} \in \mathcal{B}(\mathbb{R}).$$

Definition 2.1.8. Sei $f \colon \Omega \to \Omega'$ für einen messbaren Raum $(\Omega', \mathcal{A}')$. Dann ist

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

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

Beweis. Übung. □

Beispiel 2.1.9.

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

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

$$\sigma(f_i,\ i \in I) := \sigma\Big(\bigcup_{i\in I} \sigma(f_i)\Big) = \sigma(\{f_i^{-1}(A_i)\colon A_i \in \mathcal{A}_i,\ i \in I\}).$$

2.2 Bildmaße oder „push-forward“ eines Maßes

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

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

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

ein Maß auf $(\Omega', \mathcal{A}')$. Dieses Maß heißt „Bildmaß“ oder „push-forward“.

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

(i) $\mu_f(\emptyset) = \mu\left(f^{-1}(B)\right) = \mu(\emptyset) = 0$

(ii) Seien $A_1, A_2, ... \in \mathcal{A}'$ paarweise disjunkt, dann folgt aus der Definition und den Maßeigenschaften von μ

$$\begin{aligned}\mu_f\Big(\dot{\bigcup_{n=1}^{\infty}} A_n\Big) &\overset{\text{Def.}}{=} \mu\Big(f^{-1}\Big(\dot{\bigcup_{n=1}^{\infty}} A_n\Big)\Big)\\ &\overset{\text{Urbild}}{=} \mu\Big(\dot{\bigcup_{n=1}^{\infty}} f^{-1}(A_n)\Big)\\ &\overset{\mu \text{ Maß}}{=} \sum_{n=1}^{\infty} \mu\Big(f^{-1}(A_n)\Big) \overset{\text{Def.}}{=} \sum_{n=1}^{\infty} \mu_f(A_n).\end{aligned}$$

Damit ist μ_f auch σ-additiv.

□

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

Behauptung: $\mu_f = \lambda$. Berechnen wir dazu μ_f auf einem $\cap$-stabilen Erzeuger von $\mathcal{B}(\mathbb{R})$:

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

Weil $\mathcal{E} = \{(a,b]\colon a,b \in \mathbb{R}\}$ $\cap$-stabil ist mit $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$, gilt aufgrund von Folgerung 1.2.13 auch $\lambda = \mu_f$. In anderen Worten: Es gilt

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

für alle $x \in \mathbb{R}$, wobei $B + x := \{b + x : b \in B\}$. Man sagt, dass Lebesgue-Maß **translationsinvariant**, Verschieben von Mengen ändert das Lebesguemaß nicht.

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 $a_n \to +\infty$, $n \to \infty$, und $a_n \to -\infty$, $n \to \infty$, wie in Analysis 1 definiert. Oft schreiben wir ∞ statt $+\infty$. Auf $\overline{\mathbb{R}}$ definieren wir auch eine 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 2.3.1. Sei $(\Omega, \mathcal{A})$ ein messbarer Raum, $f\colon \Omega \to \overline{\mathbb{R}}$ heißt **messbare numerische Funktion**, falls f $(\Omega, \mathcal{B}(\overline{\mathbb{R}}))$-messbar ist.

Bemerkung 2.3.2.

(i) Jede Borel-messbare Funktion $f : \Omega \to \mathbb{R}$ ist auch eine messbare numerische Funktion, denn $f^{-1}(x \cup B) = f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ für $x \in \{\{+\infty\}, \{-\infty\}, \{+\infty, -\infty\}\}$.

(ii) Aussagen für messbare reelle Abbildungen gelten ganz analog für messbare numerische Abbildungen. So gilt etwa: $f\colon \Omega \to \overline{\mathbb{R}}$ 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 Satz 2.1.4 weil $\mathcal{E} = \{[-\infty, t]\colon t \in \overline{\mathbb{R}}\}$ die σ-Algebra $\mathcal{B}(\mathbb{R})$ erzeugt.

Definition 2.3.3. Für $a, b \in \overline{\mathbb{R}}$ definieren wir

$$\begin{aligned} a \wedge b &:= \min\{a, b\}, \\ a \vee b &:= \max\{a, b\}, \\ a^+ &:= \max\{0, a\}, \\ a^- &:= -\min\{0, a\}. \end{aligned}$$

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

Es gelten direkt aus der Definition folgende wichtige Gleichungen:

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

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

Vorlesung 9

Lemma 2.3.4. Seien f, g messbar, dann sind auch $f+g$, αf, $f \cdot g$, $f \wedge g$, $f \vee g$, $|f|$ messbar.

Beweis. Übung. □

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

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

messbar.

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

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

Der wesentliche Trick war natürlich die erste Gleichheit. Mit genau diesem Trick konnte man im vorherigen Lemma zeigen, dass $f+g$ messbar ist weil $\{f+g<t\} = \{f < t-g\}$ gilt. Genauso zeigt man auch $\{f \leq g\} \in \mathcal{A}$. Weil $\{f = g\} = \{f \leq g\} \cap \{g \leq f\}$ und $\{f \neq g\} = \{f = g\}^C$, sind auch die letzten beiden Mengen in $\mathcal{A}$. □

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

Proposition 2.3.6. Sei $f_1, f_2, ... : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ eine Folge messbarer numerischer Funktionen.

(i) Dann sind auch die punktweise definierten Funktionen

- $f_1(\omega) := \inf_{n \in \mathbb{N}} f_n(\omega)$,
- $f_2(\omega) := \sup_{n \in \mathbb{N}} f_n(\omega)$,
- $f_3(\omega) := \limsup_{n \to \infty} f_n(\omega)$,
- $f_4(\omega) := \liminf_{n \to \infty} f_n(\omega)$

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

$$f(\omega) := \lim_{n \to \infty} f_n(\omega)$$

messbar.

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

$$\{f_1 \leq t\} = \{\omega \in \Omega : \inf_{n\in\mathbb{N}} f_n(\omega) \leq t\} = \underbrace{\bigcup_{n\in\mathbb{N}} \underbrace{\{\omega : f_n(\omega) \leq t\}}_{\in\mathcal{A}}}_{\in\mathcal{A}}.$$

□

Kapitel 3

Integrationstheorie

Wir haben in Analysis 2 schon die Lebesgue-Integration im $\mathbb{R}^d$ kennengelernt. Das war damals ziemlich kompliziert, weil wir konsequent den Begriff und Eigenschaften von Maßen vermieden haben. Als Spezialfall der nun folgenden Integrationstheorie für die Borel-σ-Algebra mit dem Lebesgue-Maß $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$ ist die Theorie hier in der Stochastik 1 viel einfacher. Das liegt aber daran, dass wir schon viel Mühe in die Maßtheorie gesteckt hatten. Weil wir tatsächlich gar keine Eigenschaften des $\mathbb{R}^d$ oder des Lebesguemaß nutzen müssen, ist die folgende allgemeine Integrationstheorie auf messbaren Räumen kein bisschen komplizierter.

3.1 Das (allgemeine) Lebesgue-Integral

In diesem Abschnitt werden wir Integrale der Form $\int_\Omega f \mathrm{d}\mu$ für beliebige Maßräume $(\Omega, \mathcal{A}, \mu)$ definieren. Es sei jetzt immer $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Das Maß kann endlich sein (z.B. ein Wahrscheinlichkeitsmaß) oder unendlich sein (z.B. das Lebesguemaß). Um leichter zu folgen, kann man sich einfach den Spezialfall $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ zur Veranschaulichung vorstellen.

Definition 3.1.1. Eine messbare Abbildung $f : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ heißt **einfach** (alternativ **elementar** oder **Treppenfunktion**), falls f nur endlich viele Werte annimmt, d. h. f hat die Form

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

mit $\alpha_1, ..., \alpha_n \in \overline{\mathbb{R}}$ und

$$A_k = \alpha_k = \{f = \alpha_k\} = \{\omega \colon f(\omega)\} = f^{-1}([\alpha_k, \alpha_k]).$$

Ganz wichtig in der Theorie ist, dass alle A_k messbar sind weil f messbar ist. Weil die Mengen A_k paarweise disjunkt sind, nennt man eine Darstellung der Form (3.1) auch disjunkte Darstellung von f. Wir definieren auch noch

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

Wie in Analysis 2 sind disjunkte Darstellungen messbarer Funktionen nicht eindeutig, z. B. gilt

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

Wir definieren nun ähnlich wie in Analysis 2 das Integral nicht-negativer einfacher Funktionen, dann durch Approximation das Integral nicht-negativer messbarer Funktionen und dann durch die Zerlegen $f = f^+ - f^-$ das Integral beliebiger messbarer Funktionen.

(A) Integrale nicht-negativer einfacher Funktionen

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

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

wenn f die disjunkte Darstellung

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

hat. Weil $\alpha_k = +\infty$ sowie $\mu(A_k) = +\infty$ möglich ist, muss definiert sein, wie $+$ und $\cdot$ mit ∞ geht, siehe die Definition in Abschnitt 1.1.

Beispiel. Weil das Integral den „Flächeninhalt“ zwischen Graphen und Achse beschreiben soll, sind die Rechenregeln $+$ und $\cdot$ mit ∞ sinnvoll definiert worden:

$$\int_\mathbb{R} 0 \cdot \mathbf{1}_\mathbb{R} \,\mathrm{d}\lambda = 0 \cdot \lambda(\mathbf{1}_\mathbb{R}) = 0 \cdot (+\infty) = 0,$$

$$\int_\mathbb{R} \mathbf{1}_\mathbb{R} \,\mathrm{d}\lambda = 1 \cdot \lambda(\mathbf{1}_\mathbb{R}) = 1 \cdot (+\infty) = +\infty,$$

$$\int_\mathbb{R} (+\infty) \cdot \mathbf{1}_{[a,b]} \,\mathrm{d}\lambda = +\infty \cdot \lambda([a,b]) = +\infty \cdot (b-a) = +\infty,$$

$$\int_\mathbb{R} 3 \cdot \mathbf{1}_{[0,1]} \,\mathrm{d}\lambda = 3 \cdot \lambda([0,1]) = 3.$$

Was sollen die Integrale auch sonst sein?

Rechnen wir wie in Analysis 2 (vergleiche dort mit 12.1.9) noch nach, dass das Integral einer nicht-negativen einfachen Funktion nicht von der disjunkten Darstellung abhängt:

Lemma 3.1.3. Sei

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

mit $\alpha_k, \beta_l \geq 0$ und paarweise disjunkten A_k bzw. B_k, 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 $\bigcup_{k=1}^n A_k = \{f > 0\} = \bigcup_{l=1}^m B_k$ und der σ-Additivität von μ gilt

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

$(\star)$ gilt weil entweder $\mu(A_k \cap B_l) = 0$ oder $\mu(A_k \cap B_l) > 0$ gilt. Im ersten Fall gilt $\alpha_k \mu(A_k \cap B_l) = \beta_l \mu(A_k \cap B_l) = 0$ trivialerweise, im zweiten Fall impliziert $\mu(A_k \cap B_l) > 0$ schon $A_k \cap B_l \neq 0$ 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) $\mathbf{1}_A \in \mathcal{E}^+$

(ii) $\alpha f \in \mathcal{E}^+$ und $\int_\Omega \alpha f \, \mathrm{d}\mu = \alpha \int_\Omega f \, \mathrm{d}\mu$

(iii) $f + g \in \mathcal{E}^+$ und $\int_\Omega (f+g) \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu$

(iv) $f \leq g \Rightarrow \int_\Omega f \, \mathrm{d}\mu \leq \int_\Omega g \, \mathrm{d}\mu$.

Beweis.

(i) ✓

(ii) αf ist wegen

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

auch eine nicht-negative einfache Funktion. Das Integral berechnet sich aus der Definition:

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

(iii) Seien

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

und seien ohne Einschränkung $\bigcup\limits_{k=1}^n A_k = \Omega = \bigcup\limits_{l=1}^m B_k$. 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 mit $1 = \mathbf{1}_\Omega = \sum_{k=1}^n \mathbf{1}_{A_k} = \sum_{l=1}^m \mathbf{1}_{B_k}$ und $\mathbf{1}_{A_k} \cdot \mathbf{1}_{B_l} = \mathbf{1}_{A_k \cap B_l}$ auch

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

Also gilt

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

und damit die Behauptung.

(iv) Monotonie ✓, folgt direkt aus der Definition.

□

(B) Integral nichtnegativer messbarer numerischer Funktionen

Die Definition des Integrals nicht-negativer Funktionen läuft anders als in Analysis 2. 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 messbare $f\colon(\Omega,\mathcal{A})\to(\overline{\mathbb{R}},\mathcal{B}(\overline{\mathbb{R}}))$ mit $f\geq 0$ definieren wir

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

Warnung: Wie bei nicht-negativen einfachen Funktionen ist $\int_\Omega f\,\mathrm{d}\mu=+\infty$ ausdrücklich erlaubt!

Jetzt wollen wir diese komplizierte Definition (wie soll man damit irgendwas zeigen?) durch eine handlichere äquivalente Darstellung ersetzen. Dazu kommt zunächst eine extrem wichtige Proposition, die man unbedingt mit dem Beweis von Satz 12.3.3 in der Analysis 2 vergleichen sollte!

Proposition 3.1.6. (Darum sind messbare Funktionen so wichtig)
Für jede nichtnegative 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 $n\to\infty$.

Beweis. Wir definieren

$$f_n=\underbrace{\sum_{k=1}^{n\cdot 2^n}\frac{k}{2^n}\mathbf{1}_{\underbrace{A_k}_{\in\mathcal{A}}}+n\mathbf{1}_{\underbrace{f^{-1}((n,+\infty])}_{\in\mathcal{A}}}}_{\text{messbar}}$$

mit

$$A_k=f^{-1}\Big(\Big(\frac{k}{2^n},\frac{k+1}{2^n}\Big]\Big).$$

Fürs bessere Verständniss zeichne man die Folge f_n für das Beispiel $f : \mathbb{R} \to \overline{\mathbb{R}}$ mit $f(x) = +\infty \cdot \mathbf{1}_{(-\infty,0]} + \frac{1}{x} \cdot \mathbf{1}_{(0,+\infty)}$ hin! Die Messbarkeit von f_n folgt aus Lemma 2.3.4 und dem Fakt, dass Indikatorfunktionen $\mathbf{1}_A$ genau für messbare A messbar sind. 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.

$\square$

Vorlesung 10

Lemma 3.1.7. (Montone Konvergenz Theorem (MCT) für einfache Funktionen)
Sei $(f_n) \subseteq \mathcal{E}^+$ mit $f_n \uparrow f$, $n \to \infty$, für eine nichtnegative messbare numerische Funktion f. Dann gilt

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

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

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

„$\leq$": Folgt direkt aus der Definition

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

weil das Supremum einer Menge eine obere Schranke der Menge ist.

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

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \geq \int_\Omega g \,\mathrm{d}\mu. \tag{3.2}$$

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

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

Weil aufgrund der Definition der Mengen

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

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

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

Wegen Stetigkeit von Maßen und gilt

$$\lim_{n\to\infty} \mu(A_n \cap C_k) = \mu\Big(\bigcup_{n=1}^{\infty} (A_n \cap C_k)\Big) = \mu\Big(\underbrace{\Big(\bigcup_{n=1}^{\infty} A_n\Big)}_{=\Omega} \cap C_k\Big) = \mu(C_k).$$

Damit gilt

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

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

□

Warum war das Lemma so wichtig? Die Definition des Integrals als Supremum ist sehr unhandlich. Es hat natürlich den Vorteil, dass das Integral sofort wohldefiniert ist (wir brauchen keine Unabhängigkeit von der approximierenden Folge wie in Analysis 2), 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 $f, g\colon \Omega \to [0,\infty]$ messbar und $\alpha \geq 0$ gelten:

(i)
$$\int_\Omega \alpha f \,\mathrm{d}\mu = \alpha \int_\Omega f \,\mathrm{d}\mu,$$

(ii)
$$\int_\Omega (f+g) \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu,$$

(iii)
$$f \leq g \Rightarrow \int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega g \,\mathrm{d}\mu.$$

Beweis. Wir zeigen nur (ii), die anderen Aussagen gehen analog.

