
Stochastik

LEIF DÖRING¹

UNIVERSITÄT MANNHEIM

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

Inhaltsverzeichnis

1	Maßtheorie (Modellierung von Ereignissen und Wahrscheinlichkeiten)	3
1.1	σ -Algebren und Maße - die Grundbegriffe der Stochastik	3
1.2	Erzeuger von σ -Algebren und Dynkin-Systeme	9
1.3	Konstruktion von Maßen	16
1.4	Das Beispiel für Wahrscheinlichkeitsmaße	24
2	Abbildungen zwischen messbaren Räumen	34
2.1	Messbare Abbildungen	34
2.2	Bildmaße oder „push-forward“ eines Maßes	36
2.3	Messbare numerische Funktionen	37
3	Convergence of Measures	40
3.1	A bit of topology, measure, and integration theory	40
3.2	Weak convergence of measures - the basics	46
3.3	Relative sequential compactness of measures	51
3.4	Identification of probability laws on \mathbb{R}	55
3.5	Weak convergence on $\mathcal{B}(\mathbb{R})$	67
3.6	Applications	72
4	Brownian Motion	78
4.1	Stochastic processes in continuous time	78
4.1.1	Stochastic processes and the generic path space	78
4.1.2	Stochastic processes with continuous sample paths	91
4.2	Foundation of Brownian motion	100
4.3	Weak convergence in $C([0, \infty))$	111
4.4	Donsker’s invariance principle	116

Was machen wir, was nicht?

„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



Wahrscheinlichkeitstheorie (\rightsquigarrow Döring)



Anpassung der Modelle auf „echte“ zufällige Experimente



Mathematische Statistik (\rightsquigarrow Schlather)



Ausführung der Modelle („Zufall erzeugen“)



Stochastische Numerik (\rightsquigarrow Neuenkirch)



Anwendung auf Finanzmärkte



Finanzmathematik (\rightsquigarrow Prömel)



Anwendung auf Wirtschaftsdaten



Ökonometrie (\rightsquigarrow Trenkler, Rothe)



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



Kombinatorik (in Mannheim nicht vertreten)

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

Teil 1: Maß- und Integrationstheorie

Kapitel 1

Maßtheorie (Modellierung von Ereignissen und Wahrscheinlichkeiten)

Vorlesung 1

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

1.1 σ -Algebren und Maße - die Grundbegriffe der Stochastik

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

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

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

Die Stabilität für Vereinigungen gilt auch, wenn man nur endlich viele messbare Mengen vereinigen will. Dazu nutzt man einfach folgenden Trick: Wenn man messbare Mengen A_1, \dots, A_N vereinigen möchte, so setzt man $A_{N+1} = A_{N+2} = \dots = \emptyset$ und beachtet, dass damit wegen der σ -Additivität $\bigcup_{k=1}^N A_k = \bigcup_{k=1}^\infty A_k \in \mathcal{A}$ gilt. Merkt euch solche kleinen Tricks, der selbe Trick taucht gleich nochmal auf.

Example 1.1.2. Ist $\Omega \neq \emptyset$ eine beliebige Grundmenge, so sind folgende Mengensysteme σ -Algebren:

- $\mathcal{A}_1 = \{\emptyset, \Omega\}$
- $\mathcal{A}_2 = \mathcal{P}(\Omega)$
- $\mathcal{A}_3 = \{\emptyset, \Omega, A, A^C\}$ für $A \subseteq \Omega$ beliebig
- $\mathcal{A}_4 = \{A \subseteq \Omega \mid A \text{ oder } A^C \text{ ist abzählbar}\}$

Da $\mathcal{A}_1, \dots, \mathcal{A}_4$ Teilmengen der Potenzmenge sind, muss man jeweils nur die drei definierenden Eigenschaften einer σ -Algebra testen. Bei den ersten drei Beispielen ist das direkt, indem man alle Möglichkeiten ausprobiert. Im vierten Beispiel müssen wir nur bei der abzählbaren Vereinigung kurz nachdenken. Seien also A_1, A_2, \dots Teilmengen von Ω , die entweder abzählbar sind oder deren Komplemente abzählbar sind. Sind all diese Mengen abzählbar, so ist nach Analysis 1 auch die Vereinigung abzählbar, also ist die Vereinigung wieder in \mathcal{A}_3 . Ist eine der Mengen nicht abzählbar, sagen wir A_j , so ist das Komplement A_j^C abzählbar. Doch dann ist wegen

$$\left(\bigcup_{k=1}^\infty A_i \right)^C \stackrel{\text{de Morgan}}{=} \bigcap_{k=1}^\infty A_i^C \subseteq A_j^C$$

das Komplement der Vereinigung nach Analysis 1 abzählbar. Also ist die Vereinigung in \mathcal{A}_4 und damit \mathcal{A}_4 abgeschlossen bezüglich Vereinigungen.

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

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

Beweis. Die Strategie ist immer gleich: Man versucht die Behauptung aus den drei Regeln einer σ -Algebra herzuleiten. Da $\emptyset = \Omega^C$ gilt, gilt wegen den Eigenschaften (i) und (ii) einer σ -Algebra auch $\emptyset \in \mathcal{A}$. Mit de Morgan und den Eigenschaften (ii), (iii) der σ -Algebra gilt

$$\bigcap_{k=1}^\infty A_k = \left(\bigcup_{\substack{k=1 \\ \in \mathcal{A}, \text{ (ii)}}}^\infty A_k^C \right)^C. \quad \begin{array}{c} \underbrace{\phantom{\bigcup_{k=1}^\infty}}_{\in \mathcal{A}, \text{ (iii)}} \\ \underbrace{\phantom{\bigcup_{k=1}^\infty}}_{\in \mathcal{A}, \text{ (ii)}} \end{array}$$

Probiert die dritte Behauptung mal selber aus, eigentlich steht alles schon da. \square

Genau wie bei Vereinigungen, gilt die Abgeschlossenheit auch für endliche Schnitte. Probiert das mal selber aus, bei dem Trick nutzt ihr aber Ω statt \emptyset .

Bemerkung. Wie in Analysis 1 nutzen wir die **erweiterte Zahlengerade**

$$\overline{\mathbb{R}} = [-\infty, +\infty] := \mathbb{R} \cup \{-\infty, +\infty\}.$$

Wir definieren

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

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

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

- (i) $\mu(\emptyset) = 0$
- (ii) Sind $A_1, A_2, \dots \in \mathcal{A}$ paarweise disjunkte Mengen (d. h. $A_i \cap A_j = \emptyset$ für alle $i \neq j$), so gilt

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

Wir nenne diese Eigenschaft σ -Additivität, wobei sich das σ auf die unendliche Anzahl von Mengen bezieht.

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

Natürlich impliziert die σ -Additivität auch die endliche Additivität $\mu\left(\bigcup_{k=1}^N A_k\right) = \sum_{k=1}^N \mu(A_k)$.

Dazu wird, wie unter der Definition der σ -Algebra, $A_{N+1} = A_{N+2} = \dots = \emptyset$ gewählt. Der Begriff „Maß“ hat durchaus einen Sinn, der an Beispielen später viel klarer wird. Man misst in einem abstrakten Sinn die Größe der messbaren Mengen. Deswegen sind die zwei definierenden Eigenschaften auch klar. Malt euch einfach mal zwei Mengen in \mathbb{R}^2 hin und überlegt, warum die „Größe“ nur für disjunge Mengen die Summe der „Größen“ der einzelnen Mengen sein sollte.

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

Folgende Begrifflichkeiten werden wir ständig nutzen, um möglichst effizient formulieren zu können:

Definition 1.1.6.

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

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

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

Diskussion 1.1.8.  [Stochastische Modellierung, Nr. 1] Warum machen die Definitionen von Wahrscheinlichkeitsräumen $(\Omega, \mathcal{A}, \mathbb{P})$ für die Modellierung von zufälligen Experimenten Sinn? Wir interpretieren dazu

- $\Omega = \text{„Das kann bei dem Experiment passieren“}$, wir können aber vielleicht das Eintreten der Elementarereignisse nicht beobachten.
- $\mathcal{A} = \text{„Ereignisse, deren Eintreten (oder Nichteintreten) beobachtet werden kann.“}$ Die σ -Algebra besteht also aus den Ereignissen des Experiments, die wir beobachten können.
- $A^C = \text{„Gegenereignis“}$, also „Ereigniss A trifft nicht ein“.
- $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$ bedeutet „Wenn man das Eintreten von Ereignis A beobachten kann, dann kann man auch beobachten, dass A nicht eintritt.“
- $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$ bedeutet „Wenn man das Eintreten sowohl von Ereignis A als auch von Ereignis B beobachten kann, dann kann auch beobachten, ob eines von beiden auftritt.“ Die Interpretation der Abgeschlossenheit bezüglich endlicher Vereinigungen ist analog („Man kann beobachten, ob eines der Ereignisse eingetreten ist“). Wir lassen hier offen, warum man für die Mathematik auch abzählbare Vereinigungen erlauben muss. Hier bleibt für den Moment nur zu sagen: Es würde nicht funktionieren.
- $\mathbb{P}(A) = \text{„Wahrscheinlichkeit des Eintretens des Ereignisses } A\text{.“}$
- $\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$ bedeutet „Gegenereignis hat Gegenwahrscheinlichkeit.“ Die Gleichheit gilt, da wegen $A \cup A^C = \Omega$ auch $\mathbb{P}(A) + \mathbb{P}(A^C) = \mathbb{P}(\Omega) = 1$ gilt.