Seien $(f_n), (g_n) \subseteq \mathcal{E}^+$ mit $f_n \uparrow f$, $g_n \uparrow g$, $n \to \infty$. 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

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

□

(C) Integral messbarer numerischer Funktionen

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

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

Definition 3.1.9. f heißt μ-integrierbar, falls

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

In diesem Fall definieren wir

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu.$$

Zur Notation: Man schreibt statt

$$\int_\Omega f \,\mathrm{d}\mu$$

auch

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

Warnung: Im Gegensatz zu nichtnegativen messbare Funktionen, für die wir das Integral $+\infty$ erlauben, ist für beliebige messbare Funktionen das Integral nur definiert, wenn es endlich ist. Damit vermeiden wir definieren zu müssen, was $\infty - \infty$ bedeuten sollte.

Definition 3.1.10. Ist $f\colon \Omega \to \overline{\mathbb{R}}$ und $A \in \mathcal{A}$, so definiert man

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

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

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

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. Sind $f, g\colon \Omega \to \overline{\mathbb{R}}$ μ-integrierbar, $\alpha \in \mathbb{R}$, dann gilt

(i) αf ist μ-integrierbar und
$$\int_\Omega \alpha f \,\mathrm{d}\mu = \alpha \int_\Omega f \,\mathrm{d}\mu$$

(ii) Wenn $f + g$ sinnvoll definiert ist (d. h. kein $+\infty + (-\infty)$), so ist $f + g$ μ-integrierbar und
$$\int_\Omega (f + g) \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu$$

(iii)
$$f \leq g \Rightarrow \int_\Omega f \mathrm{d}\mu \leq \int_\Omega g \mathrm{d}\mu$$

(iv) $\triangle$-Ungleichung:

$$\Big|\int_\Omega f \,\mathrm{d}\mu\Big| \leq \int_\Omega |f| \,\mathrm{d}\mu.$$

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

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

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

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

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

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

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

Damit gelten

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

und

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

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

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

Umformen ergibt

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

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

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

Erneutes Auflösen ergibt

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

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

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

(iii) Natürlich gilt $f \leq g \Leftrightarrow g-f \geq 0$. Weil die Nullfunktion sowie $g-f$ nichtnegativ sind, gilt wegen Lemma 3.1.8 und der Definition des Integrals für einfache Funktionen (die Nullfunktion)

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

Umformen gibt die Behauptung.

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

$$\begin{aligned}\Big|\int_\Omega f\,\mathrm{d}\mu\Big| &\overset{\text{Def.}}{=} \Big|\int_\Omega f^+\,\mathrm{d}\mu - \int_\Omega f^-\,\mathrm{d}\mu\Big| \\ &\overset{\triangle}{\leq} \Big|\int_\Omega f^+\,\mathrm{d}\mu\Big| + \Big|\int_\Omega f^-\,\mathrm{d}\mu\Big| \\ &\overset{\geq 0}{=} \int_\Omega f^+\,\mathrm{d}\mu + \int_\Omega f^-\,\mathrm{d}\mu \\ &\overset{3.1.8}{=} \int_\Omega (f^+ + f^-)\,\mathrm{d}\mu = \int_\Omega |f|\,\mathrm{d}\mu.\end{aligned}$$

□

Beispiel 3.1.12. (i) Ist $(\Omega, \mathcal{A}, \mu) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, so ist das neue Integral gerade das Lebesgue-Integral aus Analysis 2. Insbesondere lassen sich alle Integrale mit den Rechenregeln aus Analysis 2 berechnen.

(ii) $(\Omega, \mathcal{A}, \mu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$, $\mu(A) = \#A$ (Zählmaß), dann ist

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

Das wird in der großen Übung besprochen und nach dem Satz über monotone Konvergenz aufgeschrieben. Allgemeine Lebesgue Integrale verallgemeinern also auch das Konzept der Reihen!

(iii) Ist $\mathbb{P}_F$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F, so heißt

$$\int_\mathbb{R} g\,\mathrm{d}F := \int_\mathbb{R} g\,\mathrm{d}\mathbb{P}_F$$

Lebesgue-Stieltjes-Integral. Damit werden wir uns später ganz ausführlich befassen, sobald wir über Erwartungswerte von Zufallsvariablen sprechen.

Wie in Analysis 2 werden Nullmengen keine Rolle bei Integralen spielen. Starten wir mit der Definition, die (aufgrund der nun bekannten Maßtheorie) hier viel einfacher ist.

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 ein Eigenschaft für alle $\omega \in \Omega$ außer einer μ-Nullmenge, so gilt die Eigenschaft μ-fast überall. Man schreibt auch μ-f.ü.

(ii) Ist μ ein Wahrscheinlichkeitsmaß, so sagt man anstelle von „μ-fast überall“ auch „μ-fast sicher“ oder μ-f.s.

Satz 3.1.15. Seien $f, g\colon \Omega \to \overline{\mathbb{R}}$ μ-integrierbar, dann gelten:

(i) f ist μ-fast überall endlich.

(ii) $f = g$ μ-fast überall impliziert

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega g \, \mathrm{d}\mu.$$

(iii) $f \geq 0$ und $\int_\Omega f \, \mathrm{d}\mu = 0$ impliziert $f = 0$ μ-fast überall.

Vorlesung 11

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

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

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) Weil aus $f = g$ fast überall auch $f^+ = g^+$ und $f^- = g^-$ fast überall folgt, impliziert die Definition des Integrals als

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega f^+ \,\mathrm{d}\mu - \int_\Omega f^- \,\mathrm{d}\mu$$

bzw.

$$\int_\Omega g \,\mathrm{d}\mu = \int_\Omega g^+ \,\mathrm{d}\mu - \int_\Omega g^- \,\mathrm{d}\mu,$$

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

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

die Nullmenge auf der f und g nicht übereinstimmen. Wir definieren die Indikatorfunktion

$$h(\omega) = (+\infty) \cdot \mathbf{1}_N(\omega) = \begin{cases} +\infty & : \omega \in N \\ 0 & : \omega \notin N \end{cases},$$

die nur die Werte 0 und $+\infty$ annimmt. Damit gilt

$$\int_\Omega h \,\mathrm{d}\mu \overset{\text{Def. Int.}}{=} (+\infty) \cdot \mu(N) \overset{\text{Ann.}}{=} (+\infty) \cdot 0 \overset{\text{Def.}}{=} 0.$$

Es gilt aufgrund der Definition von h, dass $f \leq g + h$. Das sieht merkwürdigt aus, liegt aber an folgender Fallunterscheidung: Ist $\omega \notin N$, so gilt $f(\omega) = g(\omega)$ und für $\omega \in N$ gilt aufgrund der Definition von h

$$f(\omega) \leq +\infty = g(\omega) + h(\omega).$$

Die Monotonie vom Integral impliziert nun

$$\int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega (g + h) \,\mathrm{d}\mu \overset{\text{Lin.}}{=} \int_\Omega g \,\mathrm{d}\mu + \int_\Omega h \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu.$$

Mit exakt dem selben Argument, vertausche die Rollen von f und g, folgt auch

$$\int_\Omega g \,\mathrm{d}\mu \leq \int_\Omega f \,\mathrm{d}\mu.$$

Beide Ungleichungen zusammen implizieren die Behauptung.

(iii) Sei $A_n = \{f \geq \frac{1}{n}\} = \{\omega\colon f(\omega) \geq \frac{1}{n}\}$. Damit ist

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

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

$$0 \leq \mu(\{\omega\colon f(\omega) > 0\}) = \mu\Big(\bigcup_{n=1}^\infty A_n\Big) \overset{\text{subadd.}}{\leq} \sum_{n=1}^\infty \mu(A_n) = 0.$$

Es gilt also $f = 0$ μ-fast überall.

□

Für spätere Verwendungen noch ein Satz zur Transformation von Integralen:

Satz 3.1.16. (abstrakter Transformationssatz)
Seien $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ messbare Räume, μ ein Maß auf $\mathcal{A}$, $f\colon \Omega \to \Omega'$ messbar, $g\colon \Omega' \to \overline{\mathbb{R}}$ messbar und $g \geq 0$. Dann gilt

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

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

$$\begin{tikzcd}
(\Omega,\mathcal{A},\mu) \arrow[r, "f"'] \arrow[dr, "g\circ f", "\int_\Omega g\circ f\,\mathrm{d}\mu"'] & (\Omega',\mathcal{A}',\mu_f) \arrow[d, "g"', "\int_{\Omega'} g\,\mathrm{d}\mu_f"] \\
& (\overline{\mathbb{R}},\mathcal{B}(\overline{\mathbb{R}}))
\end{tikzcd}$$

Beweis. Einmal durch die Gebetsmühle der Integrationstheorie:

(A) Ist

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

eine nichtnegative einfache Funktion, so ist auch

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

auch einfach. Weil nach Definition des push-forwards $\mu_f(A_k) = \mu(f^{-1}(A_k))$ gilt, bekommen wir

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

Damit gilt die Behauptung für einfache Funktionen.

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

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

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

□

Korollar 3.1.17. Unter den Voraussetzungen von 3.1.16 gelte jetzt noch $g\colon \Omega' \to \overline{\mathbb{R}}$. 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

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

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

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

Nun zur Berechnung der Integrale:

$$\begin{aligned} \int_{\Omega'} g \circ f \,\mathrm{d}\mu &\overset{\text{Def.}}{=} \int_{\Omega'} g^+ \,\mathrm{d}\mu_f - \int_{\Omega'} g^- \,\mathrm{d}\mu_f \\ &\overset{3.1.17}{=} \int_{\Omega} g^+ \circ f \,\mathrm{d}\mu - \int_{\Omega} g^- \circ f \,\mathrm{d}\mu \\ &= \int_{\Omega} (g \circ f)^+ \,\mathrm{d}\mu - \int_{\Omega} (g \circ f)^- \,\mathrm{d}\mu \\ &\overset{\text{Def.}}{=} \int_{\Omega} g \circ f \,\mathrm{d}\mu. \end{aligned}$$

□

3.2 Konvergenzsätze

Sei immer $(\Omega, \mathcal{A}, \mu)$ ein fester Maßraum. Gezeigt haben wir schon

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

wenn $(f_n)_{n\in\mathbb{N}}$ eine Folge nichtnegativer <u>einfacher</u> Funktionen ist, die wachsend gegen f konvergieren. Wir wollen nun die gleiche Aussage für beliebige nichtnegative wachsende Folgen zeigen.

Satz 3.2.1. (Monotone Konvergenz Theorem (MCT))
Seien $f, f_1, f_2, ...\colon \Omega \to \overline{\mathbb{R}}$ messbar und es gelte $f_1 \leq f_2 \leq ... \leq f$ sowie $f = \lim\limits_{n\to\infty} f_n$ μ-f.ü. Dann gilt

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

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

Beweis.

(i) Wir nehmen an, dass $f_1 \leq f_2 \leq ... \leq f$ und $f = \lim\limits_{n\to\infty} f_n$ nicht nur fast überall, sondern für alle $\omega \in \Omega$ gelten.

„≥“: Wegen der Monotonie des Integrals gilt

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

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

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

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

(a) $h_n \leq f_n$

(b) $h_n \uparrow f$ punktweise

Zu (a): Es gilt $g_{i,n} \leq f_n$ 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 $i = 1, ..., n$ gilt, ist auch

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

für alle festen $i \in \mathbb{N}$. Weil aber $\lim\limits_{i\to\infty} f_i = f$, ist $\lim\limits_{n\to\infty} h_n \geq f$, 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 $\lim\limits_{n\to\infty} h_n = f$ punktweise. Die Folge (h_n) ist also eine wachsende Folge von einfachen Funktionen, die zwischen (f_n) und f liegt und punktweise gegen f konvergiert. Damit bekommen wir aus dem monotone Konvergenz Theorem für einfache Funktionen durch Einschachtelung die Behauptung:

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \underset{\text{Mon.}}{\overset{(a)}{\geq}} \lim_{n\to\infty} \int_\Omega h_n \,\mathrm{d}\mu \overset{3.1.7}{=} \int_\Omega \lim_{n\to\infty} h_n \,\mathrm{d}\mu \overset{(b)}{=} \lim_{n\to\infty} \int_\Omega f \,\mathrm{d}\mu.$$

(ii) Sei N die Nullmenge, auf der die Annahme aus (i) nicht gilt. Es gilt $f_n \mathbf{1}_{N^C} \to f \mathbf{1}_{N^C}$, $n \to \infty$, für alle $\omega \in \Omega$. Wegen (i) gilt

$$\int_\Omega f_n \,\mathrm{d}\mu = \int_\Omega f_n \mathbf{1}_{N^C} \,\mathrm{d}\mu \overset{\text{(i)}, n\to\infty}{\longrightarrow} \int_\Omega f \mathbf{1}_{N^C} \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu,$$

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

□

Anwendung 3.2.2. Sei $f\colon (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\mathbb{R}))$ messbar und nichtnegativ und μ ein Maß auf $\mathcal{A}$. Dann ist

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

ein Maß auf $\mathcal{A}$.

Beweis.

$\nu \geq 0$ ✓ wegen Integral über nichtnegative Funktion

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

$$\begin{aligned}
\nu\Big(\dot{\bigcup_{n=1}^{\infty}} A_n\Big) &\overset{\text{Def.}}{=} \int_\Omega f \mathbf{1}_{\dot{\bigcup}_{n=1}^{\infty} A_n} \, \mathrm{d}\mu \\
&= \int_\Omega f \cdot \Big(\sum_{n=1}^{\infty} \mathbf{1}_{A_n}\Big) \, \mathrm{d}\mu \\
&\underset{\text{Reihe}}{\overset{\text{Def.}}{=}} \int_\Omega \lim_{N\to\infty} \sum_{n=1}^{N} f \mathbf{1}_{A_n} \, \mathrm{d}\mu \\
&\overset{3.2.1}{=} \lim_{N\to\infty} \int_\Omega \sum_{n=1}^{N} f \mathbf{1}_{A_n} \, \mathrm{d}\mu \\
&\overset{\text{Lin.}}{=} \lim_{N\to\infty} \sum_{n=1}^{N} \int_\Omega f \mathbf{1}_{A_n} \, \mathrm{d}\mu \\
&\overset{\text{Def.}}{=} \lim_{N\to\infty} \sum_{n=1}^{N} \int_\Omega \nu(A_n) \, \mathrm{d}\mu \\
&\underset{\text{Reihe}}{\overset{\text{Def.}}{=}} \sum_{n=1}^{\infty} \int_\Omega \nu(A_n) \, \mathrm{d}\mu.
\end{aligned}$$

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

Bemerkung 3.2.3.

(i) Man schreibt mit ν aus vorheriger Anwendung auch

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

und nennt f die **Radon-Nikodým-Ableitung** oder **-Dichte von ν bezüglich μ**.

(ii) Wir kennen das schon: Ist $\mathbb{P}_F$ ein Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F und Dichte f, so gilt $\frac{\mathrm{d}\mathbb{P}_F}{\mathrm{d}\lambda} = f$.

(iii) ν ist ein Wahrscheinlichkeitsmaß (d. h. $\nu(\Omega) = 1$) genau dann, wenn $\int_\Omega f \, \mathrm{d}\mu = 1$. Beachte: So war eine Dichtefunktion auf $\mathbb{R}$ definiert.

Anwendung 3.2.4. Kommen wir nochmal zurück zu Beispiel 3.1.12, (ii), und bestätigen die Behauptung. Sei $f \geq 0$ und μ das Zählmaß auf $\mathbb{N}$

$$\begin{aligned}
\int_{\mathbb{N}} f \, d\mu &= \int_{\mathbb{N}} f \sum_{k=0}^{\infty} \mathbf{1}_{\{k\}} \, d\mu \\
&= \int_{\mathbb{N}} \lim_{n\to\infty} f \sum_{k=0}^{n} \mathbf{1}_{\{k\}} \, d\mu \\
&\overset{3.2.1}{=} \lim_{n\to\infty} \int_{\mathbb{R}} \sum_{k=0}^{n} f \, \mathbf{1}_{\{k\}} \, d\mu \\
&\overset{\text{Lin.}}{=} \lim_{n\to\infty} \sum_{k=0}^{n} \int_{\mathbb{R}} f(k) \, \mathbf{1}_{\{k\}} \, d\mu \\
&\overset{\text{Def. Int.}}{=} \lim_{n\to\infty} \sum_{k=0}^{n} f(k)\mu(\{k\}) \\
&\overset{\text{Zählmaß}}{=} \sum_{k=0}^{\infty} f(k).
\end{aligned}$$

Vorlesung 12

Satz 3.2.5. (Lemma von Fatou)
Sei (f_n) ein Folge *nichtnegativer* messbarer numerischer Funktionen auf $(\Omega, \mathcal{A})$, (f_n) muss dabei nicht konvergieren. Dann gilt

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

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

Wenn sogar (f_n) fast überall konvergiert und $(\int_\Omega f_n \,\mathrm{d}\mu)$ konvergiert, gilt damit

$$\int_\Omega \lim_{n\to\infty} f_n \,\mathrm{d}\mu \leq \lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu.$$

Die Ungleichung „$\leq$“ gilt in den Konvergenzsätzen also mit weniger Annahmen als 3.2.1 und gleich in 3.2.6.

Beweis. Definiere $g_n := \inf_{k\geq n} f_k$, g_n ist messbar für alle $n \in \mathbb{N}$, $g_n \leq f_n$ für alle $k \geq n$ und $\lim_{k\to\infty} g_k = \liminf_{n\to\infty} f_n$. 3.2.1 gibt

$$\begin{aligned}\int_\Omega \liminf_{n\to\infty} f_n \,\mathrm{d}\mu &= \int_\Omega \lim_{k\to\infty} g_k \,\mathrm{d}\mu \overset{3.2.1}{=} \lim_{k\to\infty} \int_\Omega g_k \,\mathrm{d}\mu \\ &= \liminf_{k\to\infty} \int_\Omega g_k \,\mathrm{d}\mu \overset{\text{Mon.}}{\leq} \liminf_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu.\end{aligned}$$

□

Satz 3.2.6. (Dominierte Konvergenz Theorem (DCT))
Seien $f, f_1, f_2, \ldots\colon \Omega \to \overline{\mathbb{R}}$ messbar. Es sollen gelten

(a) $\lim_{n\to\infty} f_n = f$ μ-fast überall,

(b) $|f_n| \leq g$ μ-fast überall für alle $n \in \mathbb{N}$,

für eine beliebige μ-integrierbare nichtnegative messbare numerische Funktion g. Dann sind $f, f_1, f_2, \ldots$ μ-integrierbar und

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

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 wieder mit der Hilfsfolge $(f_n \mathbf{1}_N)$ für die Nullmenge $N = \{(f_n) \text{ konvergiert nicht}\}$ den Fall der μ-f.ü. Konvergenz zeigen.

Der Beweis beruht auf einer elementaren Erkenntniss: Wenn $|f_n| \leq g$ gilt, so gilt auch $f_n \leq g$ 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 nichtnegative Funktionen bekommen und Fatou anwenden können.

(i)

$$\begin{aligned}\int_\Omega f \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu &\overset{\text{Lin.}}{=} \int_\Omega (f+g) \,\mathrm{d}\mu \\ &\overset{\text{Ann.}}{=} \int_\Omega \big(\lim_{n\to\infty} f_n + g\big) \,\mathrm{d}\mu \\ &= \int_\Omega \big(\liminf_{n\to\infty} f_n + g\big) \,\mathrm{d}\mu \\ &\overset{3.2.5}{\leq} \liminf_{n\to\infty} \int_\Omega (f_n+g) \,\mathrm{d}\mu \\ &\overset{\text{Lin.}}{=} \liminf_{n\to\infty} \Big(\int_\Omega f_n \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu\Big) \\ &= \liminf_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu + \int_\Omega g \,\mathrm{d}\mu.\end{aligned}$$

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

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

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

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

also

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

Beide Schritte zusammen ergeben

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

Also stimmen lim inf und lim sup überein und geben nach Analysis 1 den Grenzwert

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

□

Korollar 3.2.7. 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 C, und $f_n \to f$, $n \to \infty$, μ-f.ü., so gilt

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

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

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

□

3.3 Zufallsvariablen und Erwartungswerte

Damit wir genug Zeit haben, das Rechnen in der Stochastik in den Übungen zu üben, schieben wir einen ersten Blick in die Stochastik rein. Mit unserem bisherigen Wissen der Maß- und Integrationstheorie können wir über reelwertige Zufallsexperimente und Erwartungswerte sprechen und etliches ausrechnen.

Zur Erinnerung: Ein Wahrscheinlichkeitsmaß $\mathbb{P}_F$ auf $\mathcal{B}(\mathbb{R})$ mit Verteilungsfunktion F beschreibt ein reellwertiges Zufallsexperiment, bei dem die „gezogene Zahl" in $(a, b]$ liegt mit Wahrscheinlichkeit $\mathbb{P}_F((a, b]) = F(b) - F(a)$. Hat F eine Dichte f, so ist $\mathbb{P}_F((a, b]) = \int_a^b f(x)dx$. Ist F diskret, so gilt $\mathbb{P}_F((a, b]) = \sum_{a_k\in(a,b]} p_k$. Ein paar Integralen geben wir jetzt Namen und überlegen uns anschließend, wie man die Integrale in vielen Beispielen berechnen kann.

Definition 3.3.1.

(i)
$$\int_{\mathbb{R}} x \,\mathrm{d}\mathbb{P}_F(x)$$
heißt **Erwartungswert (EW)** von $\mathbb{P}_F$ oder Erwartungswert der Verteilungsfunktion F, wenn das Integral definiert ist.

(ii)
$$\int_{\mathbb{R}} x^k \,\mathrm{d}\mathbb{P}_F(x)$$
heißt **k-tes Moment** von $\mathbb{P}_F$ oder k-tes Moment der Verteilungsfunktion F, wenn das Integral definiert ist.

(iii)
$$\int_{\mathbb{R}} e^{\lambda x} \,\mathrm{d}\mathbb{P}_F(x)$$
heißt **exponentielles Moment** von $\mathbb{P}_F$ oder exponentielles Moment der Verteilungsfunktion F.

Beachte: Alle Integrale über nichtnegative Integranden sind definiert, das Integral könnte aber $+\infty$ sein. Damit sind exponentielle Momente und gerade Momente immer in $[0, +\infty]$ definiert. Man sagt auch ein Integral „existiert" statt ein Integral ist definiert.

Satz 3.3.2. (Integrale von Dichten)
Sei $\mathbb{P}_F$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und F habe Dichte f, d. h.

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

Dann gilt für $g \colon \mathbb{R} \to \overline{\mathbb{R}}$ messbar

$$\int_\Omega g \,\mathrm{d}\mathbb{P}_F \text{ existiert} \quad \Leftrightarrow \quad \int_\Omega g(x)f(x)\,\mathrm{d}x \text{ existiert}$$

und wenn die Integrale existieren ist

$$\int_\Omega g \,\mathrm{d}\mathbb{P}_F(x) = \int_\Omega g(x)f(x)\,\mathrm{d}x.$$

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

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

Warum gilt $(*)$, also $\mathbb{P}_F(A) = \int_A f(x)\,\mathrm{d}x$? Ist $A = (a, b]$, so gilt $\mathbb{P}_F((a,b)) = \int_a^b f(x)\,\mathrm{d}x$ weil F Dichte f hat. In 3.2.2 haben wir gezeigt, dass $\nu(A) = \int_A f(x)\,\mathrm{d}x$ 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

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

aufgrund der Linearität des Integrals.

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

$$\int_{\mathbb{R}} g \,\mathrm{d}\mathbb{P}_F \overset{3.2.1}{=} \lim_{n\to\infty} \int_{\mathbb{R}} g_n \,\mathrm{d}\mathbb{P}_F = \lim_{n\to\infty} \int_{\mathbb{R}} g_n(x) f(x)\,\mathrm{d}x \overset{3.2.1}{=} \int_{\mathbb{R}} g(x) f(x)\,\mathrm{d}x.$$

(ii) Für g beliebig zerlegt man g in $g^+ - g^-$ und wende das Vorherige an. Insbesondere zeigt dies sofort die Äquivalenz der Definiertheit der Integrale.

□

Hier ist eine kleine aber wichtige Bemerkung: Warum ist der erste Teil des Satzes wichtig? Natürlich ist das Integral über die Dichte besser zu handeln, das können wir mit einfacher Analysis ausrechnen. Wenn wir den Satz nicht hätten, wie würden wir zeigen, dass $\int_{\mathbb{R}} g(x)\,\mathrm{d}\mathbb{P}_F$ existiert oder nicht existiert? Unklar. Mit dem Satz betrachten wir einfach nur das Integral $\int_{\mathbb{R}} g(x) f(x)\,\mathrm{d}x$. Wenn dabei was schief geht, dann geht es auch für $\int_{\mathbb{R}} g\,\mathrm{d}\mathbb{P}_F$ schief und wir müssen nichts mehr tun (das sehen wir später zum Beispiel beim Erwartungswert der Cauchy-Verteilung). Wenn jedoch das Integral $\int_{\mathbb{R}} g(x) f(x)\,\mathrm{d}x$ existiert, dann ist das gerade $\int_{\mathbb{R}} g\,\mathrm{d}\mathbb{P}_F$.

Satz 3.3.3. (Integrale bei diskreten Verteilungen)
Sei $\mathbb{P}_F$ ein Maß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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 $N \in \mathbb{N} \cup \{+\infty\}$, $a_1, ..., a_N \in \mathbb{R}$ und $\sum_{k=1}^N p_k = 1$. Dann gilt für $g : \mathbb{R} \to \overline{\mathbb{R}}$ messbar

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

und wenn das Integral definiert ist, gilt

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

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

Vorlesung 13

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 $g = g\mathbf{1}_{\{a_1,\dots,a_N\}}$ $\mathbf{P}_F$-fast überall. Es folgt dann

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

(b) $N = +\infty$ funktioniert im Prinzip genauso, wir müssen nur einmal monotone Konvergenz für die wachsende Folge von messbaren Funktionen $f_n := \sum_{k=1}^n g(k)\mathbf{1}_{\{a_k\}}$ nutzen, um Integral und Summe zu vertauschen:

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

(ii) Sei nun g messbar, aber nicht mehr nichtnegativ. Wir zerlegen wie immer in Positiv- und Negativteil und erinnern daran, dass das Integral per Definition nur definiert ist (d. h. existiert) wenn die Integrale über Positiv- und Negativteil endlich sind. Die beiden können wir dann mit dem Fall $g \geq 0$ behandeln. Es gilt also wegen Teil

(i)

$$\begin{aligned}\int_{\mathbb{R}} g\,\mathrm{d}\mathbb{P}_F \text{ ist definiert} \quad &\Leftrightarrow \quad \int_{\mathbb{R}} g^+\,\mathrm{d}\mathbb{P}_F < \infty, \int_{\mathbb{R}} g^-\,\mathrm{d}\mathbb{P}_F < \infty\\ &\overset{g^+\geq 0}{\underset{g^-\geq 0}{\Leftrightarrow}} \quad \sum_{k=1}^N p_k g^+(a_k) < \infty, \sum_{k=1}^N p_k g^-(a_k) < \infty\\ &\Leftrightarrow \quad \sum_{k=1}^N |g(a_k)| p_k < \infty,\end{aligned}$$

weil $|g(a_k)| = g^+(a_k) + g^-(a_k)$. Wenn das Integral definiert ist, gilt folglich

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

□

Warum der abstrakte Erwartungswert überhaupt Erwartungswert heißt, wird aus Beispielen am klarsten ersichtlich. Jeder hat eine Vorstellung, was der „Erwartungswert" eines Würfelwurfs sein sollte, nämlich 3,5. Beim gleichverteilten Ziehen aus $[0,1]$ „erwarten" wir im Mittel $\frac{1}{2}$. Das kommt natürlich bei unserer Modellierung auch raus:

Beispiel 3.3.4.

(i) Erwartungswert des Würfelexperiments: Seien dazu wieder $a_1 = 1, ..., a_6 = 6$ und $p_1 = ... = p_6 = \frac{1}{6}$. Dann gilt

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

(ii) Erwartungswert der uniformen Verteilung $\mathcal{U}([0,1])$ auf $[0,1]$ mit Dichte $f(x) = \mathbf{1}_{[0,1]}$. Also ist mit dem Satz

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

Warnung 3.3.5. Es ist nicht immer der Fall, dass eine Verteilungsfunktion diskret ist oder eine Dichte hat. In diesen Fällen gibt es keine einfach Formel für $\int_{\mathbb{R}} g\,\mathrm{d}\mathbb{P}_F$! Woran erkennt man sofort, ob eine Verteilungsfunktion diskret ist, eine Dichte hat, oder keines von beiden ist? Es gibt nur drei Fälle:

- Ist F überall stetig, so gibt es eine Dichte.

- Hat F ausschließlich konstante Abschnitte mit Sprüngen dazwischen, so ist F diskret (die a_k sind die Stellen der Sprünge, die p_k sind die Höhen der Sprünge), siehe die Bildchen in 1.4.7.
- Hat F sowohl Sprünge, als auch Bereiche, in denen F stetig wächst, so hat F weder eine Dichte noch ist F diskret. Ein Beispiel ist auf Übungsblatt 7.