Damit haben wir den Sinn der Definitionen einer σ -Algebra und eines Maßes hoffentlich großteils motiviert.

Als Beispiel modellieren wir den Wurf eines Würfels gemäß obiger Interpretation. Wir wählen $\Omega = \{1, \dots, 6\}$ weil das zufällige Experiment (Würfel werfen) sechs Möglichkeiten hat. Die Zahlen spielen hier keine Rolle, es geht nur darum, dass es sechs Elementarereignisse des Experiments gibt. Wir könnten die Elementarereignisse zum Beispiel auch $\Omega = \{\omega_1, \dots, \omega_6\}$ nennen. Als σ -Algebra der beobachtbaren Ereignisse nehmen wir $\mathcal{A} = \mathcal{P}(\Omega)$ weil wir alle Ereignisse des Würfelwurfs beobachten können. Ein Ereigniss $A \in \mathcal{A}$ bedeutet „Eine der Zahlen in A ist gewürfelt worden“. Weil unser Würfel fair sein soll, legen wir $\mathbb{P}(\{1\}) = \dots = \mathbb{P}(\{6\}) = \frac{1}{6}$ fest. Die Wahrscheinlichkeiten aller weiteren Ereignisse sind automatisch festgelegt, indem das Ereigniss in die disjunkten Elementarereignisse zerlegt wird, z. B. die Wahrscheinlichkeit eine gerade Zahl zu würfeln:

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

Nicht jedes zufällige Experiment ist so einfach wie das Würfeln (endlich viele Möglichkeiten), für kompliziertere zufällige Experimente (z. B. die Temperatur morgen in Mannheim) brauchen wir leider viel kompliziertere Modelle. Gehen wir also zurück zu allgemeinen Maßräumen.

Zunächst eine kleine Folgerung der Definition des Maßes, die im Sinne von „ $\mu(A) = \text{Größe von } A$ “ von Mengen total Sinn macht.

Lemma 1.1.9.  [Monotonie und Subadditivität] Es sei μ ein Maß auf einer σ -Algebra \mathcal{A} , dann gelten:

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

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

$$\mu(B) \stackrel{\mu \geq 0}{\leq} \mu(B) + \mu(A \setminus B) = \mu(B \cup A \setminus B) = \mu(A),$$

wobei wir beide definierenden Eigenschaften des Maßes genutzt haben.

(ii) wird in den Tutorien oder der großen Übung diskutiert. \square

Zu beachten ist, dass bei der Subadditivität nicht gefordert wird, dass die Mengen disjunkt sind (sonst würde schließlich Gleichheit gelten!). Die Ungleichheit entsteht dadurch, dass bei nicht-disjunkten Mengen der Schnitt mehrfach gezählt wird.

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

Bemerkung. Sind μ_1, μ_2 Maße auf \mathcal{A} und $a, b \geq 0$, so ist auch die Summe

$$a\mu_1 + b\mu_2 : A \mapsto a\mu_1(A) + b\mu_2(A)$$

ein Maß auf \mathcal{A} . Summen von Maßen nennt man auch **Mischung**. Sind \mathbb{P}_1 und \mathbb{P}_2 Wahrscheinlichkeitsmaße und zusätzlich $a + b = 1$, so ist die Mischung $\mathbb{P} = a\mathbb{P}_1 + b\mathbb{P}_2$ wieder ein Wahrscheinlichkeitsmaß.

Kommen wir nun zu ein paar Beispielen:

Beispiel 1.1.10. [endliche Gleichverteilung] Sei $\#\Omega < \infty$ und $\mathcal{A} = \mathcal{P}(\Omega)$. Dann heißt das Maß $\mu(A) = \frac{\#A}{\#\Omega}$ Gleichverteilung auf Ω . Checkt mal selber, dass μ die Eigenschaften von Maßen erfüllt. Weil $\mu(\Omega) = 1$ gilt, würde man \mathbb{P} statt μ schreiben. Der Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ ist ein Modell für das zufällige Experiment, in dem aus $\#\Omega$ vielen Elementen jedes Element mit der selben Wahrscheinlichkeit gezogen wird, zum Beispiel Lotto.

Beispiel 1.1.11. [abzählbare Verteilungen, Zählmaß] Sei Ω abzählbar, z. B. $\Omega = \mathbb{N}$. Wir wählen $\mathcal{A} = \mathcal{P}(\Omega)$ und eine Folge $(p_k)_{k \in \Omega}$ nicht-negativer Zahlen. Definieren wir

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

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

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

Die p_k werden auch Gewichte, oder Wahrscheinlichkeitsgewichte, genannt.

Beispiel 1.1.12. [Poissonverteilung auf \mathbb{N}] Hier ist ein konkretes Beispiel zu der vorherigen Klasse von Beispielen, die Poissonverteilung. Für ein $\lambda > 0$ (λ nennt man Parameter der Verteilung) sei $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ für $k \in \mathbb{N}$. Es gelten dann

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

Also definiert $\mathbb{P}(A) = e^{-\lambda} \sum_{k \in A} \frac{\lambda^k}{k!}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Man nennt \mathbb{P} auch **Poissonverteilung mit Parameter λ** .

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

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

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

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

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

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

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

Weitere wichtige Beispiele wie die geometrische Verteilung und die Binomialverteilung 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 beachtet, dass endliche Maße und Wahrscheinlichkeitsmaße sehr eng beieinander liegen: Durch $\mathbb{P}(A) := \frac{\mu(A)}{\mu(\Omega)}$ kann ein endliches Maß immer zu einem Wahrscheinlichkeitsmaß „normiert“ werden.

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

Vorlesung 2

Satz 1.1.14. [Stetigkeit von Maßen] Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $(A_n)_{n \in \mathbb{N}}$ eine Folge messbarer Mengen, so gelten:

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

Beweis.

- (i) Definiere

$$A'_1 := A_1, \quad A'_2 := A_2 \setminus A'_1, \quad A'_n := A_n \setminus A'_{n-1}, \quad n \geq 3.$$

Malt euch auf jeden Fall eine Skizze, um die Bedeutung dieser Mengen zu verstehen! Weil die A'_n paarweise disjunkt sind und $A_n = \bigcup_{k=1}^n A'_k$ gilt, folgt

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

(ii) Die Behauptung sofort aus (i) weil

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

Weil μ endlich ist gilt für alle messbaren Mengen

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

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

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

□

Für den Moment ist noch nicht so klar, wie wichtig die Stetigkeit von Maßen ist. Es wird aber kaum eine Vorlesung in der Stochastik 1 geben, in der die Stetigkeit von Maßen nicht auftaucht.

Beispiel 1.1.15. [Gegenbeispiel zu (ii) mit $\mu(\Omega) = \infty$] In dem Beweis haben wir die Endlichkeit des Maßes deutlich benutzt, sonst stände auf beiden Seiten ein unendlicher Summand. Folgendes Beispiel zeigt, dass die Aussage bei unendlichen Maßen tatsächlich schief gehen kann. Sei dazu $\Omega = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ und μ das Zählmaß aus Beispiel 1.1.11, also

$$\mu(A) = \sum_{k \in A} 1 = \#A.$$

Wegen $\mu(\mathbb{N}) = \#\mathbb{N} = +\infty$ ist das Zählmaß unendlich. Mit $A_n = \{n, n+1, \dots\}$ gilt $\mu(A_n) = +\infty$ für alle $n \in \mathbb{N}$ und $A_n \downarrow A = \emptyset$. Weil $\mu(\emptyset) = 0$ und $\mu(A_n) = +\infty$ für alle $n \in \mathbb{N}$ gilt, ist die Aussage von Satz 1.1.14 hier falsch.

1.2 Erzeuger von σ -Algebren und Dynkin-Systeme

Bisher waren die Beispiele sehr einfach. Die Angelegenheit wird aber sehr viel schwieriger, wenn wir überabzählbare Grundmengen Ω betrachten, \mathbb{R} zum Beispiel. Wir werden in diesem Abschnitt überlegen, wie man trotzdem Maße auf sehr komplizierten σ -Algebren \mathcal{A} auf Ω verstehen kann. Der Trick wird sein, die Maße nur auf relativ einfachen Teilmengen von \mathcal{A} anzuschauen. Das ist ein wenig wie in der linearen Algebra, dort müssen lineare Abbildungen auch nur auf einer Basis definiert werden. Um das Konzept dieser einfachen Teilmengen (sogenannte Erzeuger) verstehen zu können, braucht es ein wenig Vorarbeit.

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

Beweis. Ruft euch zunächst in Erinnerung, dass σ -Algebren über Ω Mengen sind, Mengen von Teilmengen von Ω . Mengen kann man schneiden, also macht die Aussage grundsätzlich Sinn. Sei nun \mathcal{A}_i , $i \in I$, eine Menge von σ -Algebren über Ω und $\mathcal{A} := \bigcap_{i \in I} \mathcal{A}_i$. Wir checken die drei Eigenschaften einer σ -Algebra für \mathcal{A} :

- (i) $\Omega \in \mathcal{A}$ ist klar weil $\Omega \in \mathcal{A}_i$ für alle $i \in I$ und damit ist Ω auch im Durchschnitt.
- (ii) Sei $A \in \mathcal{A}$, also ist $A \in \mathcal{A}_i$ für alle $i \in I$. Weil alle \mathcal{A}_i σ -Algebren sind, ist auch $A^C \in \mathcal{A}_i$ für alle $i \in I$ und damit ist A^C im Durchschnitt der \mathcal{A}_i . Folglich ist \mathcal{A} abgeschlossen bezüglich Komplementbildung.
- (iii) Sei (A_n) eine Folge von Mengen in \mathcal{A} , also $A_n \in \mathcal{A}_i$ für alle i und n . Weil das alles σ -Algebren sind, gilt

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

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

□

Wie an anderen Stellen der Mathematik (z. B. bei Unterverktorräumen) ist der Schnitt oft strukturerhaltend. Die Vereinigung meistens nicht, das ist auch bei σ -Algebren so:

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

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

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

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

Beweis. Existenz:

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

erfüllt die geforderten Eigenschaften.

Eindeutigkeit: Sei \mathcal{A}' eine weitere solche σ -Algebra. Dann gilt $\mathcal{A} \subseteq \mathcal{A}'$ weil \mathcal{A} der Schnitt über alle solche \mathcal{A}' ist und der Schnitt von Mengen in jeder Menge enthalten ist, über die geschnitten wird. Weil \mathcal{A}' die Eigenschaft (ii) erfüllt, ist auch $\mathcal{A}' \subseteq \mathcal{A}$. Damit ist $\mathcal{A} = \mathcal{A}'$ gezeigt und es gibt nur eine σ -Algebra mit den Eigenschaften (i) und (ii). □

Die σ -Algebra aus dem Korollar hat eine enorme Bedeutung in der Stochastik, ohne sie wäre der Rest dieses Abschnittes nicht machbar. Geben wir dieser σ -Algebra also einen Namen:

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

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

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

Warnung: Der Erzeuger einer σ -Algebra ist nicht eindeutig. Das wird gleich anhand der Borel- σ -Algebra deutlich werden.

In folgendem Beispiel wird eine offensichtliche kleine Beobachtung genutzt: Aus der Definition folgt sofort $\mathcal{E} \subseteq \sigma(\mathcal{E})$, $\sigma(\mathcal{E})$ ist schließlich die kleinste σ -Algebra, die \mathcal{E} enthält. Merkt euch das, so wie in folgendem Beispiel wird das öfters benutzt.

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

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

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

Beispiel 1.2.6. **[Das Beispiel - die Borel- σ -Algebra]** Wir kommen jetzt zu dem mit Abstand wichtigstem Beispiel einer σ -Algebra der Stochastik. Sei $\Omega = \mathbb{R}^d$ und $\mathcal{E} = \{O \subseteq \mathbb{R}^d: O \text{ offen}\}$. Dann heißt $\mathcal{B}(\mathbb{R}^d) := \sigma(\mathcal{E})$ die **Borel- σ -Algebra auf \mathbb{R}^d** . Die Mengen in $\mathcal{B}(\mathbb{R}^d)$ heißen **Borelmengen**.

Anhand der Borel- σ -Algebra spielen wir mal mit den neuen Begriffen rum. Wenn ihr folgende Überlegungen verstanden habt, habt ihr einen großen Schritt geschafft!

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

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

Wir müssen folgendes verstehen: Wie zeigt man allgemein für zwei Mengensysteme $\mathcal{E}, \mathcal{E}' \subseteq \mathcal{P}(\Omega)$, dass $\sigma(\mathcal{E}) = \sigma(\mathcal{E}')$ gilt? **Trick:** 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) $\sigma(\sigma(\mathcal{E})) = \sigma(\mathcal{E})$, „Idempotenz“
- (ii) $\mathcal{E} \subseteq \mathcal{E}' \Rightarrow \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{E}')$, „Monotonie“

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 zur Borel- σ -Algebra. Wir zeigen mit obigem Trick nur $\sigma(\mathcal{E}_4) = \mathcal{B}(\mathbb{R}^d)$, den Rest macht ihr in den Übungen. Es ist klar, dass $\mathcal{E}_4 \subseteq \sigma(\{O \subseteq \mathbb{R}^d : O \text{ offen}\})$, 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 $O \subseteq \mathbb{R}^d$ irgendwelche $a_1, \dots, a_d, b_1, \dots, b_d \in \mathbb{Q}$ mit $x \in (a_1, b_1) \times \dots \times (a_d, b_d) \subseteq O$. Damit gilt für jede offene Menge $O \subseteq \mathbb{R}^d$

$$O = \bigcup_{x \in O} \{x\} \subseteq \bigcup_{\text{abz. viele}} (a_1, b_1) \times \dots \times (a_d, b_d) \subseteq O,$$

also

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

weil abzählbare Vereinigungen von Mengen einer σ -Algebra (hier die offenen Quadern) wieder in der σ -Algebra sind. Wir haben jetzt beide Richtungen von (1.1) gezeigt und damit $\sigma(\mathcal{E}_4) = \sigma(\{O \subseteq \mathbb{R}^d : O \text{ offen}\}) = \mathcal{B}(\mathbb{R}^d)$ bewiesen.

Warnung: Die Borel- σ -Algebra ist wahnsinnig groß! Sie enthält alle offenen Mengen, abgeschlossene Mengen, kompakte Mengen, jegliche Arten von Quadern, alle abzählbare Vereinigungen von solchen und so weiter und so weiter. Es gibt keine Chance die Borel- σ -Algebra explizit hinzuschreiben, wir haben keine Ahnung, aus welchen Mengen die Borel- σ -Algebra wirklich besteht. Im Folgenden wollen wir deshalb, soweit möglich, Maße auf der Borel- σ -Algebra durch einen ihrer Erzeuger untersuchen, also indem wir Maße nur auf offene Mengen, kompakte Mengen oder verschiedene Quadern untersuchen. Um zu zeigen, dass man das machen kann, müssen wir etwas arbeiten. Ein Hilfsmittel dafür sind sogenannte Dynkin-Systeme:

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

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

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

Beispiel 1.2.8.

- Jede σ -Algebra ist ein Dynkin-System, die Definition einer σ -Algebra fordert mehr.
- Sind μ_1, μ_2 endliche Maße auf einem messbaren Raum (Ω, \mathcal{A}) mit $\mu_1(\Omega) = \mu_2(\Omega)$, so ist $\mathcal{M} = \{A \in \mathcal{A}: \mu_1(A) = \mu_2(A)\}$ ein Dynkin-System. Warum?
 - (i) Klar, das nehmen wir an.
 - (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 $A^C \in \mathcal{M}$. Beachtet, dass hier die Annahme der Endlichkeit wie bei der Stetigkeit der Maße benutzt wird!
 - (iii) Seien $A_1, A_2, \dots \in \mathcal{M}$ paarweise disjunkt, es gilt also $\mu_1(A_n) = \mu_2(A_n)$ für alle $n \geq 1$. Damit folgt wegen der σ -Additivität von Maßen

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

$$\text{also ist } \bigcup_{k=1}^{\infty} A_k \in \mathcal{M}.$$

Die Definitionen von σ -Algebra und Dynkin-System sind recht ähnlich. Um zu zeigen, dass ein Dynkin-System sogar eine σ -Algebra ist, nutzt man oft folgenden Proposition:

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

Beweis. Die eine Richtung ist wegen Lemma 1.1.3 easy. Überlegt mal schnell.