In vielen Beispielen müssen wir gar nicht rechnen, sondern sehen das Ergebniss 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 $\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 0$. Das können wir direkt ausnutzen, um viele Erwartungswerte direkt als 0 zu erkennen:

Lemma 3.3.6. Hat F Dichte f und existiert der Erwartungswert von $\mathbb{P}_F$, so gilt

$$f \text{ achsensymmetrisch} \quad \Rightarrow \quad \int_{\mathbb{R}} x\,\mathrm{d}\mathbb{P}_F(x) = 0.$$

Beweis. Ist f achsensymmetrisch, so ist $h(x) := xf(x)$ punktsymmetrisch weil dann $h(-x) = -xf(-x) = -xf(x) = -h(x)$. Wegen 3.3.2 ist also

$$\int_{\mathbb{R}} x\,\mathrm{d}\mathbb{P}_F(x) = \int_{\mathbb{R}} xf(x)\,\mathrm{d}x \overset{\text{punktsym.}}{=} 0.$$

□

Warnung 3.3.7. Wir müssen in Lemma 3.3.6 auf jeden Fall annehmen, dass der Erwartungswert existiert. Beispielsweise hat die Cauchy-Verteilung eine achsensymmetrische Dichte, der Erwartungswert 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:

$$\begin{aligned}
\int_{\mathbb{R}} x^+\,\mathrm{d}\mathbb{P}_F(x) &= \int_0^\infty x\,\mathrm{d}\mathbb{P}_F(x) \\
&\overset{3.3.2}{=} \frac{1}{\pi}\int_0^\infty x\frac{1}{1+x^2}x \\
&= \frac{1}{\pi}\int_0^\infty x\frac{1}{1/x+x}\,\mathrm{d}x \\
&\geq \frac{1}{\pi}\int_1^\infty \frac{1}{1/x+x}\,\mathrm{d}x \\
&\geq \frac{1}{\pi}\int_1^\infty \frac{1}{1+x}\,\mathrm{d}x \\
&\overset{3.2.1}{=} \frac{1}{\pi}\lim_{N\to\infty}\int_1^N \frac{1}{1+x}\,\mathrm{d}x \\
&\overset{\text{Hauptsatz}}{=} \frac{1}{\pi}\lim_{N\to\infty}\Big[\ln(1+x)\Big]_1^N = +\infty.
\end{aligned}$$

Genauso ist

$$\int_{\mathbb{R}} x^-\,\mathrm{d}\mathbb{P}_F(x) = \frac{1}{\pi}\int_{-\infty}^0 -x\frac{1}{1+x^2}\,\mathrm{d}x = \frac{1}{\pi}\int_0^\infty x\frac{1}{1+x^2}\,\mathrm{d}x = +\infty.$$

Damit ist $\int_{\mathbb{R}} x\,\mathrm{d}\mathbb{P}_F(x) = \int_{\mathbb{R}} x^+\,\mathrm{d}\mathbb{P}_F(x) - \int_{\mathbb{R}} x^-\,\mathrm{d}\mathbb{P}_F(x) = +\infty - (+\infty)$ nicht definiert. Für die Cauchy verteilt existiert der Erwartungswert also nicht!

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:

$$\begin{aligned}\int_1^\infty \frac{1}{x^p}\,\mathrm{d}x &\overset{3.2.1}{=} \lim_{N\to\infty}\int_1^N \frac{1}{x^p}\,\mathrm{d}x\\ &= \lim_{N\to\infty}\begin{cases}\Big[\ln(x)\Big]_1^N & : p=1\\ \frac{1}{1-p}\Big[x^{1-p}\Big]_1^N & : p\neq 1\end{cases}\\ &= \lim_{N\to\infty}\begin{cases}\ln(N) & : p=1\\ \frac{1}{1-p}(N^{1-p}-1) & : p\neq 1\end{cases}\\ &= \begin{cases}+\infty & : p\leq 1\\ \frac{1}{p-1}<\infty & : p>1\end{cases}.\end{aligned}$$

Unsere Heuristik für die Integrierbarkeit bei unendlich ist es, grob mit der Funktion $\frac{1}{x}$ zu vergleichen. Fällt unser Integrand deutlich schneller gegen 0, ist vermutlich ist vermutlich $\int_1^\infty f(x)\,\mathrm{d}x<\infty$. Fällt der Integrand langsamer als $\frac{1}{x}$ gegen 0 ist sicherlich $\int_1^\infty f(x)\,\mathrm{d}x=+\infty$.
Hier sind ein paar Beispiele:

(i) Nochmal die Cauchy-Verteilung: $\frac{1}{\pi}x\frac{1}{1+x^2}$ ist bei ∞ ungefähr wie $\frac{1}{x}$, das ist aber *nicht* integrierbar. Die Heuristik sagt uns also auf einen Blick, dass etwas schief geht. Um daraus ein sauberes Argument zu machen, müssen wir leider doch die Abschätzung aus 3.3.7 durchgehen.

(ii) Für welches β hat $\mathrm{Exp}(\lambda)$ ein endliches exponentielle Moment, wann ist also

$$\int_{\mathbb{R}} e^{\beta x}e^{-\lambda x}\mathbf{1}_{[0,\infty)}(x)\,\mathrm{d}x=\int_0^\infty e^{x(\beta-\lambda)}\,\mathrm{d}x$$

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

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{\mathbb{R}} e^{\beta x}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x$$

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 abschätzen können. Die Ungleichung gibt uns eine Schranke, wieviel Masse der Verteilung mindestens in $[-a,a]$ liegt. Weil man auch sagt, „wieviel Masse in $[-a,a]$ konzentriert ist“ nennt man solche Ungleichungen **Konzentrationsungleichungen**.

Proposition 3.3.9. (Markov-Ungleichung für Polynome)
Sei $\mathbb{P}_F$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, sodass für eine gerade natürliche Zahl $2k$ das $2k$-te Moment endlich ist. Dann gilt

$$\mathbb{P}_F([-a,a]) \geq 1 - \frac{\int_{\mathbb{R}} x^{2k}\,\mathrm{d}\mathbb{P}_F(x)}{a^{2k}}.$$

Gleichbedeutend (Gegenereigniss) gilt

$$\mathbb{P}_F([-a,a]^C) \leq \frac{\int_{\mathbb{R}} x^{2k}\,\mathrm{d}\mathbb{P}_F(x)}{a^{2k}}.$$

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 $g(x)=x^k$ und die Menge $[-a,a]^C$. Für $x \in [-a,a]^C$ gilt natürlich $|x|>a$ und damit, k ist per Annahme gerade, $x^k=|x|^k \geq a^k$. Dann müssen wir nur noch die Monotonie vom Integral nutzen:

$$\begin{aligned}\int_{\mathbb{R}} x^k\,\mathrm{d}\mathbb{P}_F(x) &\overset{\text{Mon.}}{\geq} \int_{\mathbb{R}} x^k \mathbf{1}_{[-a,a]^C}(x)\,\mathrm{d}\mathbb{P}_F(x)\\ &\geq \int_{\mathbb{R}} \underbrace{a^k \mathbf{1}_{[-a,a]^C}(x)}_{\text{einfach}}\,\mathrm{d}\mathbb{P}_F(x)\\ &\overset{\text{Def.}}{=} a^k \mathbb{P}_F([-a,a]^C).\end{aligned}$$

Auflösen gibt

$$\mathbb{P}_F([-a,a]^C) \leq \frac{\int_{\mathbb{R}} x^{k}\,\mathrm{d}\mathbb{P}_F(x)}{a^{k}}.$$

Die zweite Ungleichung ist die Gegenwahrscheinlichkeit, weil für Wahrscheinlichkeitsmaße $\mathbb{P}_F(B)=1-\mathbb{P}_F(B^C)$ gilt. □

Anwendung 3.3.10. (Vergleiche Übungsblatt 4, Aufgabe 2)
Aufgabe: für $\mu=0$ finde $a>0$ mit $\mathbb{P}_F([-a,a]) \geq 0.99$, wobei F die Normalverteilung $\mathcal{N}(0,\sigma^2)$ ist. Damit uns die Markovungleichung 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,

$$\int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = \sigma^2 \quad \text{und} \quad \int_{\mathbb{R}} x^8 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,\mathrm{d}x = 105\sigma^8.$$

Um die Konzentration der Normalverteilung in $[-a,a]$ abzuschätzen, verwenden wir Proposition 3.3.9 einmal mit $k=2$ und einmal mit $k=8$:

$$\mathbb{P}_F([-a,a]) \geq 1 - \frac{\sigma^2}{a^2}, \quad \text{sowie} \quad \mathbb{P}_F([-a,a]) \geq 1 - \frac{105\sigma^8}{a^8}.$$

Wir probieren jetzt mal mit dem ganz konkreten Beispiel $\sigma = 0.1$ aus, welche Abschätzung besser ist. Einsetzen und Umformen gibt beim zweiten Moment

$$1 - \frac{(0.1)^2}{a^2} \geq 0.99 \quad \Leftrightarrow \quad a \geq 1$$

und beim 8.ten Moment

$$1 - \frac{105(0.1)^8}{a^8} \geq 0.99 \quad \Leftrightarrow \quad a \geq \frac{105^{\frac{1}{8}} \cdot 0.1}{(0,01)^{\frac{1}{8}}} \approx 0.32.$$

Zusammenfassend gibt uns die Markovungleichung für $k = 2$ die Information, dass mindestens 0.99 der Masse von $\mathcal{N}(0, 0.1)$ in $[-1, 1]$ liegt. Viel besser ist aber die Aussage für $k = 8$: Mindestens 0.99 der Masse von $\mathcal{N}(0, 0.1)$ liegt schon in $[-0.32, 0.32]$! Das bedeutet einfach nur, dass in diesem Beispiel die Markovungleichung für $k = 8$ viel genauer an dem echten Wert liegt als für $k = 2$. Diese kleine Abschätzung zum Spass bestätigt quantitativ, was wir aus den Bildchen in **??** qualitativ schon erraten konnten: Für kleines σ ist $\mathcal{N}(0, \sigma^2)$ sehr stark um den Ursprung herum konzentriert. Um etwas Gefühl zu bekommen, ersetzt einfach mal $\sigma = 0.1$ durch $\sigma = 0.01$ und setzt die aufgelöste Formel in einen Taschenrechner ein. Die Konzentration um den Ursprung wird dann noch viel stärker.

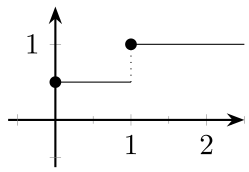
Vorlesung 14

Diskussion 3.3.11. So weit so gut, aber was machen wir mit sehr einfachen zufälligen Experimenten, die wie Würfeln nur endlich viele „Werte“ zulassen? Das mit Maßen auf $\mathcal{B}(\mathbb{R})$ zu modellieren, ist natürlich völlig übertrieben! Beispielsweise für den Münzwurf haben wir zwei Modellierungsvarianten diskutiert:

Variante 1:

$$\Omega = \{\text{Kopf}, \ \text{Zahl}\},$$
$$\mathcal{A} = \mathcal{P}(\Omega),$$
$$\mathbb{P}(\{\text{Kopf}\}) = \mathbb{P}(\{\text{Zahl}\}) = \frac{1}{2}, \mathbb{P}(\Omega) = 1, \mathbb{P}(\emptyset) = 0.$$

Variante 2: $\Omega' = \mathbb{R}$, $\mathcal{A}' = \mathcal{B}(\mathbb{R})$, $\mathbb{P}' = \mathbb{P}_F$, wobei $\mathbb{P}_F$ das Maß mit diskreter Verteilungsfunktion F ist:



Beide Modelle modellieren ein Experiment, bei dem zwei Ereignisse jeweils mit Wahrscheinlichkeit $\frac{1}{2}$ auftreten. Der Vorteil des zweiten Modells ist, dass wir reelle Zahlen bekommen, wir können damit also einen Erwartungswert definieren. Da wir den Ereignissen die Werte 0 und 1 geben, ist der Erwartungswert $\frac{1}{2}$. In der zweiten Variante haben wir zwei Dinge auf einmal modelliert:

- Was passiert? (Welches zufällige Ereigniss passiert beim Würfeln)
- Was wird ausgezahlt? (Was wird für Kopf bzw. Zahl ausgezahlt)

Normalerweise haben wir natürlich viel Freude an viel zu komplizierten mathematischen Modellen. In diesem Fall, werden wir aber Variante 1 weiterverfolgen. Wir müssen uns aber noch Gedanken über Übersetzung der Ereignisse in Auszahlungen machen. Dafür kommen messbare Abbildungen $X : \Omega \to \mathbb{R}$ ins Spiel.

Ab jetzt: Trenne in der Modellierung von „Was passiert?“ (also Ereignisse und Wahrscheinlichkeiten) und „Was wird ausgezahlt?“.

Definition 3.3.12. Ein **stochastisches Modell** besteht aus

(i) einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$,

(ii) einer $(\mathcal{A}, \mathcal{B}(\mathbb{R}))$-messbaren Abbildung $X : \Omega \to \mathbb{R}$.

(i) beschreibt die zufälligen Ereignisse, (ii) beschreibt die „Auszahlung“. X wird auch **Zufallsvariable (ZV)** genannt.

Definition 3.3.13. Sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$.

(i) Die **Verteilung der Zufallsvariablen** X ist definiert als

$$\mathbb{P}_X(B) := \mathbb{P}(X \in B) \overset{\text{Not.}}{=} \mathbb{P}(\{\omega \in \Omega\colon X(\omega) \in B\}) \overset{\text{Not.}}{=} \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}).$$

Unabhängig von dem zugrundeliegenden Wahrscheinlichkeitsraums ist $\mathbb{P}_X$ ist ein Maß auf $\mathcal{B}(\mathbb{R})$ und zwar der push-forward von X.

(ii) Die **Verteilungsfunktion der Zufallsvariablen** X ist definiert als

$$F_X(t) := \mathbb{P}_X((-\infty, t]) \overset{\text{Def.}}{=} \mathbb{P}(X \le t), \quad t \in \mathbb{R}.$$

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 $\mathbb{P}_X$“. F_X heißt auch die „Verteilungsfunktion von X“. Weil so viele Indizes natürlich nerven, werden wir immer $X \sim F$ schreiben, wenn wir meinen, dass X gemäß F verteilt ist. 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 $\mu(\{f \le t\})$ als Abkürzung für $\mu(\{\omega : f(\omega) \le t\})$ gewöhnen müssen. In der Stochastik gehen wir noch einen Schritt weiter und lassen die Klammern auch noch weg. Wir schreiben daher immer

$$\mathbb{P}(X < t), \quad \mathbb{P}(X \in (a, b]), \quad \mathbb{P}(X = a)$$

und so weiter. Das liest sich als „Wahrscheinlichkeit, dass X kleiner als t ist“ auch ziemlich natürlich.

Definition 3.3.14. Zwei Zufallsvariablen X und Y heißen **identisch verteilt**, falls $F_X = F_Y$ bzw. $\mathbb{P}_X = \mathbb{P}_Y$. Haben zwei Zufallsvariablen die gleiche Verteilungsfunktion, so sehen wir sie als gleichwertig an, d. h. die ZV zusammen mit den zugrunde liegenden Wahrscheinlichkeitsräumen beschreiben das gleiche Experiment.

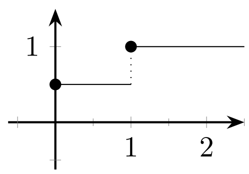
Beispiel. Zurück zum Münzwurf mit Auszahlung 1 für Zahl und 0 für Kopf. Wir schreiben dazu zwei stochatische Modelle hin. Zum einen sei

$$\begin{aligned} \Omega &= \{\text{Kopf}, \ \text{Zahl}\}, \\ \mathcal{A} &= \mathcal{P}(\Omega), \\ \mathbb{P}(\{\text{Kopf}\}) = \mathbb{P}(\{\text{Zahl}\}) &= \frac{1}{2}, \mathbb{P}(\Omega) = 1, \mathbb{P}(\emptyset) = 0, \end{aligned}$$

mit Zufallsvariablen (Auszahlungsfunktion) definiert als

$$X(\text{Kopf}) = 0, \quad X(\text{Zahl}) = 1.$$

Zum anderen sei wieder $\Omega' = \mathbb{R}$, $\mathcal{A}' = \mathcal{B}(\mathbb{R})$ und $\mathbb{P}' = \mathbb{P}_F$ mit der Verteilungsfunktion



Wie wir schon lange wissen, ist $\mathbb{P}_F$ eine Summe von Dirac-Maßen, nämlich $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Da wir die Auszahlung schon direkt im Modell modelliert haben, zahlen wir genau den Betrag aus, der gezogen wird. Das machen wir mit der Zufallsvariablen $Y(\omega) = \omega$. Berechnen wir nun die Verteilungen der Zufallsvariablen X und Y:

$$\begin{aligned}\mathbb{P}_X(B) \overset{\text{Def.}}{=} \mathbb{P}'(X \in B) &= \begin{cases} \frac{1}{2} & : 0 \in B, 1 \notin B \text{ oder } 1 \in B, 0 \notin B \\ 1 & : 0, 1 \in B \\ 0 & : 0 \notin B, 1 \notin B \end{cases} \\ &= \frac{1}{2}\delta_0(B) + \frac{1}{2}\delta_1(B)\end{aligned}$$

sowie

$$\mathbb{P}_Y(B) \overset{\text{Def.}}{=} \mathbb{P}(Y \in B) = \mathbb{P}(\{\omega \in \mathbb{R} : Y(\omega) \in B) = \mathbb{P}(\{\omega \in \mathbb{R} : \omega \in B) \overset{\text{Def.}}{=} \mathbb{P}_F(B).$$

Weil $\mathbb{P}_F = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$, gilt also $\mathbb{P}_X = \mathbb{P}_Y$ und damit auch $F_X = F_Y$. Damit sind X und Y identische verteilte Zufallsvariablen und wir sehen die beiden stochastischen Modelle als gleichwertige Modelle für den Münzwurf an.