Sei nun \mathcal{D} ein \cap -stabiles Dynkin-System. Wir basteln etwas rum, bis wir die Vereinigungseigenschaft gezeigt haben. Das sieht vielleicht lästig aus, ist aber eine gute Möglichkeit, euch an die Rechentricks zu gewöhnen.

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

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

- (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, bekommen wir mit (i)

$$A \cup B = A \cup (\underbrace{B \setminus (A \cap B)}_{\in \mathcal{D}, \text{ wegen (i)}}) \in \mathcal{D}.$$

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

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

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

Aus (ii) folgt $B_n \in \mathcal{D}$ und dann mit (i) auch $C_n \in \mathcal{D}$. Mit der Definition der Dynkin-Systeme folgt dann

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

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

□

Vorlesung 3

Wir haben letztes Mal gezeigt, dass der Schnitt von σ -Algebren wieder eine σ -Algebra ist. Exakt genauso zeigt man, dass auch der Schnitt von Dynkin-Systemen wieder ein Dynkin-System ist. Daraus motiviert definieren wir auch wieder von Teilmengen von $\mathcal{P}(\Omega)$ erzeugte Dynkin-Systeme:

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

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

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

Der nächste Satz sieht harmlos aus, ist aber ein enorm mächtiges Werkzeug. Deshalb bekommt er auch den großen Namen „Hauptsatz“. Tiefer werden wir nicht in die Theorie der Dynkin-Systeme eintauchen, nach dem Hauptsatz kommt die für uns relevante Anwendung.

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

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

Dabei nutzten wir, dass der Schnitt über mehr Mengen natürlich kleiner ist als der Schnitt über weniger Mengen.

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

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

(a) Definiere dazu zunächst

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

für ein 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 aus den Regeln des Dynkin-Systems $d(\mathcal{E})$

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

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

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

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

□

Klar, aus dem Beweis nimmt man nicht so richtig viel Verständniss mit. Aber bevor ihr den Beweis einfach weglassst: Nur durch das Durcharbeiten von solch komischen Argumenten bekommt ihr in Kombination mit den Übungsaufgaben „Rechenroutine“ mit den relevanten Mengenoperationen. Nun kommen wir zu der wesentlichen Anwendung von Dynkin-Systemen, nur deshalb sprechen wir überhaupt über Dynkin-Systeme! Mit Dynkin-Systemen können wir ganz einfach zeigen, dass die Gleichheit von Maßen schon aus der Gleichheit der Maße auf einem \cap -stabilen Erzeuger folgt. Wenn wir zum Beispiel an die wahnsinnig große Borel- σ -Algebra $\mathcal{B}(\mathbb{R})$ denken, macht all das schnell Sinn. Um die Gleichheit von zwei Maßen auf der ganzen Borel- σ -Algebra zu zeigen reicht es, die Gleichheit auf einem der vielen Erzeuger zu checken.

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

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

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

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

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

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

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

Weil rechts und links das selbe steht, müssen alle Teilmengenrelationen sogar Gleichheiten sein. Damit gilt $\mathcal{M} = \mathcal{A}$ und das ist die Aussage des Satzes. □

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

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

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

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

Geben wir der genutzten Erweiterung endlicher Maße einen Namen:

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

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

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

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

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

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

□

So, nun endlich ein richtig konkretes Beispiel!

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

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

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

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

Die ersten beiden Eigenschaften folgen aus der Definition von Wahrscheinlichkeitsmaßen und der Monotonie von Maßen. Die weiteren Eigenschaften folgen aus der Stetigkeit von Maßen, damit dürft ihr euch in den Übungen auseinandersetzen. Um die gerade bewiesenen Sätze anzuwenden, zeigen wir folgende Behauptung:

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

In Worten: Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$ sind durch ihre Verteilungsfunktion eindeutig festgelegt. Die Behauptung folgt aus Satz 1.2.12 mit $\mathcal{E} = \{(-\infty, t]: t \in \mathbb{R}\} \subseteq \mathcal{P}(\mathbb{R})$. Checken wir dazu die benötigten Eigenschaften:

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

Genauso beweist man auch die Aussage des nächsten Beispiels:

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

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

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

1.3 Konstruktion von Maßen

Gerade haben wir gesehen, dass endliche Maße auf $\mathcal{B}(\mathbb{R})$ schon auf den Intervallen (also durch die Verteilungsfunktionen) eindeutlich festgelegt sind. Das ist eine Eindeutigkeitsaussage. Jetzt drehen wir das Ganze um und untersuchen Existenzaussagen. Am Ende soll folgendes rauskommen: Wenn wir eine geeignete Mengenfunktion (also eine Funktion auf Mengen, so wie ein Maß) nur auf den Intervallen definieren, dann gibt es auch ein passendes Maß auf $\mathcal{B}(\mathbb{R})$. Dazu brauchen wir einiges an Handwerkszeug.

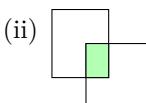
Definition 1.3.1. $\mathcal{S} \subseteq \mathcal{P}(\Omega)$ heißt **Semiring**, falls

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

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

Beispiel. $\mathcal{Q} := \{Q \subseteq \mathbb{R}^2 : Q \text{ Quader}\}$ ist ein Semiring. Checken wir anschaulich die definierenden Eigenschaften:

- (i) $\emptyset \in \mathcal{Q}$ weil $\emptyset = (0, 0) \times (0, 0)$



(ii)

C_1	C_8	
C_2	B	C_7
	C_5	C_6
C_3	C_4	

Natürlich kann man das formell sauber hinschreiben, das gibt uns aber keinen Verständnis Mehrwehrt.

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 Quadern entfernt, bleibt eine disjunkte Vereinigung von Quadern übrig. Mit Induktion kriegen wir das auch für allgemeine Semiringe hin:

Lemma 1.3.2. ► [Eine kleine Indexschlacht] Es gilt in einem Semiring auch (iii)': Sind $B_1, \dots, B_r, A \in \mathcal{S}$ mit B_1, B_2, \dots, B_r paarweise disjunkt, so existieren $C_1, \dots, C_n \in \mathcal{S}$ paarweise disjunkt mit

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

Beweis. Vollständige Induktion bezüglich r .

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

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

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

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

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

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

Also gilt

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

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

□

Vorlesung 4

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

Lemma 1.3.3. ► [Eine ziemlich große Indexschlacht] Sei \mathcal{S} ein Semiring und $\mu: \mathcal{S} \rightarrow [0, \infty]$ eine Mengenfunktion mit

- $\mu(\emptyset) = 0$

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

Dann gilt

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

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

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

Beweis.

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

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

Damit gilt wegen der geforderten Additivität von μ

$$\mu(B) = \mu(A \cup (B \setminus A)) = \mu(A \cup C_1 \cup \dots \cup C_m) = \mu(A) + \mu(C_1) + \dots + \mu(C_m) \geq \mu(A).$$

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

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

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

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

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

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

Damit gelten

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

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

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

Alles zusammen ergibt die Subadditivität:

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

□

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

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

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

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

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

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

Proposition 1.3.6.

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

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

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

- (i) Für $Z \subseteq \Omega$ gilt

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

Damit ist $\Omega \in \mathcal{A}_{\mu^*}$ gezeigt.

- (ii) Sei $A \in \mathcal{A}_{\mu^*}$, dann erfüllt (+ kommutieren und $(A^C)^C = A$ nutzen) auch A^C die definierende Eigenschaft von \mathcal{A}_{μ^*} . Also ist \mathcal{A}_{μ^*} abgeschlossen unter Komplementbildung.

- (iii) Wir zeigen zunächst die Abgeschlossenheit bezüglich Vereinigungen von zwei Mengen. Seien also $A_1, A_2 \in \mathcal{A}_{\mu^*}$. Sei $Z \subseteq \Omega$ beliebig, dann folgt mit $Z' := Z \cap (A_1 \cup A_2)$

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

Folglich gilt auch

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

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

- (iv) Per Induktion folgt aus (iii) die Abgeschlossenheit bezüglich Vereinigungen endlich vieler Mengen.

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

$$A'_1 = A_1, \quad A'_2 = A_2 \setminus A'_1, \quad \text{und} \quad A'_n = A_n \setminus (A'_1 \cup \dots \cup A'_{n-1}), \quad n \geq 3,$$

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

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

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

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

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

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

Einsetzen in die Induktionsvoraussetzung gibt

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

Damit ist der Induktionsschritt gezeigt.

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

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

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

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

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

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

Weil Z beliebig war, ist damit

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

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

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

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

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

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

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

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

Vorlesung 5

Satz 1.3.7.  **[Fortsetzungssatz von Carathéodory]** Sei \mathcal{S} ein Semiring und $\mu: \mathcal{S} \rightarrow [0, \infty]$ eine Mengenfunktion mit

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

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

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

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

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

Beachte: Weil per Definition (Analysis 1) $\inf \emptyset = +\infty$ gilt, ist $\mu^*(A)$ auch definiert, wenn man A nicht durch Mengen aus \mathcal{S} überdecken kann.

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

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

(i) $\mu^*(\emptyset) = 0$ ist klar, weil $\emptyset \in \mathcal{S}$ und $\mu(\emptyset) = 0$.

(ii) Monotonie folgt direkt aus der Definition.

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

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

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

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

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

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

Weil ε beliebig gewählt wurde, gilt damit die Subadditivität von μ^* . An dieser Stelle eine kleine Anmerkung: Eine Überdeckung durch eine abzählbare Vereinigung von abzählbaren Vereinigungen ist gleichbedeutend zu nur einer abzählbaren Vereinigung. Das Stichwort ist Cantors Diagonalverfahren, siehe Analysis 1 (oder irgendeine andere Quelle). Wir können Paare von natürlichen Zahlen auf verschiedenen Weisen zählen: Zeilenweise, Spaltenweise, oder wie eine Schlange im Diagonalverfahren.

(b) Wir zeigen $\mathcal{S} \subseteq \mathcal{A}_{\mu^*}$. Dazu müssen wir

$$\mu^*(Z) = \mu^*(Z \cap S) + \mu^*(Z \cap S^C), \quad \text{für alle } Z \subseteq \Omega, S \in \mathcal{S},$$

nachrechnen.

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

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

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

Es gelten

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

sowie analog

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

Mit den Definitionen folgt

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

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

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

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

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

„ \leq “:

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

für alle $A \in \mathcal{S}$.

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

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

Folglich gilt

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

□

Wir fassen nun den Existenz- und den Eindeutigkeitssatz zusammen, das gibt folgendes zentrale Theorem:

Satz 1.3.8.  **[Existenz und Eindeutigkeit von Maßen]** Es sei (Ω, \mathcal{A}) ein messbarer Raum, \mathcal{E} ein Semiring mit $\sigma(\mathcal{E}) = \mathcal{A}$ und $\mu: \mathcal{E} \rightarrow [0, \infty]$ eine Mengenfunktion mit

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

Dann existiert genau ein Maß $\bar{\mu}$ auf $\mathcal{A} = \sigma(\mathcal{E})$, so dass $\bar{\mu}(A) = \mu(A)$ für alle $A \in \mathcal{E}$.

Beweis. Existenz folgt direkt aus 1.3.7, Eindeutigkeit folgt direkt aus Satz 1.2.13. \square

Merkt euch für immer folgendes, das hilft euch, die verschiedenen Konzepte einzuordnen:

Existenz ist Carathéodory mit äußeren Maßen, Eindeutigkeit folgt aus Dynkin-Systemen!

Anschließend an die abstrakte Theorie wollen wir nun als Beispiel die Borel- σ -Algebra auf \mathbb{R} diskutieren. Hier wird alles viel klarer werden und das ist gerade die Anwendung, die wir für die Stochastik brauchen.

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

Das Konzept von Verteilungsfunktionen von Wahrscheinlichkeitsmaßen auf $\mathcal{B}(\mathbb{R})$ ist euch bereits bei Dynkin-Systemen und in den Übungen über den Weg gelaufen. Definieren wir nun abstrakt, welche reellen Funktionen wir Verteilungsfunktionen nennen wollen:

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

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

Der Hauptsatz über Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$ ist nun folgender bijektiver Zusammenhang zu Verteilungsfunktionen:

Satz 1.4.2.  **[Wahrscheinlichkeitsmaße aus Verteilungsfunktionen]** Für jede Verteilungsfunktion F gibt es **genau ein** Wahrscheinlichkeitsmaß \mathbb{P}_F und $\mathcal{B}(\mathbb{R})$ mit

$$\mathbb{P}_F((-\infty, t]) = F(t), \quad t \in \mathbb{R}.$$

Man sagt dann, „ \mathbb{P}_F ist gemäß F verteilt“ oder „ \mathbb{P}_F hat Verteilung F “ und schreibt $\mathbb{P}_F \sim F$.

Beweis. Eindeutigkeit: Das haben wir uns schon in Beispiel 1.2.15 überlegt.

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

$$\mathcal{S} = \{(a, b] : a \leq b\},$$

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

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

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

(i) $\mu(\emptyset) = \mu((a, a]) = F(a) - F(a) = 0$

(ii) Für die σ -Additivität von μ auf \mathcal{S} seien $(a_n, b_n] \in \mathcal{S}$ paarweise disjunkt mit

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

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

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

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

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

wobei wir $F(a_k) = F(b_{k-1})$ genutzt haben. Nun aber zurück zum allgemeinen Fall: Wir [Vorlesung 6](#) zeigen

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

denn das ist gerade die σ -Additivität

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

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

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

$$F(b) - F(a) \geq F(b_N) - F(a_1) \stackrel{\text{Teleskop}}{=} \sum_{k=1}^N (F(b_k) - F(a_k))$$

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

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

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

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

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

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

gilt, gilt auch

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

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

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

Daraus folgt dann

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

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

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

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

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

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

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

Achtung, das Argument werden wir jetzt immer wieder nutzen!

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

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

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

□

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

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

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

Beweis. Übung, ziemlich analog zum vorherigen Beweis. Wir checken die Voraussetzungen von Satz 1.3.8, hier ist eine Skizze: Betrachte

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

\mathcal{S} ist ein Semiring.

$$\mu(Q) := \text{Volumen}(Q) = \prod_{k=1}^d (b_k - a_k)$$

ist eine σ -additive Mengenfunktion auf \mathcal{S} (die σ -Additivität zeigt man mit Kompaktheit im \mathbb{R}^d , genau wie im letzten Satz). Mit der Folge $E_n := (-n, n] \times \dots \times (-n, n] \in \mathcal{S}$ haben wir eine Folge mit endlichem Volumen, die gegen \mathbb{R}^d wächst. Das ist die dritte Eigenschaft von Satz 1.3.8, es gibt also eine eindeutige Fortsetzung von μ auf $\mathcal{B}(\mathbb{R}^d)$. Das Maß nennen wir λ . Es fehlt noch, dass λ ein unendliches Maß ist. Aber auch das geht mit den Argumenten des vorherigen Beweises:

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

□

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

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

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

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

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

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

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

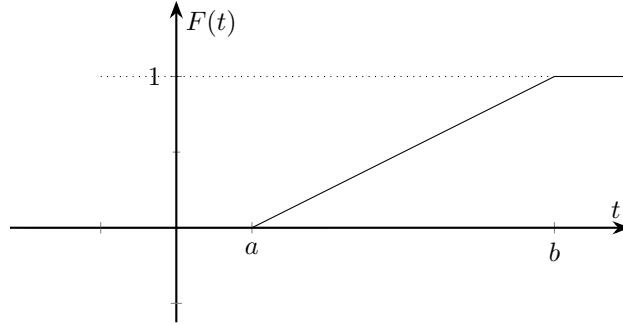
Beispiel 1.4.6. Für $a < b$ sei

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

oder anders geschrieben als

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

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



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

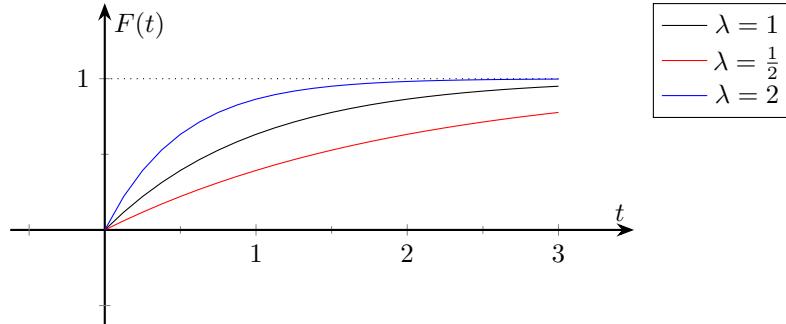
Beispiel 1.4.7. ▶ Für $\lambda > 0$ sei

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

oder anders geschrieben als

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

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



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

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

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

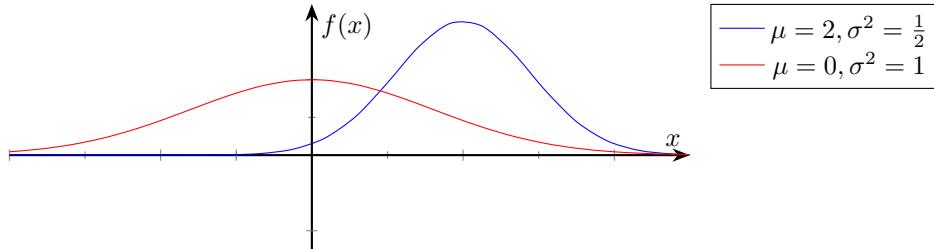
Beachte: Solch eine Integralfunktion F erfüllt automatisch die Eigenschaften einer Verteilungsfunktion (siehe große Übung)! Ist umgekehrt F von der Form (1.8), so heißt f **Dichte** von F . Verteilungsfunktionen mit Dichten nennt man auch **absolutstetig**, Wahrscheinlichkeitsmaße auf $\mathcal{B}(\mathbb{R})$ mit absolutstetiger Verteilungsfunktion nennt man **absolutstetige Maße**.