Definition 3.3.15. (i) Eine Zufallsvariable X heißt **diskret**, falls F_X eine diskrete Verteilungsfunktion ist. Dann heißen die Sprungstellen $a_1, ..., a_N$ „Werte, die X annimmt“ und $p_1, ..., p_N$ „Wahrscheinlichkeiten, dass X den entsprechenden Wert annimmt“. Die Formulierung rechtfertigt sich aus folgender kleiner Rechnung:

$$\begin{aligned}\mathbb{P}(X = a_k) &\overset{\text{Maß}}{=} \mathbb{P}(X \leq a_k) - \mathbb{P}(X < a_k) \\ &= \mathbb{P}(X \leq a_k) - \lim_{s \uparrow a_k} \mathbb{P}(X \leq s) \\ &= F_X(a_k) - F_X(a_k-) = p_k.\end{aligned}$$

(ii) Eine Zufallsvariable heißt **absolutstetig** (oder nur stetig) mit Dichte f, falls die Verteilungsfunktion F_X von X die Dichte f hat, es gilt also

$$\mathbb{P}(X \in (a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a) = \int_a^b f(x)dx.$$

Um uns das Leben leicht zu machen, sollten wir im Kopf immer wie folgt denken: Wenn X gemäß F verteilt ist und F absolut stetig ist, so gilt $\mathbb{P}(X \in (a, b]) = \int_a^b f(x)\,dx$. Ist F diskret, so gilt $\mathbb{P}(X = a_k) = p_k$ für irgendwelche $a_1, ...$ und Wahrscheinlichkeiten p_k, die sich auf 1 summieren. 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.

Beispiel 3.3.16. Wir kennen die meisten Beispiele schon:

Diskrete ZV	Stetige ZV
$X \sim \text{Poi}(\lambda)$	$X \sim \mathcal{N}(\mu, \sigma^2)$
$X \sim \text{Bin}(n, p)$	$X \sim \text{Cauchy}(s, t)$
$X \sim \text{Ber}(p)$	$X \sim \text{Exp}(\lambda)$
$X \sim \text{Geo}(\lambda)$	$X \sim \mathcal{U}([a, b])$
$X \sim$ Gleichv. auf endl. Ω	

Im Appendix gibt es eine Übersicht über die Verteilungen und ein paar Animationen zum Rumspielen.

Satz 3.3.17. (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 $(\Omega, \mathcal{A}, \mathbb{P})$, so dass $X \sim F$.

Beweis. Wir geben zunächst einen allgemeinen Beweis. Dazu sei der Wahrscheinlichkeitsraum $\mathbb{R} = \Omega$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, $\mathbb{P} = \mathbb{P}_F$ und die Zufallsvariable $X(\omega) = \omega$. Beachte: $X(\omega) = \omega$ ist eine stetige Abbildung von $\mathbb{R}$ nach $\mathbb{R}$ und damit auch messbar. Berechnen wir die Verteilungsfuntion dieses stochastischen Modells:

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}_F(X \in B) = \mathbb{P}_F(\{\omega \colon X(\omega) \in B\}) \overset{\text{Def. } X}{=} \mathbb{P}_F(B).$$

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 3.3.18. Alle diskreten ZV hätten auch *ohne* Maßtheorie konstruiert werden können! Sei dazu F eine diskrete Verteilungsfunktion, die an N-vielen Stellen a_i um p_i nach oben springt. Sei nun $\Omega = \{\omega_1, ..., \omega_N\}$ eine beliebige Menge mit N Elementen. Auf Ω wählen wir als σ-Algebra $\mathcal{A} = \mathcal{P}(\Omega)$. Das Maß definieren wir, indem wir es auf den Elementarereignissen als $\mathbb{P}(\omega_i) = p_i$ definieren und mit der σ-Additivität für beliebiges $A \in \mathcal{A}$ fortsetzen, d. h. $\mathbb{P}(A) = \sum_{\omega_k \in A} \mathbb{P}(\{\omega_k\})$. Als Zufallsvariable wählen wir die Abbildung $X(\omega_i) := a_i$. 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 $\mathbb{P}(X = a_i) = p_i$, also ist X gemäß F verteilt. Dieser direkte Beweis 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 3.3.17 zu nutzen.

Um die Begriffe zu üben, schauen wir uns eine kleine Rechnung an. Wir behaupten, dass $Y \sim \operatorname{Exp}(1)$, wenn $Y = -\ln(U)$ und $U \sim \mathcal{U}([0,1]$. Probieren wir die Begriffe aus und berechnen definitionsgemäß die Verteilungsfunktion von Y durch Auflösen:

$$F_Y(t) = \mathbb{P}(Y \leq t) \overset{\text{Def.}}{=} \mathbb{P}(-\ln(U) \leq t) = \mathbb{P}(U \geq \exp(-t)) = 1 - \mathbb{P}(U \leq \exp(-t)).$$

Wenn wir jetzt die Verteilungsfunktion $F_U(t) = t\mathbf{1}_{[0,1]}(t)$ 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.

Definition 3.3.19. Ist X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$, so heißen, falls die Integrale existieren,

(i)
$$\mathbb{E}[X] := \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega) \qquad \textbf{Erwartungswert} \text{ von X,}$$

(ii)
$$\mathbb{E}[X^k] := \int_\Omega X^k(\omega)\,\mathrm{d}\mathbb{P}(\omega) \qquad k\textbf{-tes Moment} \text{ von X,}$$

(iii)
$$\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] \qquad \textbf{Varianz} \text{ von X,}$$

(iv)
$$\mathbb{E}[e^{\lambda X}] := \int_\Omega e^{\lambda X(\omega)}\,\mathrm{d}\mathbb{P}(\omega) \qquad \textbf{exponentielles Moment} \text{ von X.}$$

Allgemein definiert man für $g : \mathbb{R} \to \overline{\mathbb{R}}$ messbar

$$\mathbb{E}[g(X)] := \int_\Omega g(X(\omega))\,\mathrm{d}\mathbb{P}(\omega),$$

falls das Integral definiert ist.

Zur Notation: Wegen der allgemeinen Äquivalenzen

$$\begin{aligned} \int f\,\mathrm{d}\mu \text{ ist def.} \quad &\overset{\text{Def.}}{\Leftrightarrow} \quad \int f^+\,\mathrm{d}\mu < \infty, \int f^-\,\mathrm{d}\mu < \infty \\ &\overset{\text{Übung}}{\Leftrightarrow} \quad \int |f|\,\mathrm{d}\mu < \infty \end{aligned}$$

schreiben wir meistens bequemer „$\mathbb{E}[|g(X)|] < \infty$“ anstelle von „$\mathbb{E}[g(X)]$“ existiert.

Vorlesung 15

Bemerkung. Die Notation $\mathbb{E}[g(X)]$ ist etwas unglücklich weil das Integral nicht nur g und X abhängt, sondern auch von $\mathbb{P}$. Daher sollte man eher $\mathbb{E}_{\mathbb{P}}[g(x)]$ schreiben, man lässt aber meisten das $\mathbb{P}$ aus Faulheit weg.

Lemma 3.3.20. 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,

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) \,\mathrm{d}\mathbb{P}_F(x)$$

für messbare numerische Abbildungen $g\colon \mathbb{R} \to \overline{\mathbb{R}}$. Wie immer gilt die Gleichheit, wenn eine Seite (und damit die die andere Seite) definiert ist.

Beweis. Mit dem Transformationssatz 3.1.16 gilt in einem Schaubild

$$\begin{tikzcd}
(\Omega, \mathbb{P}) \arrow[r, "X"] \arrow[dr, "g\circ X"'] & (\mathbb{R}, \mathbb{P}_X) \arrow[d, "g"] \\
& \overline{\mathbb{R}}
\end{tikzcd}$$

wobei $\mathbb{P}_X$ der push-forward von X ist. Weil definitionsgemäß $\mathbb{P}_X = \mathbb{P}_F$ ist, gibt das sauber ausgeschrieben

$$\mathbb{E}[g(X)] \overset{\text{Def.}}{=} \int_{\Omega} g(X(\omega)) \,\mathrm{d}\mathbb{P}(\omega) \overset{3.1.16}{=} \int_{\mathbb{R}} g(x) \,\mathrm{d}\mathbb{P}_X(x) \overset{\mathbb{P}_X = \mathbb{P}_F}{=} \int_{\mathbb{R}} g(x) \,\mathrm{d}\mathbb{P}_F(x).$$

$\square$

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 Verteilungsfunktion F_X von X ab. Das ist der Grund, warum man sich üblicherweise nur für die Verteilung 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. Dann gilt

$$\mathbb{E}[g(X)] = \mathbb{E}[g(Y)]$$

für alle $g\colon \mathbb{R} \to \overline{\mathbb{R}}$ messbar.

Jetzt wissen wir auch schon, wie wir $\mathbb{E}[g(X)]$ berechnen können, das haben wir nämlich schon gemacht:

Satz 3.3.21. (Berechnungsregeln)
Sei X eine Zufallsvariable mit Verteilungsfunktion F, die auf irgendeinem Wahrscheinlichkeitsraum definiert ist.

(i) Hat F Dichte f, so gilt

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) \,\mathrm{d}x.$$

(ii) Ist F diskret mit Sprungstellen $a_1, ..., a_N$ und Sprunghöhen $p_a, ..., p_N$, so gilt

$$\mathbb{E}[g(X)] = \sum_{k=1}^{N} p_k \, g(a_k).$$

Im diskreten Fall kann man den Satz noch intuitiver umformulieren (nutzt das!): Ist X eine Zufallsvariable, die endlich oder abzählbar viele Werte $a_1, ..., a_N$ annimmt, so gilt die einfache Formel

$$\mathbb{E}[g(X)] = \sum_{k=1}^{N} \mathbb{P}(X = a_k) \, g(a_k).$$

Beweis. Dazu kombinieren wir nur die Formel aus Satz 3.3.20 mit den Formeln aus Satz 3.3.2 und Satz 3.3.3. □

Beispiel 3.3.22.

- Für $X \sim \mathcal{N}(\mu, \sigma^2)$ gilt $\mathbb{E}[X] = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x = \mu$
- Für $X \sim \mathrm{Ber}(p)$ gilt $\mathbb{E}[X] = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p$
- Für $X \sim \mathrm{Poi}(\lambda)$ 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.

Proposition 3.3.23. (Rechenregeln für den Erwartungswert)
Seien X, Y Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\mathbb{E}[|X|], \mathbb{E}[|Y|] < \infty$. Dann 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., d. h. $\mathbb{P}(X = \alpha) = 1$, so ist $\mathbb{E}[X] = a$

(iv) $\mathbb{P}(X \in A) = \mathbb{E}[\mathbf{1}_A(X)]$

Beweis. Wir müssen nur beachten, dass Erwartungswerte per Definition Integrale sind. Dann können wir die Rechenregeln für Integrale direkt anwerden. 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:

$$\mathbb{E}[X] = \int_\Omega X(\omega) \, \mathrm{d}\mathbb{P}(\omega) = \int_\Omega \underbrace{\alpha \mathbf{1}_\Omega}_{\text{einfach}} \, \mathrm{d}\mathbb{P} \overset{\text{Def.}}{=} \alpha \mathbb{P}(\Omega) = \alpha$$

(iv) Hier müssen wir nur die Definitionen im Kopf klar bekommen:

$$\mathbb{E}[\mathbf{1}_A(X)] = \int_\Omega \mathbf{1}_A(X(\omega)) \, \mathrm{d}\mathbb{P}(\omega) \overset{\text{Trafo}}{=} \int_{\mathbb{R}} \underbrace{\mathbf{1}_A(x)}_{\text{einfach}} \, \mathrm{d}\mathbb{P}_X(x) = \mathbb{P}_X(A) \overset{\text{Def.}}{=} \mathbb{P}(X \in A)$$

□

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 $\mathbb{P}(X = 7) = 1$. Viel interessanter wäre zum Beispiel, jeden Tag um 7 Uhr die Temperatur zu messen. Die entsprechende Zufallsvariable wäre nicht fast sicher konstant.

Korollar 3.3.24. Ist $\mathbb{E}[X^2] < \infty$, so ist $\mathbb{V}(X) < \infty$ und es gilt

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

Es gilt $\mathbb{V}(X) = 0$ genau dann, wenn X fast sicher den gleichen Wert annimmt, dieser ist dann $\mathbb{E}[X]$.

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! □

Satz 3.3.25. (Konvergenzsätze für Zufallsvariablen)
Seien $X, X_1, X_2, ...$ Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$.

(i) MCT: Gilt $0 \leq X_1 \leq X_2 \leq ... \leq X$ und $\lim\limits_{n\to\infty} X_n = X$ $\mathbb{P}$-fast sicher, so gilt

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

(ii) DCT: Gilt $|X_n| \leq C$ $\mathbb{P}$-fast sicher für alle $n \in \mathbb{N}$ und gilt $\lim\limits_{n\to\infty} X_n = X$ $\mathbb{P}$-fast sicher, so gilt

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Beweis. Weil $\mathbb{E}[X_n] \overset{\text{Def.}}{=} \int_\Omega X_n \, \mathrm{d}\mathbb{P}$, ist das gerade Satz 3.2.1 und Korollar 3.2.7. □

Definition 3.3.26. Sei X eine Zufallsvariable, so heißt $\mathcal{M}_X(t) := \mathbb{E}[e^{tX}]$ 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 M_X aber auch gar nicht definiert.

Beispiel.

- 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 $X \sim \mathrm{Poi}(\lambda)$, so ist

$$\mathcal{M}_X(t) = \mathbb{E}[e^{tk}] = \sum_{k=0}^{\infty} e^{tX} \mathbb{P}(X = k) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} \overset{\text{Übung}}{=} e^{\lambda(e^t - 1)}$$

für alle $t \in \mathbb{R}$.

- Sei $X \sim \text{Cauchy}(s, t)$, so ist $\mathcal{M}_X$ nirgends definiert!

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 3.3.27. Sei X eine Zufallsvariable, für die $\mathcal{M}_X(t)$ für ein $\epsilon > 0$ in $(-\varepsilon, \varepsilon)$ definiert ist. Dann ist $\mathcal{M}_X$ an der Stelle 0 unendlich oft differenzierbar und es gilt

$$\mathbb{E}[X^n] = \mathcal{M}_X^{(n)}(0),$$

wobei $\mathcal{M}_X^{(n)}(0)$ die n-te Ableitung an der Stelle 0 ist.