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

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

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

Verteilungsfunktion der (**standard**) **Normalverteilung**. Das Maß \mathbb{P}_F nennt man dann auch (**standard**) normalverteilt und man schreibt $\mathbb{P}_F \sim \mathcal{N}(0, 1)$. In der großen Übung wird diskutiert, dass für $\mu \in \mathbb{R}$ und $\sigma^2 \geq 0$ auch $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ eine Dichtefunktion ist. Die zugehörige Verteilung nennt man auch normalverteilt und schreibt $\mathcal{N}(\mu, \sigma^2)$.



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

Das Gegenstück zu absolutstetigen Verteilungen sind sogenannte diskrete Verteilungen:

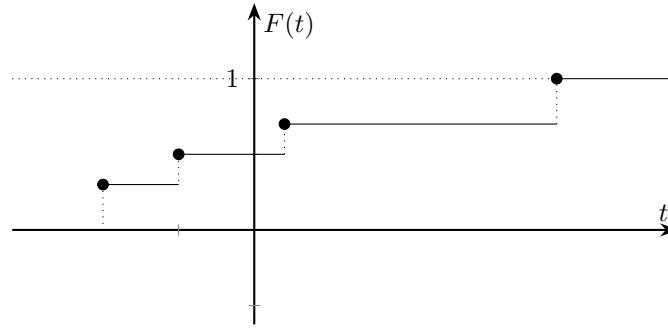
Beispiel 1.4.10. Für $a_1, \dots, a_N \in \mathbb{R}$, $N \in \mathbb{N}$ oder $N = +\infty$, mit $p_1, \dots, p_N \geq 0$ und $\sum_{k=1}^N p_k = 1$ ist

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

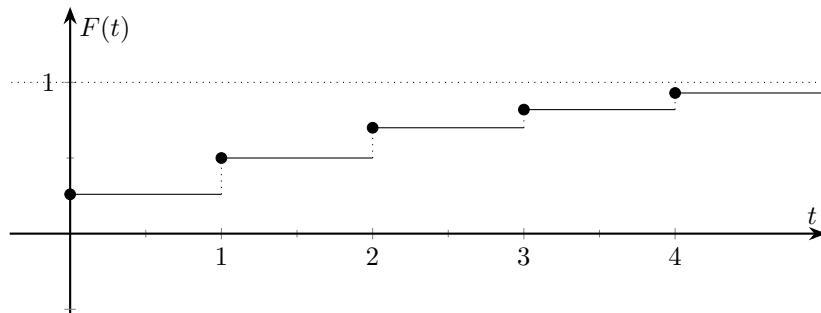
eine Verteilungsfunktion. Die zugehörigen Maße \mathbb{P}_F werden (**endliche**) **diskrete Verteilungen** genannt. In den Übungen zeigt ihr, dass die Maße im diskreten Fall ganz einfach angegeben werden können, es sind Mischungen aus Dirac-Maßen an den Stellen a_1, \dots, a_N :

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

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



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



Vorlesung 7

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

Diskussion 1.4.11. ► [Stochastische Modellierung, Nr. 2] Das Modellieren von endlich vielen Möglichkeiten ist relativ einfach, siehe Diskussion 1.1.8. Man kommt recht natürlich auf die Eigenschaften der σ -Algebra und des Maßes. Zur Erinnerung war das gleichverteilte Ziehen aus einer endlichen Menge modelliert durch den endlichen Zustandsraum Ω (=Möglichkeiten zum Ziehen), $\mathcal{A} = \mathcal{P}(\Omega)$ und der diskreten Gleichverteilung $\mathbb{P}(A) = \frac{\#A}{\#\Omega}$.

Das Modellieren von Experimenten mit unendlich vielen Möglichkeiten ist dagegen schwieriger. Wie modelliert man zum Beispiel das Ziehen aus dem Intervall $[0, 1]$, sodass kein Bereich von $[0, 1]$ bevorzugt wird? Wenn wir beobachten wollen, ob eine feste Zahl gezogen wurde oder nicht, müssen die einelementigen Mengen $\{t\}$ in der σ -Algebra sein. Wenn kein Element bevorzugt werden soll, also $\mathbb{P}(\{t\})$ für alle t gleich sein soll, führt die Unendlichkeit automatisch zu $\mathbb{P}(\{t\}) = 0$ für alle $t \in [0, 1]$. Warum das? Wenn man irgendeine Folge (a_n) unterschiedlicher Zahlen in $[0, 1]$ wählt, z. B. $a_n = \frac{1}{n}$, und $\mathbb{P}(\{t\}) =: c$ für alle t setzt, so gilt wegen der σ -Additivität von Maßen

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

also $c = 0$. Hier sehen wir deutlich den Unterschied zur Gleichverteilung auf endlichen Mengen, die einfache Definition durch Einpunktmengen führt zu nichts! Im Gegensatz zum endlichen Fall legen wir für gleichverteilten Zufall in $[0, 1]$ jetzt fest, dass die Wahrscheinlichkeit von Teilintervallen von $[0, 1]$ nur von der Länge abhängen soll. Das führt zur Forderung $\mathbb{P}((a, b]) = b - a = F(b) - F(a)$ für $a < b$ aus $[0, 1]$, wobei F die Verteilungsfunktion aus Beispiel 1.4.6 ist. Da wir als mathematisches

Modell des zufälligen Ziehens eine σ -Algebra und ein Maß haben wollen, wählen wir nun die kleinste σ -Algebra die all diese Intervalle enthält (die Borel- σ -Algebra) und darauf ein Maß, das den Intervallen die geforderten Wahrscheinlichkeiten gibt. Aufgrund des Fortsetzungssatzes gibt es so ein Maß, das ist gerade $\mathcal{U}([0, 1])$.

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

Bemerkung 1.4.12.

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

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

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

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

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

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

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

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

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

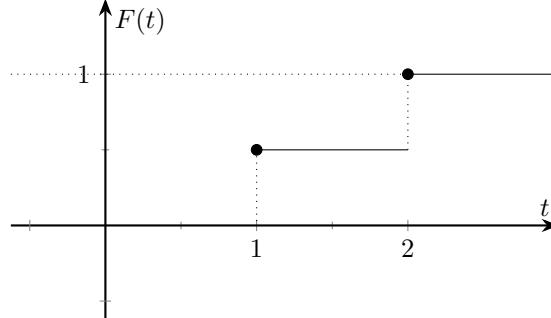
wobei $F(t-) := \lim_{s \uparrow t} F(s)$ der Linksgrenzwert aus der Analysis ist. Konsequenz: Ist F stetig in t , so hat die Einpunktmenge $\{t\}$ keine Masse. Hierzu beachte man, dass F an jeder Stelle rechtsstetig ist, die Stetigkeit somit äquivalent zu $F(t) = F(t-)$ ist. Insbesondere haben alle einpunktigen Mengen keine Masse, sofern F eine stetige Funktion ist (z. B. bei $\mathcal{U}([a, b])$, $\text{Exp}(\lambda)$, $\mathcal{N}(\mu, \sigma^2)$).

Klingt komisch, oder? Ist es aber nicht. Hier sehen wir, warum Maße erst auf überabzählbaren Mengen wirklich spannend werden:

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

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

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



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

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

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

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

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

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

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

was monoton wachsend in λ ist.

Der Fall mit Dichten: Die obige Diskussion können wir für Verteilungsfunktionen mit Dichten noch konkretisieren. Sei dazu F eine Verteilungsfunktion mit Dichte f , also $F(t) = \int_{-\infty}^t f(x)dx$.