Vorlesung 16

Beweis.

(a) Wir zeigen zunächst, dass $\mathcal{M}_X$ in $(-\varepsilon, \varepsilon)$ eine Potenzreihe ist. Nach Analysis 1 ist

$$e^{tX} = \sum_{k=0}^{\infty} \frac{(tX)^n}{k!} = \lim_{m\to\infty} \sum_{k=0}^{m} \frac{(tX)^n}{k!} =: \lim_{m\to\infty} S_m.$$

Die Zufallsvariable e^{tX} kann also als Grenzwert der Folge (S_m) von Zufallsvariablen geschrieben werden. Um gleich dominierte Konvergenz zu nutzen, brauchen wir einen integrierbare Majorante S für die Folge (S_m). Das ist gar nicht so schwer:

$$|S_m| \overset{\text{Def.}}{=} \Big|\sum_{k=0}^{m} \frac{(tX)^n}{k!}\Big| \overset{\triangle}{\leq} \sum_{k=0}^{m} \Big|\frac{(tX)^n}{k!}\Big| \leq \sum_{k=0}^{\infty} \Big|\frac{(tX)^n}{k!}\Big| =: S.$$

Wegen $S = e^{|tX|} \leq e^{tX} + e^{-tX}$ ist S integrierbar:

$$\mathbb{E}[|S|] = \mathbb{E}[e^{|tX|}] \underset{\text{Lin.}}{\overset{\text{Mon.}}{\leq}} \mathbb{E}[e^{tX}] + \mathbb{E}[e^{-tX}] \overset{\text{Def.}}{=} \mathcal{M}_X(t) + \mathcal{M}_X(-t) < \infty$$

nach Annahme. Jetzt kann also dominierte Konvergenz angewandt werden. Es gilt dann

$$M_X(t) = \mathbb{E}[\lim_{m\to\infty} S_m] \overset{\text{DCT}}{=} \lim_{m\to\infty} \mathbb{E}[S_m] \overset{\text{Lin.}}{=} \lim_{m\to\infty} \sum_{k=0}^{m} \frac{t^k \mathbb{E}[X^k]}{k!} = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}[X^k]}{k!}.$$

Damit ist M_X in $(-\epsilon, \epsilon)$ eine Potenzreihe mit Koeffizienten $a_k = \frac{\mathbb{E}[X^k]}{k!}$.

(b) Aus Analysis 1 wissen wir (so haben wir die Taylor-Koeffizienten bestimmt), dass M_X in $(-\epsilon, \epsilon)$ unendlich oft differenzierbar ist und die Reihe gliedweise differenziert werden kann. n-faches Ableiten gibt dann $M_X^{(n)}(0) = \mathbb{E}[X^n]$.

□

Beispiel 3.3.28.

- Für $X \sim \mathcal{N}(\mu, \sigma^2)$ ergibt die explizite Formel der momenterzeugenden Funktion mit dem Satz

$$\begin{aligned}\mathbb{E}[X] &= \mathcal{M}'_X(0) = \exp\Big(\mu \cdot 0 + \frac{\sigma^2 t^2}{2}\Big) \cdot (\mu + \sigma^2 \cdot 0) = \mu,\\ \mathbb{E}[X^2] &= \mathcal{M}''_X(0) = \mu^2 + \sigma^2.\end{aligned}$$

  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 $X \sim \mathrm{Poi}(\lambda)$ 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 $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda^2 + \lambda - \lambda^2$.

Proposition 3.3.29. (Jensen'sche Ungleichung)
Ist X eine Zufallsvariable mit $\mathbb{E}[|X|] < \infty$ und $\varphi\colon \mathbb{R} \to \mathbb{R}$ *konvex* mit $\mathbb{E}[\varphi(X)] < \infty$, so gilt

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

Beweis. 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:

$$\varphi'(x_0)\mathbb{E}[X] + b = \mathbb{E}[\varphi'(x_0)X] + \mathbb{E}[b] = \mathbb{E}[\varphi'(x_0)X + b] \overset{\text{Mon.}}{\leq} \mathbb{E}[\varphi(x)].$$

Zusammen folgt die Behauptung. □

Beispiel. Als Merkregel für das „≤“ in Proposition 3.4.2 nimmt man $\varphi(x) = x^2$. Weil

$$0 \leq \mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \overset{\text{Üb.}}{=} \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

muss $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$ gelten. Also muss in 3.4.2 „≤“ und nicht „≥“ 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 3.3.30. (Markov- und Tschebyscheff-Ungleichung)
Sei X eine Zufallsvariable und $h\colon \mathbb{R} \to (0, \infty)$ wachsend, dann gelten für $a > 0$

(i)
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[h(X)]}{h(a)} \qquad \text{(Markov-Ungleichung)}$$

(ii)
$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[h(|X|)]}{h(a)} \qquad \text{(Markov-Ungleichung)}$$

(iii)
$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}(X)}{a^2} \qquad \text{(Tschebyscheff-Ungleichung)}$$

Beweis.

(i) Definiere $A = [a, \infty)$, dann gilt weil h wachsend ist

$$\begin{aligned}
\mathbb{E}[h(X)] &\overset{\text{Mon.}}{\geq} \mathbb{E}[h(X) \cdot \mathbf{1}_A(X)] \\
&\geq \mathbb{E}[h(a) \cdot \mathbf{1}_A(X)] \\
&\overset{\text{Lin.}}{=} h(a) \cdot \mathbb{E}[\mathbf{1}_A(X)] \\
&\overset{3.3.23,(iv)}{=} h(a) \cdot \mathbb{P}(X \geq a).
\end{aligned}$$

Durchteilen gibt die Abschätzung.

(ii) Genau wie (i).

(iii) Benutze $h(x) = x^2$ in (ii) mit der Zufallsvariablen $X - \mathbb{E}[X]$.

□

Beispiel. Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so gilt $\mathbb{P}(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$.

Bevor wir wieder in die abstrakte Maßtheorie gehen, noch ein echtes Anwendungsbeispiel von Zufallsvariablen.

Beispiel 3.3.31. Eine Website wird im Mittel pro Stunde zweimal geklickt. Wie groß ist die Wahrscheinlichkeit, dass die Website 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. Sobald wir über Unabhängigkeit sprechen, sehen wir aber, dass die Poissonverteilung sinnvoll ist. Wir nehmen also 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. Um nun die Aufgabe zu lösen, müssen wir für eine Poi(2)-verteilte

Zufallsvariable $\mathbb{P}(X \geq 5)$ berechnen. Das geht ganz einfach:

$$\begin{aligned}\mathbb{P}(X \geq 5) &= 1 - \mathbb{P}(X \leq 4)\\ &= 1 - \sum_{k=0}^{4} \mathbb{P}(X = 4)\\ &= 1 - e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{4}{2} + \frac{8}{6} + \frac{16}{24}\right) \approx 0,053.\end{aligned}$$

Hierbei haben wir genutzt, dass für eine diskrete Zufallsvariable immer $\mathbb{P}(X \in A) = \sum_{k\in A} \mathbb{P}(X = k)$ gilt. Das folgt natürlich aus der σ-Additivität von Maßen weil $\mathbb{P}(X \in A)$ nur die Kurzschreibweis für das Bildmaß (push-forward) $\mathbb{P}_X(A)$ ist.

3.4 L^P-Räume

Sei jetzt wieder $(\Omega, \mathcal{A}, \mu)$ ein beliebiger messbarer Raum und $f\colon \Omega \to \overline{\mathbb{R}}$ 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:

$$\int_\Omega |f| \,\mathrm{d}\mu = 0 \quad\Leftrightarrow\quad |f| = 0 \ \mu\text{-fast überall} \quad\Leftrightarrow\quad f = 0 \ \mu\text{-fast überall.}$$

Satz 3.4.1. (Hölder-Ungleichung)
Seien $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$\int_\Omega |fg| \,\mathrm{d}\mu \leq \Big(\int_\Omega |f|^p \,\mathrm{d}\mu\Big)^{\frac{1}{p}} \Big(\int_\Omega |g|^q \,\mathrm{d}\mu\Big)^{\frac{1}{q}}.$$

Beweis. (i) Alle auftretenden integranden sind messbar und nichtnegativ, also sind alle Integrale definiert, $+\infty = +\infty$ ist aber möglich.

(ii) Wir erinnern an die Young-Ungleichung aus Analysis 2 (das ist gerade die Konvexität des ln): Für $\alpha, \beta > 0$ gilt

$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}.$$

(iii) Ist ein Faktor der rechten Seite 0 oder $+\infty$, so ist nichts zu zeigen. Das ist sofort klar für $+\infty$, aber auch der Fall 0 ist nicht schwer: Wenn nämlich $(\int_\Omega |f|^p \,\mathrm{d}\mu)^{1/p} = 0$ 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.

(iv) Definiere

$$\sigma = \Big(\int_\Omega |f|^p \,\mathrm{d}\mu\Big)^p > 0 \quad\text{und}\quad \tau = \Big(\int_\Omega |g|^q \,\mathrm{d}\mu\Big)^q > 0$$

sowie die messbaren Abbildungen

$$\alpha(\omega) = \frac{|f(\omega)|}{\sigma} \quad\text{und}\quad \beta(\omega) = \frac{|g(\omega)|}{\tau}$$

für alle $\omega \in \Omega$. Mit Young folgt

$$\frac{|f(\omega)g(\omega)|}{\sigma\tau} \leq \frac{|f(\omega)|^p}{\sigma^p p} + \frac{|g(\omega)|^q}{\tau^q q}.$$

Integrieren beider Seiten gibt wegen der Monotonie des Integrals

$$\frac{1}{\sigma\tau}\int_\Omega |fg| \,\mathrm{d}\mu \leq \frac{\int_\Omega |f|^p \,\mathrm{d}\mu}{\sigma^p p} + \frac{\int_\Omega |g|^q \,\mathrm{d}\mu}{\tau^q q} = \frac{1}{p} + \frac{1}{q} = 1.$$

Durchmultiplizieren gibt die Höldersche Ungleichung.

$\square$

Vorlesung 16

Beweis.

(a) Wir behaupten: $\mathcal{M}_X$ ist eine Potenzreihe in $(-\varepsilon,\varepsilon)$. Warum? Nach Analysis 1 ist

$$e^{tX}=\sum_{k=0}^{\infty}\frac{(tX)^n}{k!}=\lim_{m\to\infty}\sum_{k=0}^{m}\frac{(tX)^n}{k!}=:\lim_{m\to\infty}S_m.$$

Es gilt

$$|S_m|\overset{\text{Def.}}{=}\Big|\sum_{k=0}^{m}\frac{(tX)^n}{k!}\Big|\overset{\triangle}{\leq}\sum_{k=0}^{m}\Big|\frac{(tX)^n}{k!}\Big|\leq\sum_{k=0}^{\infty}\Big|\frac{(tX)^n}{k!}\Big|=S.$$

Wegen $e^{|tX|}\leq e^{tx}+e^{-tx}$ gilt

$$\mathbb{E}[S]=\mathbb{E}[e^{|tX|}]\underset{\text{Lin.}}{\overset{\text{Mon.}}{\leq}}\mathbb{E}[e^{tx}]+\mathbb{E}[e^{-tx}]\overset{\text{Def.}}{=}\mathcal{M}_X(t)+\mathcal{M}_X(-t)<\infty\;\forall t\in(-\varepsilon,\varepsilon).$$

Also ist S eine obere Schranke für (S_m); S ist integrierbar. Deshalb kann 3.2.6 angewendet werden:

Inhalt...

□

Proposition 3.4.2. (Jensen'sche Ungleichung)
Ist X eine ZV mit $\mathbb{E}[|X|]<\infty$ und $\phi\colon\mathbb{R}\to\mathbb{R}$ *konvex* mit $\mathbb{E}[\phi(X)]<\infty$, so gilt

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

Beispiel. Sei $\phi(x)=x^2$, dann gilt

$$\mathbb{E}[X]^2\leq\mathbb{E}[X^2].$$

Damit hat man eine Merkregel für das „$\leq$“ in 3.4.2, denn wir wissen für $\mathbb{V}(X)$:

$$0\leq\mathbb{V}(X)=\mathbb{E}[(X-\mathbb{E}[X])^2]\overset{\text{Üb.}}{=}\mathbb{E}[X^2]-\mathbb{E}[X]^2.$$

Beweis. Wir beweisen nur für ein differenzierbares ϕ. Wegen der Konvexität gilt für ein $x_0\in\mathbb{R}\ \exists b\in\mathbb{R}$ mit

- $\phi'(x_0)x+b\leq\phi(x)\ \forall x\in\mathbb{R}$
- $\phi'(x_0)x_0+b\leq\phi(x_0)$

Wähle $x_0=\mathbb{E}[X]$.

$$\begin{aligned}\phi(\mathbb{E}[X])=\phi(x_0)&=\phi'(x_0)x_0+b=\phi'(x_0)\mathbb{E}[X]+b\\=\mathbb{E}[\phi'(x_0)X]+\mathbb{E}[b]&\overset{\text{Lin.}}{=}\mathbb{E}[\phi'(x_0)X+b]\overset{\text{Mon.}}{\leq}\mathbb{E}[\phi(x)].\end{aligned}$$

□

Satz 3.4.3. (Markov- und Tschebyscheff-Ungleichung)
Sei X eine ZV und $h\colon\mathbb{R}\to(0,\infty)$ wachsend, dann gilt für $a>0$

(i)
$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[h(X)]}{h(a)} \quad \text{(Markov)}$$

(ii)
$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[h(|X|)]}{h(a)} \quad \text{(Markov)}$$

(iii)
$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{V}(X)}{a^2} \quad \text{(Tschebyscheff)}$$

Beweis.

(i) Definiere $A = [a, \infty)$, dann gilt

$$\mathbb{E}[h(X)] \overset{\text{Mon.}}{\geq} \mathbb{E}[h(X) \cdot \mathbf{1}_A(X)] \geq \mathbb{E}[h(a) \cdot \mathbf{1}_A(X)] \overset{\text{Lin.}}{=} h(a) \cdot \mathbb{E}[\mathbf{1}_A(X)] = h(a) \cdot \mathbb{P}(X \geq a).$$

(ii) Genau wie (i).

(iii) Benutze $h(X) = x^2$ in (i) mit der ZV $X - \mathbb{E}[X]$.

□

Beispiel. Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so gilt $\mathbb{P}(|x - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$.

Beispiel 3.4.4. Eine Website wird im Mittel pro Stunde zweimal geklickt. Wie groß ist die Wahrscheinlichkeit, dass die Website in einer Stunde fünfmal geklickt wird?
Unsere ZV ist hier $X = \#\text{Klicks in einer Stunde}$, eine sinnvolle Verteilung für X wäre $X \sim \text{Poi}(\lambda)$, d. h. $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$. Dann wäre $\lambda = 2$, denn der EW der Poisson-Verteilung gerade λ ist. Dann wissen wir

$$\begin{aligned}\mathbb{P}(X \geq 5) = 1 - \mathbb{P}(X \leq 4) &\overset{\sigma\text{-Add.}}{=} 1 - \sum_{k=0}^{4} \mathbb{P}(X = 4) \\ &= 1 - e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{4}{2} + \frac{8}{6} + \frac{16}{24}\right) \approx 0{,}053.\end{aligned}$$

3.5 L^P Räume

Sei $(\Omega, \mathcal{A}, \mu)$ ein messbarer Raum und $f\colon \Omega \to \mathbb{R}$ sein $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$-messbar. Erinnerung: Wegen 3.1.15 gilt $\int_\Omega |f| \,\mathrm{d}\mu = 0 \Leftrightarrow f = 0$ μ-fast überall $\Leftrightarrow |f| = 0$ μ-fast überall.

Satz 3.5.1. (Hölder-Ungleichung)
Seien $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$\int_\Omega |fg| \,\mathrm{d}\mu \leq \Big(\int_\Omega |f|^p \,\mathrm{d}\mu\Big)^{\frac{1}{p}} \Big(\int_\Omega |g|^q \,\mathrm{d}\mu\Big)^{\frac{1}{q}}.$$

Beide Seiten können den Wert $+\infty$ annehmen.

Beweis. (i) Alle auftretenden Integrale sind messbar und nichtnegativ, also sind alle Integrale definiert, $+\infty = +\infty$ ist aber möglich.

(ii) Erinnerung an die Young-Ungleichung aus Analysis 2 (das ist gerade die Konvexität des ln):
$$\alpha\beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q},\ \alpha, \beta > 0$$

(iii) Ist ein Faktor der rechten Seite 0 oder $+\infty$, so ist nichts zu zeigen. Das ist klar für $+\infty$, wenn $(\int_\Omega |f|^p \,\mathrm{d}\mu)^{1/p}) = 0$, so ist $f = 0$ μ-fast überall, also ist $|fg| = 0$ μ-fast überall. Dann ist auch die linke Seite Null.

(iv) Definiere
$$\sigma = \Big(\int_\Omega |f|^p \,\mathrm{d}\mu\Big)^p > 0,\ \tau = \Big(\int_\Omega |g|^q \,\mathrm{d}\mu\Big)^q > 0$$
und
$$\alpha(\omega) = \left(\frac{|f(\omega)|}{\sigma}\right),\ \beta(\omega) = \left(\frac{|g(\omega)|}{\tau}\right).$$
Mit Young folgt
$$\frac{|fg|}{\sigma\tau} \leq \frac{|f|^p}{\sigma^p p} + \frac{|g|^q}{\tau^q q}.$$
Integrieren gibt
$$\frac{1}{\sigma\tau}\int_\Omega |fg| \,\mathrm{d}\mu \leq \frac{\int_\Omega |f|^p \,\mathrm{d}\mu}{\sigma^p p} + \frac{\int_\Omega |g|^q \,\mathrm{d}\mu}{\tau^q q} = \frac{1}{p} + \frac{1}{q} = 1.$$

□

Vorlesung 17

Korollar 3.5.2.

(i) Der Fall $p = q = 2$ heißtauch Cauchy-Schwarz:

$$\left(\int_\Omega |fg| \,\mathrm{d}\mu\right)^2 \leq \int_\Omega |f|^2 \,\mathrm{d}\mu \int_\Omega |f|^2 \,\mathrm{d}\mu.$$

(ii) Ist μ ein Wahrscheinlichkeitsmaß, so gilt wegen $\mu(\Omega) = 1$

$$\begin{aligned}\int_\Omega |f| \,\mathrm{d}\mu &= \int_\Omega |f \cdot 1| \,\mathrm{d}\mu \\ &\leq \left(\int_\Omega |f|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} \left(\int_\Omega \underbrace{|1|^q}_{=1\cdot\mathbf{1}_\Omega} \,\mathrm{d}\mu\right)^{\frac{1}{q}} \\ &= \left(\int_\Omega |f|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} \cdot 1,\end{aligned}$$

also

$$\left(\int_\Omega |f| \,\mathrm{d}\mu\right)^p \leq \int_\Omega |f|^p \,\mathrm{d}\mu$$

für alle $p > 1$.

Satz 3.5.3. (Minkowski-Ungleichung)
Sei $p \geq 1$, so gilt

$$\left(\int_\Omega |f+g|^p \,\mathrm{d}\mu\right)^p \leq \left(\int_\Omega |f|^p \,\mathrm{d}\mu\right)^p + \left(\int_\Omega |g|^p \,\mathrm{d}\mu\right)^p.$$

Beide Seiten können den Wert $+\infty$ annehmen.

Beweis. Wie in Analysis 2, folgt aus Hölder und der Young Ungleichung. Wir zeigen die stärkere Ungleichung

$$\left(\int_\Omega (|f|+|g|)^p \,\mathrm{d}\mu\right)^p \leq \left(\int_\Omega |f|^p \,\mathrm{d}\mu\right)^p + \left(\int_\Omega |g|^p \,\mathrm{d}\mu\right)^p. \tag{3.3}$$

Tatsächlich impliziert (3.3) Minkowski weil die linke Seite von Minkowski kleiner ist als die linke Seite von (3.3). 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 (3.3) natürlich mit Gleichheit.

(b) Sei nun $p > 1$. Ist die rechte Seite $+\infty$, so gilt (3.3). 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 (3.3):

$$\begin{aligned}
\int_\Omega (|f|+|g|)^p \,\mathrm{d}\mu &= \int_\Omega (|f|+|g|)^{p-1}(|f|+|g|)\,\mathrm{d}\mu \\
&\overset{\text{ausm.}}{\underset{\text{+lin.}}{=}} \int_\Omega (|f|+|g|)^{p-1}|f|\,\mathrm{d}\mu + \int_\Omega (|f|+|g|)^{p-1}|g|\,\mathrm{d}\mu \\
&\overset{2\times}{\underset{\text{Hölder}}{\leq}} \left(\int_\Omega |f|^p\,\mathrm{d}\mu\right)^{\frac{1}{p}} \left(\int_\Omega (|f|+|g|)^{(p-1)q}\,\mathrm{d}\mu\right)^{\frac{1}{q}} \\
&\quad + \left(\int_\Omega |g|^p\,\mathrm{d}\mu\right)^{\frac{1}{p}} \left(\int_\Omega (|f|+|g|)^{(p-1)q}\,\mathrm{d}\mu\right)^{\frac{1}{q}} \\
&\overset{\text{auskl.}}{=} \left(\left(\int_\Omega |f|^p\,\mathrm{d}\mu\right)^{\frac{1}{p}} + \left(\int_\Omega |g|^p\,\mathrm{d}\mu\right)^{\frac{1}{p}}\right) \left(\int_\Omega |f|+|g|\,\mathrm{d}\mu\right)^{1-\frac{1}{p}}.
\end{aligned}$$

In der letzten Gleichung haben wir genutzt, dass $1-\frac{1}{p}=\frac{1}{q}$ und $(p-1)q=1$ aufgrund der Voraussetzung an p und q gelten. Rübermultiplizieren des zweiten Faktors gibt dann (3.3).

$\square$

Definition 3.5.4. $f\colon \Omega \to \overline{\mathbb{R}}$ heißt p-**fach integrierbar**, falls $\int_\Omega |f|^p\,\mathrm{d}\mu < \infty$. Statt 2-fach integrierbar sagt man **quadratintegrierbar**, statt 1-fach integrierbar sagt man **integrierbar**. Natürlich passt die Notation zu unserer ursprünglichen Definition der Integrierbarkeit:

$$f \text{ int.} \overset{\text{alte}}{\underset{\text{Def.}}{\Leftrightarrow}} \int_\Omega f^+\,\mathrm{d}\mu < \infty,\ \int_\Omega f^-\,\mathrm{d}\mu < \infty \overset{\text{Übung}}{\Leftrightarrow} \int_\Omega |f|\,\mathrm{d}\mu < \infty \overset{\text{neue}}{\underset{\text{Def.}}{\Leftrightarrow}} f \text{ int.}$$

Schauen wir uns ein ganz konkretes Beispiel für die neue Definition an.

Beispiel 3.5.5.

(i) Für $\Omega=[1,\infty)$, $\mathcal{A}=\mathcal{B}([1,\infty))$, $\mu=\lambda_{|[1,\infty)}$ ist $\int_\Omega f\,\mathrm{d}\mu=\int_1^\infty f\,\mathrm{d}\mu$. Für $f(x)=\frac{1}{x^a}$ ist f p-fach integrierbar genau dann, wenn $p>\frac{1}{a}$.

(ii) Für $\Omega=[0,1]$, $\mathcal{A}=\mathcal{B}([0,1])$, $\mu=\lambda_{|[0,1]}$ ist $\int_\Omega f\,\mathrm{d}\mu=\int_0^1 f\,\mathrm{d}\mu$. Für

$$f(x)=\begin{cases}\frac{1}{x^a} & : x\in(0,1]\\ +\infty & : x=0\end{cases}$$

ist f p-fach integrierbar genau dann wenn $p<\frac{1}{a}$. 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 3.5.6.

$$\mathcal{L}^p(\mu) := \left\{ f\colon \Omega\to\mathbb{R} \;\middle|\; \int_\Omega |f|^p\,\mathrm{d}\mu < \infty \right\}.$$

Manachmal schreibt man auch $\mathcal{L}^p(\Omega,\mathcal{A},\mu)$.

Lemma 3.5.7. Mit punktweiser Addition und Skalarmultiplikation ist $\mathcal{L}^p(\mu)$ ein reeller Vektorraum, sogar ein Untervektorraum der messbaren Funktionen, d. h.

(i) $f, g \in \mathcal{L}^p(\mu) \Rightarrow f + g \in \mathcal{L}^p(\mu)$,

(ii) $\alpha \in \mathbb{R}$, $f \in \mathcal{L}^p(\mu) \Rightarrow \alpha f \in \mathcal{L}^p(\mu)$,

(iii) $0 \in \mathcal{L}^p(\mu)$, wobei 0 die konstante Nullfunktion ist.

Beweis. Messbare Funktionen mit punktweiser Addition und skalarer Multiplikation geben einen Vektorraum. Die Eigenschaften (i)-(iii) bedeuten, dass $\mathcal{L}^p(\mu)$ ein Untervektorraum ist. Wir prüfen also nur die dafür benötigten Eigenschaften:

(i) $\int_\Omega |0|^p \,\mathrm{d}\mu = 0 < \infty$

(ii) αf ist messbar und $\int_\Omega |\alpha f|^p \,\mathrm{d}\mu \overset{\text{Lin.}}{=} |\alpha|^p \int_\Omega |f|^p \,\mathrm{d}\mu < \infty$ weil $f \in \mathcal{L}^p(\mu)$ angenommen wurde. Also gilt auch $\alpha f \in \mathcal{L}^p(\mu)$.

(iii) $f + g$ ist messbar und wegen Minkowski gilt

$$\left(\int_\Omega |f+g|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} \leq \left(\Big(\underbrace{\int_\Omega |f|^p \,\mathrm{d}\mu}_{<\infty}\Big)^{\frac{1}{p}} + \Big(\underbrace{\int_\Omega |g|^p \,\mathrm{d}\mu}_{<\infty}\Big)^{\frac{1}{p}}\right)^p < \infty.$$

Also ist $f + g \in \mathcal{L}^p(\mu)$.

□

Lemma 3.5.8.

$$||f||_p = \Big(\int_\Omega |f|^p \,\mathrm{d}\mu\Big)^{\frac{1}{p}}$$

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) $0 \leq ||f||_p < \infty$ (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! □

Warnung: $|| \cdot ||_p$ ist *keine* Norm! Jedes f mit $\mu(\{f \neq 0\}) = 0$ erfüllt

$$||f||_p = \Big(\int_\Omega |f|^p \,\mathrm{d}\mu\Big)^{\frac{1}{p}} = 0.$$

Wenn f auf einer μ-Nullmenge ungleich 0 ist, so ist $f \neq 0$. Also ist die Definitheit nicht erfüllt.

Definition 3.5.9. Seien $f, f_1, f_2, ... \in \mathcal{L}^p(\mu)$. Wir sagen $(f_n)_{n\in\mathbb{N}}$ konvergiert in $\mathcal{L}^p(\mu)$ gegen f, falls $||f_n - v||_p \to 0$, $n \to \infty$. Man schreibt dann auch $f_n \overset{\mathcal{L}^p(\mu)}{\longrightarrow} f$, $n \to \infty$.

Bemerkung 3.5.10. Grenzwerte von Folgen in $\mathcal{L}^p(\mu)$ sind nicht eindeutig. Es ist also möglich, dass $f_n \overset{\mathcal{L}^p(\mu)}{\longrightarrow} f$, $n \to \infty$ und $f_n \overset{\mathcal{L}^p(\mu)}{\longrightarrow} g$, $n \to \infty$, aber $f \neq g$. $\rightsquigarrow$ Übungsaufgabe.

Definition 3.5.11. $(f_n)_{n\in\mathbb{N}} \subseteq \mathcal{L}^p(\mu)$ heißt Cauchyfolge in $\mathcal{L}^p(\mu)$, falls

$$\forall \varepsilon > 0\, \exists N \in \mathbb{N}\colon ||f_n - f_m||_p < \varepsilon\, \forall n, m \geq N.$$

Satz 3.5.12. Jede Cauchyfolge in $\mathcal{L}^p(\mu)$ hat einen Grenzwert in $\mathcal{L}^p(\mu)$.

$\mathcal{L}^p(\mu)$ ist sozusagen vollständig, abgesehen davon, dass $\mathcal{L}^p(\mu)$ kein normierter Vektorraum ist.

Beweis. Sei $(f_n)_{n\in\mathbb{N}}$ eine Cauchyfolge in $\mathcal{L}^p(\mu)$. Wir müssen ein $f \in \mathcal{L}^p(\mu)$ finden, so dass $f_n \overset{\mathcal{L}^p(\mu)}{\longrightarrow} f$, $n \to \infty$. Wir basteln uns zunächst ein f und zeigen dann, dass dieses f die zwei Eigenschaften erfüllt. Weil $(f_n)_{n\in\mathbb{N}}$ eine Cauchyfolge ist, existiert eine Teilfolge $(f_{n_k})_{k\in\mathbb{N}}$ mit $||f_{n_{k+1}} - f_{n_k}|| \leq \frac{1}{2^k}$ für alle $k \in \mathbb{N}$. Damit definierten wir

$$g_k = f_{n_{k+1}} - f_{n_k} \quad \text{und} \quad g = \sum_{k=1}^{\infty} |g_k|.$$

Wie die $\triangle$-Ungleichung in $\mathbb{R}$ gilt auch die $\triangle$-Ungleichung in $\mathcal{L}^p(\mu)$ für ∞ viele Summanden $\rightsquigarrow$ Übungsaufgabe. Daher gilt

$$||g||_p \overset{\text{Def.}}{=} \Big|\Big| \sum_{k=1}^{\infty} |g_k| \Big|\Big|_p \leq \sum_{k=1}^{\infty} ||g_k||_p = \sum_{k=1}^{\infty} ||f_{n_{k+1}} - f_{n_k}||_p \leq \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

Also ist $g \in \mathcal{L}^p(\mu)$ und damit ist g μ-fast überall endlich. Die Reihe $\sum_{k=1}^{\infty} g_k(\omega)$ ist also für μ-fast alle $\omega \in \Omega$ absolut konvergent. Nach Analysis 1 ist also auch die Reihe $\sum_{k=1}^{\infty} g_k(\omega)$ für μ-fast alle $\omega \in \Omega$ konvergent. Jetzt definieren wir uns ein f:

$$f(\omega) := \begin{cases} f_{n_1}(\omega) + \sum\limits_{k=1}^{\infty} g_k(\omega) & : \sum\limits_{k=1}^{\infty} g_k(\omega) \text{ konvergent} \\ 0 & : \text{ sonst} \end{cases}.$$

Wir zeigen jetzt, dass $f \in \mathcal{L}^p(\mu)$ und $f_n \overset{\mathcal{L}^p(\mu)}{\longrightarrow} g$, $n \to \infty$.