Weil F stetig ist, haben alle einpunktigen Mengen keine Masse. Aber wie können wir an f direkt sehen, wo die Masse verteilt ist? Ist f stetig, so folgt aus dem Hauptsatz der Analysis $F'(t) = f(t)$ für alle $t \in \mathbb{R}$. Folglich impliziert ein an der Stelle t großes f ein in t stark wachsendes F und damit viel Masse um t . Andersrum impliziert ein an der Stelle t kleines f ein in t wenig wachsendes F und damit wenig Masse um t . Im Extremfall impliziert natürlich $f = 0$ in $(a, b]$ auch F konstant in $(a, b]$ und damit wird keine Masse auf $(a, b]$ verteilt. Wir merken uns grob

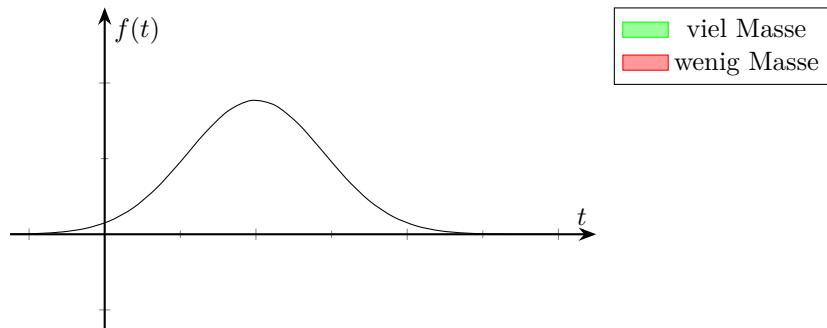
„Hat F eine Dichte, so ist viel Masse dort, wo f groß ist.“

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

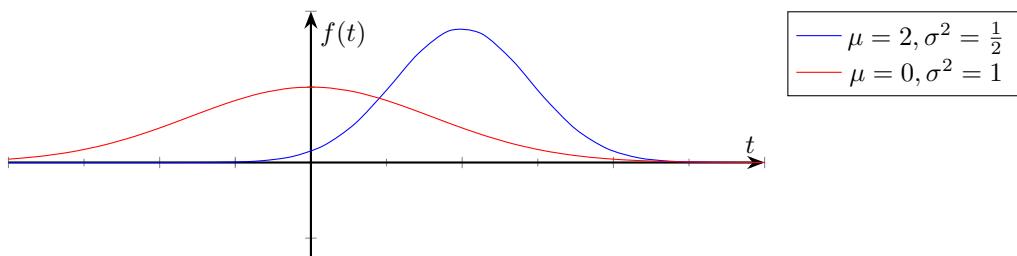
$$\mathbb{P}_F((a, b]) = F(b) - F(a) = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx = \int_a^b f(x)dx,$$

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

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



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



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

Kapitel 2

Abbildungen zwischen messbaren Räumen

Vorlesung 8

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

Mengen	Abbildungen
Gruppen	Homomorphismen
Vektorräume	Lineare Abbildungen
Metrische Räume	stetige Abbildungen

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

2.1 Messbare Abbildungen

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

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

Es gibt verschiedene Notationen für messbare Abbildungen. Man nutzt synonym

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

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

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

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

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

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

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

Beweis.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

□

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

Beispiel 2.1.6.

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

- Indikatorfunktionen

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

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

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

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

Bemerkung 2.1.7. Wir erinnern daran, dass

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

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

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

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

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

für alle $t \in \mathbb{R}$ oder

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

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

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

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

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

Schreibt euch das nächste Beispiel einmal selber hin:

Beispiel 2.1.9.

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

Die nächste Definition verallgemeinert die erzeugte σ -Algebra von einer auf beliebig viele Funktionen. Der Begriff ist für die Stochastik 1 nicht besonders wichtig, ist für die Finanzmathematik aber essentiell. Mit der Definition wird das Konzept von Information mathematisiert.

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

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

die kleinste σ -Algebra auf Ω , bezüglich derer alle f_i messbar sind. Man spricht auch hier von der **von den $f_i, i \in I$, erzeugte σ -Algebra**.

2.2 Bildmaße oder „push-forward“ eines Maßes

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

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

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

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

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

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

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

Damit ist μ_f auch σ -additiv.

□

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

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

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

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

$$\lambda(B) = \lambda(B + a), \quad \forall a \in \mathbb{R}, B \in \mathcal{B}(\mathbb{R}),$$

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

2.3 Messbare numerische Funktionen

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

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

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

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

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

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

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

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

Bemerkung 2.3.3. ►

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

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

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

sowie

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

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

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

Es gelten direkt aus der Definition folgende wichtige Identitäten

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

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

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

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

messbar, also in \mathcal{A} .

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

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

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

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

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

Eine kleine Warnung: Wir müssen beim Addieren von numerischen Funktionen aufpassen, dass die Addition wohldefiniert ist. Es darf niemals $+\infty + (-\infty)$ auftauchen, das ist nicht definiert worden. Man sollte also immer schreiben, „ $f + g$ (wenn die Addition wohldefiniert ist)“. Weil solche Probleme in der Stochastik 1 keine ernsthafte Rolle spielen, sind wir hier bewusst etwas unsauber, um nicht von den wichtigsten Punkten abzulenken.

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

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

(i) Dann sind auch die punktweise definierten Funktionen

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

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

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

messbar.

Beweis. Der Beweis wird in der großen Übung diskutiert, hier nur für g_1 . Wegen Bemerkung 2.1.7 reicht es, für alle $t \in \mathbb{R}$, $\{g_1 < t\} \in \mathcal{A}$ zu zeigen. Die Mengen $\{g_1 < t\}$ werden wieder geschrieben als abzählbare Vereinigungen, Komplemente, Schnitte, etc. von aufgrund der Voraussetzung messbaren Mengen:

$$\begin{aligned} \{g_1 < t\} &= \{\omega \in \Omega : \inf_{n \in \mathbb{N}} f_n(\omega) < t\} \\ &= \{\omega \in \Omega : f_n(\omega) < t \text{ für ein } n \in \mathbb{N}\} \\ &= \bigcup_{n \in \mathbb{N}} \{\omega \in \Omega : f_n(\omega) < t\} \\ &= \underbrace{\bigcup_{n \in \mathbb{N}} \underbrace{\{f_n < t\}}_{\in \mathcal{A}}}_{\in \mathcal{A}}. \end{aligned}$$

Für g_2, \dots, g_4 muss man sich überlegen, was $\{g_i < t\}$ eigentlich bedeutet und das dann in abzählbare Vereinigungen, Komplemente, Schnitte, etc. messbarer Mengen der Form $\{f_n \in \dots\}$ umschreiben. Probiert es aus! Jetzt ist ein guter Moment zu wiederholen, wie \liminf , \limsup definiert sind. \square

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

Kapitel 3

Convergence of Measures

Lecture 11

In this chapter the technical tools for proving Donsker's famous invariance principle will be developed. This main result of Chapter 4 will be to prove weak convergence of scaled random walks with second moment jumps towards the Brownian motion. But what does weak convergence of a sequence of stochastic processes mean? The convergence will be defined to be weak convergence of the processes interpreted as path-valued random variables. We already met the notion of weak convergence of real-valued random variables in Section ??, in this chapter weak convergence will be generalised to much more general state spaces. To do so, a bit of Functional Analysis needs to be combined with measure theory and a bit of probability. Along the way we will also prove a couple of theorems on the characterisation of random variables through moments.

3.1 A bit of topology, measure, and integration theory

To get started let us recall from analysis some topological concepts that will lead us naturally to general Borel- σ -algebras. We will always work with metric or normed spaces E , metrics will typically be denoted by d and norms $\|\cdot\|$. A central object of basic topology are open and closed sets:



Definition 3.1.1. Suppose (E, d) is a metric space and

$$B_\varepsilon(x) := \{y \in E : d(x, y) < \varepsilon\}, \quad \varepsilon > 0,$$

are the ε -balls around x .

- $A \subseteq E$ is called **open** if for all $x \in A$ there is some $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A$.
- $A \subseteq E$ is called **closed** if A^c is open.

The set of all open sets is also the **topology** of E and denoted by τ . Any open set containing a ball $B_\varepsilon(x)$ is called an (open) **neighbourhood** of x .

- $A \subseteq E$ is called **bounded** if there is some $r > 0$ such that $d(x, y) \leq r$ for all $x, y \in A$.

To study various properties of subsets of metric spaces very quickly one needs to think about more fine properties. Since there are many equivalent ways of redefining the following definitions it is likely that you might have seen different notions in your basic analysis lectures.