(i)

$$|f| \overset{\triangle}{\leq} |f_{n_1}| + |\sum_{k=1}^{\infty} g_k| \overset{\triangle}{\leq} \underbrace{\underbrace{|f_{n_1}|}_{\in\mathcal{L}^p(\mu)} + \underbrace{\sum_{k=1}^{\infty} |g_k|}_{\in\mathcal{L}^p(\mu)}}_{\in\mathcal{L}^p(\mu)}$$

(ii) Wegen Teleskop gilt

$$f_{n_k} \overset{\text{Tel.}}{=} f_{n_1} + \sum_{l=1}^{k-1} g_l \to f,\ n \to \infty\ \mu\text{-f. ü.}$$

$$|f_{n_k} - f|^p \leq (|f_{n_k}| + |f|)^p \overset{\text{Def.}}{\triangle} \Big(|f_{n_k}| + \sum_{l=1}^{k-1} |g_k| + |f|\Big)$$
$$\leq (|f_{n_1}| + g + |f|) =: h^p \quad \text{(unabhängig von k)}$$

h ist integrierbare Majorante, daher mit 3.2.6.

$$\lim_{k\to\infty} ||f_{n_k} - f||_p \overset{\text{Def.}}{=} \lim_{k\to\infty} \left(\int_\Omega |f_{n_k} - f|^p\right)^{\frac{1}{p}} \overset{3.2.6}{=} \left(\int_\Omega \underbrace{\lim_{k\to\infty} |f_{n_k} - f|^p}_{0\ \mu\text{-f. ü.}}\right)^{\frac{1}{p}} = 0.$$

Also gilt $f_{n_k} \overset{\mathcal{L}^p(\mu)}{\longrightarrow} f$, $k \to \infty$. Teilfolge loswerden:

$$||f_n - f||_p \overset{\triangle}{=} ||f_n - f_{n_k}||_p - ||f_{n_k} - f||_p \to 0,\ n \to \infty.$$

□

Wenn $\mathcal{L}^p(\mu)$ ein normierter Vektorraum wäre, könnten wir jetzt wunschlos glücklich sein. Denn dann wäre $\mathcal{L}^p(\mu)$ ein Banachraum. Wir müssen aber noch den Defekt der Definitheit der Halbnorm reparieren. Dazu habt ihr in der Linearen Algebra das Konzept der Quotientenräume kennengelernt. Das nutzen wir hier aus, indem wir einfach alle Abbildungen die μ-fast überall 0 sind, als gleiches Objekt identifizieren. Dazu wählen wir die Äquivalenzrelation

$$f \sim g \quad :\Leftrightarrow \quad f = g\ \mu\text{-fast überall}$$

und den Quotientenraum, der aus den Äquivalenzklassen

$$[f]_\sim := \{g \in \mathcal{L}^p(\mu) : f \sim g\} = \{g \in \mathcal{L}^p(\mu) : f = g\ \mu\text{-fast überall}\}$$

besteht. Wie in der linearen Algebra werden die Vektorraumoperationen auf dem Quotientenraum durch die Repräsentanten definiert:

$$[f] + [g] := [f + g] \quad \text{und} \quad \alpha[f] := [\alpha f].$$

Genauso verfahren wir mit der Norm: Wir definieren auch noch eine Norm:

$$||[f]||_p := ||f||_p.$$

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 Banachraum:

Satz 3.5.13. $L^p(\mu)$ ist mit gerade definierten Operationen ein Banachraum.

Beweis. Wir zeigen, dass $L^p(\mu)$ mit den definierten Operationen ein normierter Vektorraum ist und dass dieser vollständig 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:

$$\begin{aligned} ||[f]||_p = 0 &\overset{\text{Def.}}{\Leftrightarrow} ||f||_p = 0 \\ &\overset{\text{Def.}}{\Leftrightarrow} \int_\Omega |f|^p \,\mathrm{d}\mu = 0 \\ &\Leftrightarrow |f|^p = 0\ \mu\text{-fast überall} \\ &\Leftrightarrow f = 0\ \mu\text{-fast überall} \\ &\Leftrightarrow [f] = [0] \end{aligned}$$

Damit ist $||\cdot||_p$ eine Norm.

(c) Die Vollständigkeit folgt direkt aus der Definition der Norm und Satz 3.5.12 weil

$$||[f_n] - [f_m]||_p = ||[f_n - f_m]|| \overset{\text{Def.}}{=} ||f_n - f_m||_p$$

gilt.

$\square$

Vorlesung 18

3.6 Produktmaße und Fubini

Ziel: Konstruiere ein Modell, um unabhängige ZV zu konstruieren. Im Folgenden seien $(\Omega_1, \mathcal{A}_1, \mu_1)$, $(\Omega_2, \mathcal{A}_2, \mu_2)$ Maßräume und $\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) \colon \omega_1 \in \Omega_1,\ \omega_2 \in \Omega_2\}$.

Definition 3.6.1.

(i) $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\{(A_1, A_2) \colon A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2\})$ heißt **Produkt-σ-Algebra** auf $\Omega_1 \times \Omega_2$

(ii) Ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ heißt **Produktmaß**, falls $\mu((A_1, A_2)) = \mu_1(A_1) \cdot \mu_1(A_2)$.

Beispiel. $\Omega_1 = \Omega_2 = \mathbb{R}$, $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{B}(\mathbb{R})$, $\mu_1 = \mu_2 = \lambda$. Produktmaß: $\mu = \mu_1 \otimes \mu_2$ ist das 2-dimensionale Lebesguemaß. Weil Quader $Q = (a_1, b_1] \times (a_2, b_2] \in \mathcal{A}_1 \times \mathcal{A}_2$ muss für μ gelten $\mu(Q) = \lambda((a_1, b_1]) \cdot \lambda((a_2, b_2]) = (b_1 - a_1)(b_1 - a_1) =$ „Fläche von Q“.

Satz 3.6.2. (Konstruktion Produktmaß)
Sind μ_1, μ_2 σ-endlich, so existiert ein endliches Maß $\mu = \mu_1 \otimes \mu_2$ auf $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ mit $\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2)\ \forall A_1 \in \mathcal{A}_1,\ A_2 \in \mathcal{A}_2$.

Der „einfache“ Beweis wäre, die Voraussetzungen von 1.3.12 zu checken. Wir machen einen anderen, aus dem Fubini sofort folgt:

Beweis. Eindeutigkeit: Aus Dynkin-System, 1.2.13. $\mathcal{S} = \{A_1 \times A_2 \colon A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$ ist $\cap$-stabiler Erzeuger von $\mathcal{A}_1 \otimes \mathcal{A}_2$. Weil μ_1, μ_2 σ-endlich 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 $\mu(E_n^1) < \infty$, $\mu(E_n^2) < \infty\ \forall n \in \mathbb{N}$. Sei $E_n = E_n^1 \times E_n^2$, dann gelten

- $E_n \uparrow \Omega$
- $\mu(E_n) \overset{\text{Ann.}}{=} \mu_1(E_n^1) \cdot \mu_2(E_n^2) < \infty$.

Damit kann 1.2.13 angewendet werden und die Eindeutigkeit folgt.
Existenz:

$$\overline{\mu} := \int_{\Omega_2} \mu_1(A_{\omega_2})\, \mathrm{d}\mu_2,$$

wobei $A_{\omega_2} = \{\omega \in \Omega_2 \colon (\omega_1, \omega_2) \in \mathcal{A}\}$. Alternativ:

$$\mu := \int_{\Omega_1} \mu_2(A_{\omega_1})\, \mathrm{d}\mu_1.$$

Wir zeigen: mu ist ein Produktmaß auf $\mathcal{A}_1 \times \mathcal{A}_2$. Derselbe Beweis zeigt, dass $\overline{\mu}$ ein Produktmaß ist ($\Rightarrow \mu = \overline{\mu}$ wegen der Eindeutigkeit).

(i) $mu \colon \mathcal{A} \to [0, \infty]$ weil μ_2 ein Maß ist (deshalb nichtnegativ) und Integrale über nichtnegative Funktionen nichtnegativ sind.

(ii) $\mu(\emptyset) = 0$ ✓ weil A_{ω_1} auch leere Menge ist

(iii) σ-Additivität: Seien dazu $A^1, A^2, ... \in \mathcal{A}$ paarweise disjunkt. Sei

$$A := \dot{\bigcup_{n=1}^{infty}} A_n,$$

dann ist

$$A_{\omega_1} = \dot{\bigcup_{n=1}^{\infty}} A^n_{\omega_1},$$

$$\begin{aligned}\mu\Big(\dot{\bigcup_{n=1}^{\infty}} A^n_{\omega_1}\Big) &\overset{\text{Def.}}{=} \int_{\Omega_1} \mu_2\Big(\Big(\dot{\bigcup_{n=1}^{\infty}} A^n_{\omega_1}\Big)_{\omega_1}\Big)\,\mathrm{d}\mu \\ = \int_{\Omega_1} \mu_2\Big(\dot{\bigcup_{n=1}^{\infty}} A^n_{\omega_1}\Big)\,\mathrm{d}\mu &\overset{\mu_2\ \sigma\text{-endlich}}{=} \int_{\Omega_1} \sum_{n=1}^{\infty} \mu_2(A^n_{\omega_1})\,\mathrm{d}\mu \\ &\overset{3.2.1}{=} \sum_{n=1}^{\infty} \int_{\Omega_1} \mu_2(A^n_{\omega_1})\,\mathrm{d}\mu \overset{\text{Def.}}{=} \sum_{n=1}^{\infty} \mu(A^n).\end{aligned}$$

Produktmaß: $A = A_1 \times A_2$. Weil

$$(A_1, A_2)_{\omega_1} = \begin{cases} \emptyset & ,\omega_1 \notin A_1 \\ A_2 & ,\omega_1 \in A_1 \end{cases}$$

folgt

$$\begin{aligned}\mu(A) &= \int_{\Omega_1} \mu_2(A)\,\mathrm{d}\mu(\omega_1) = \int_{\Omega_1} \mu_2(A_2)\mathbf{1}_{A_1}(\omega_1)\,\mathrm{d}\mu(\omega_1) \\ &\overset{\text{Lin.}}{=} \mu_2(A_2)\int_{\Omega_1} \mathbf{1}_{A_1}(\omega_1)\,\mathrm{d}\mu(\omega_1) = \mu_2(A_2)\cdot\mu_1(A_1)\end{aligned}$$

Damit μ sinnvoll definiert ist, müssen wir zwei Sachen zeigen

(i) $A_{\omega_1} \in A_2\ \forall A \in \mathcal{A}$

(ii) $\omega_1 \to \mu_2(A_{\omega_1})$ muss messbar sein.

zu (i): Sei $\mathcal{F} = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 \colon A_{\omega_1} \in \mathcal{A}_2\}$. Wir zeigen: $\mathcal{F} = \mathcal{A}_1 \otimes \mathcal{A}_2 = \mathcal{A}$. $\mathcal{F} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$ klar, es fehlt noch $\mathcal{A}_1 \otimes \mathcal{A}_2 \subseteq \mathcal{F}$. Wir behaupten, $\mathcal{F}$ ist eine σ-Algebra.

(i) $\Omega \in \mathcal{F}$, weil $\Omega_{\omega_1} = \Omega_2 \in \mathcal{F}$

(ii) Sei $A \in \mathcal{F}$, dann ist $A^C \in \mathcal{F}$, weil $(A^C)_{\omega_1} = (A_{\omega_1})^C \in \mathcal{A}_2$

(iii) Genauso mit abzählbaren Vereinigungen:

$$\Big(\dot{\bigcup_{n=1}^{\infty}} A^n\Big)_{\omega_1} = \dot{\bigcup_{n=1}^{\infty}} \underbrace{A^n_{\omega_1}}_{\in\mathcal{A}_2} \in \mathcal{A}_2$$

Weil $\mathcal{S} \subseteq \mathcal{F}$ und $\sigma(\mathcal{S}) = \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$, ist $\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{S}) \subseteq \sigma(\mathcal{F}) = \mathcal{F} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$. Also folgt Gleichheit.

Zu (ii): 1. Schritt: $\mu_2(\Omega_2) < \infty$: Wir zeigen $\mathcal{A}_1 \otimes \mathcal{A}_2 \subseteq \mathcal{F} := \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : \omega_1 \to \mu_2(A_{\omega_1}) \text{ messbar}\}$. Weil „$\supseteq$“ per Definition gilt, wären wir fertig. Es gilt $\mathcal{S} \subseteq \mathcal{F}$ weil

$$\mu_2((A_1 \times A_2)_{\omega_1}) = \underbrace{\mu_2(A_2)\, \underbrace{\mathbf{1}_{A_1}(\omega_1)}_{\text{messbar}}}_{\text{messbar}}.$$

$\mathcal{F}$ ist ein Dynkin-System.

- $\Omega \in \mathcal{F}$ klar, weil $\omega_1 \to \mu_2(\Omega_{\omega_1}) \equiv \mu_2(\Omega_2)$
- Sei $A \in \mathcal{F}$, dann gilt

$$\omega_1 \to \mu_2((A^C)_{\omega_1}) = \mu_2(A^C_{\omega_1}) \overset{\text{Maß}}{=} \underbrace{\mu_2(\Omega_2)\, \underbrace{\mu_2(A_{\omega_1})}_{\text{messbar}}}_{\text{messbar}}.$$

  Also gilt auch $A^C \in \mathcal{F}$.
- Sind $A^1, A^2, ...$ paarweise disjunkt.

$$\begin{aligned}
\omega_1 \to \mu_2\Big(\Big(\dot{\bigcup_{n=1}^{\infty}} A^n\Big)_{\omega_1}\Big) &= \mu_2\Big(\dot{\bigcup_{n=1}^{\infty}} A^n_{\omega_1}\Big) \\
&\overset{\sigma\text{-add.}}{=} \sum_{n=1}^{\infty} \underbrace{\mu(A^n_{\omega_1})}_{\text{messbar}} \quad (\text{weil } A^n_{\omega_1} \text{ nach Annahme messbar}) \\
&= \underbrace{\lim_{m\to\infty} \underbrace{\sum_{n=1}^{m} \mu(A^n_{\omega_1})}_{\text{messbar}}}_{\text{messbar}} \quad (\text{Grenzwerte messb. Funktionen messb.})
\end{aligned}$$

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \overset{\text{Def.}}{=} \sigma(\mathcal{S}) = d(\mathcal{S}) \subseteq d(\mathcal{F}) = \mathcal{F} \subseteq \mathcal{A}_1 \otimes \mathcal{A}_2$$
$$\Rightarrow \mathcal{F} = \mathcal{A}_1 \otimes \mathcal{A}_2.$$

□

Korollar 3.6.3. Sind $(\Omega_i, \mathcal{A}_i, \mu_i)_{i=1,...,n}$ σ-endliche Maßräume, so existiert genau ein Maß $\mu_1 \otimes ... \otimes \mu_n$ auf der Produkt-σ-Algebra $\mathcal{A} = \mathcal{A}_1 \otimes ... \otimes \mathcal{A}_n$ auf $\Omega_1 \times ...\Omega_n = \sigma(\{A_1 \times ... \times A_n : A_i \in \mathcal{A}_i\})$ mit $\mu((A_1 \times ... \times A_n)) = \mu_1(A_1) \cdot ... \cdot \mu_n(A_n)$.

Beweis. Induktion. □

Definition. Sind alle $(\Omega_i, \mathcal{A}_i, \mu_i)$ gleich, so schreibt man $\mu^{\otimes n}$ statt $\mu \otimes ... \otimes \mu$. $\mu^{\otimes n}$ heißt **n-faches Produktmaß** von μ.