Definition 3.1.2. Suppose (E, d) is a metric space and $A \subseteq E$.

- $x \in A$ is called an **inner point** if there is $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq A$.



- The set of inner points of A is denoted by \dot{A} and is called **interior of A** .
- The **closure** \bar{A} is the smallest closed set containing A (the intersection of all closed sets containing A).
- $\partial A := \bar{A} \setminus \dot{A}$ is called the **boundary** of A .
- A is called **dense** in E if $\bar{A} = E$.

Please keep in mind the important facts that arbitrary unions of open sets are open and finite intersections of open sets are open. The empty set and the entire space are both open and closed.



Definition 3.1.3. Suppose (E, d) is a metric space.

- A sequence $(x_n)_{n \in \mathbb{N}} \subseteq E$ **converges** in E towards x if $\lim_{n \rightarrow \infty} d(x_n, x) = 0$, i.e. for every $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for all $n \geq N$.
- A sequence $(x_n)_{n \in \mathbb{N}} \subseteq E$ is called a **Cauchy-sequence** in E if for all $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geq N$.
- If all Cauchy-sequences in E converge, then E is called **complete**.

This section is mostly about convergence so let us recall the important connection of closeness and convergence. A set A is closed if and only if all converging sequences $(x_n) \subseteq A$ converge to some $x \in A$. Formulated differently, \bar{A} consists precisely of those elements which are the limits of sequences in A . Just check yourself some examples for intervals in \mathbb{R} !

In the discussion of conditional expectations $\mathbb{E}[X|Y]$ we have strongly used that \mathbb{R} contains the countable dense subset \mathbb{Q} . Once we combine general metric spaces with measures theory it won't come as a surprise (think of the definition of a measure and σ -algebras) that the existence of countable dense subsets should be useful.



Definition 3.1.4. A metric space (E, d) is called **separable** if there is a countable and dense subset of E .

Here is a small but useful fact from topology. A metric space (E, d) is separable if and only if the topology τ has a countable base \mathcal{B} . A base \mathcal{B} is a subset of open sets such that all open sets can be written as a union of sets from the base. The concept of a base is a bit similar to a generator of a σ -algebra but somewhat simpler as we are only allowed to take countable unions of events in σ -algebras but arbitrary unions of open sets for a base.



The set of all intervals with rational end-points is a countable base of the topology of \mathbb{R} induced by the usual norm.

There is a special class of metric spaces that turned out to be most useful in Functional Analysis and probability theory. This is the typical setting in which the general theory of Markov processes can be developed.



Definition 3.1.5. A **Polish metric space** is a complete and separable metric space.

The word Polish in the definition honours the school of Polish mathematicians from the 20th century that was responsible for the development of almost all tools of Functional Analysis. There are a few prime examples of Polish metric spaces that we will encounter again and again:

$$(\mathbb{R}, |\cdot|), \quad (\mathbb{R}^d, |\cdot|), \quad (\mathbb{C}, |\cdot|), \quad \text{and} \quad (C([0, 1]), \|\cdot\|_\infty),$$

where $C([0, 1])$ are the continuous real-valued functions on $[0, 1]$.



Definition 3.1.6. Suppose (E, d) is a metric space and $A \subseteq E$.

- A is called **compact** if every covering of A by open sets has a finite subcovering,
 A is called **sequentially compact** if every sequence in A has a subsequence that converges to a limit in A .
- A is called **relatively compact** if \bar{A} is compact, A is called **relatively sequentially compact** if all sequences in A have a subsequence with limit in \bar{A} .

Different characterisations of compactness have been discussed for $(\mathbb{R}^d, |\cdot|)$ in analysis, in particular, compactness and sequential compactness are equivalent and compact sets are precisely the closed and bounded sets (Heine-Borel Theorem). The Heine-Borel equivalence fails in most other metric spaces but we still know from analysis (check it!) that

$$A \text{ is (relatively) compact} \iff A \text{ is (relatively) sequentially compact},$$

holds in all metric spaces. In complete metric spaces a more general version of Heine-Borel can be formulated using the concept of total boundedness:



Definition 3.1.7. A set $A \subseteq E$ is called **totally bounded** if for all $\varepsilon > 0$ there are finitely many points $x_1, \dots, x_n \in A$ with $A \subseteq \bigcup_{k=1}^n B_\varepsilon(x_k)$.

Using the triangle inequality it is easy to see that totally bounded sets are always bounded, i.e. are covered by a large single ball. It is not generally the case that bounded sets are also totally bounded.



Check that boundedness and totally boundedness are equivalent in $(\mathbb{R}^d, |\cdot|)$ but are not equivalent in all infinite sets with the discrete metric (i.e. $d(x, y) = \frac{1}{2}$ for all $x \neq y$).

We will see later that totally-bounded and bounded are not the same in $(C([0, 1], \|\cdot\|_\infty))$.¹



Proposition 3.1.8. Suppose (E, d) is a complete metric space and $A \subseteq E$, then

$$A \text{ is compact} \iff A \text{ is closed and totally bounded.}$$

Proof. " \Rightarrow ": A totally bounded follows from the compactness definition by covering A with ε -balls around all elements, A closed follows from sequential compactness since in metric spaces all subsequences converge to the same limit as their convergent sequences.

" \Leftarrow ": Take a sequence $(x_n) \subseteq A$. For each $m \in \mathbb{N}$ take a finite covering $B_{\frac{1}{m}}(y_1^m), \dots$ of A . By the finiteness there must be a subsequence which lies eventually in one of the balls $B_1(y_k^1)$. Similarly, from this subsequence we extract a further subsequence which lies eventually in one of the $B_{\frac{1}{2}}(y_k^2)$. A diagonal argument gives a subsequence with $x_n \subseteq B_{\frac{1}{n}}(y_k^n)$. This is Cauchy and converges by completeness. Hence, A is sequentially compact. \square

It is not too hard to come up with counter examples to the second equivalence in non-complete spaces. For instance, take the normed space $(\mathbb{Q}, |\cdot|)$ and $A = (-\sqrt{2}, \sqrt{2}) \cap \mathbb{Q}$. Then A is closed, totally bounded but not compact.

¹nicht vergessen

As always in mathematics we are interested in the natural mappings between objects, mappings that respect the structure of the objects. For σ -algebra these were the measurable mappings, for metric spaces these are the continuous mappings:



Definition 3.1.9. A mapping between two metric spaces (E, d_E) and (F, d_F) is called **continuous** if preimages of all open sets in F are open in E . We use the notation

- $C(E, F) := \{f: E \rightarrow F \mid f \text{ continuous}\}$
- $C(E) := \{f: E \rightarrow \mathbb{R} \mid f \text{ continuous}\}$
- $C_b(E) := \{f: E \rightarrow \mathbb{R} \mid f \text{ continuous, bounded}\}$
- $C_c(E) := \{f: E \rightarrow \mathbb{R} \mid f \text{ continuous, compact support}\}$

A compactly supported function is defined as in basic analysis as a function such that $\{x \in E : f(x) = 0\}$ is contained in a compact set.

Recall that there are other equivalent ways of defining continuity via ε - δ formalism

$$\forall \varepsilon > 0 \exists \delta > 0 : d_E(x, y) < \delta \Rightarrow d_F(f(x), f(y)) < \varepsilon$$

and sequences

$$x_n \rightarrow x, n \rightarrow \infty \implies f(x_n) \rightarrow f(x), n \rightarrow \infty.$$

We will always use the most convenient formulation. Here is a not so simple exercise:



$(C_b([0, \infty)), \|\cdot\|_\infty)$ is not separable but $(C_c([0, \infty)), \|\cdot\|_\infty)$ is separable.

Now we come to the fun part, extending the concepts from measure theory on the Borel- σ -algebra of \mathbb{R} to general metric spaces. Recall that $\mathcal{B}(\mathbb{R})$ was defined to be the smallest σ -algebra containing all open sets (or closed sets, or intervals, etc.), hence, it is quite clear how we should proceed for general metric spaces.



Definition 3.1.10. For a metric space (E, d) we call

$$\mathcal{B}(E) = \sigma(\{O \subseteq E : O \text{ open}\})$$

the **Borel- σ -algebra** on E .

Of course, since σ -algebras are closed under taking complements $\mathcal{B}(E)$ is also generated by all closed sets. It is also instructive to check the following exercise to link topology and measure theory. The proof is precisely the same that shows that $\mathcal{B}(\mathbb{R})$ is also generated by all intervals.



If (E, d) is separable, then $\mathcal{B}(E) = \sigma(\{B_\varepsilon(x) : x \in E, \varepsilon > 0\})$.

Measurable mappings between two metric spaces will always be with respect to the corresponding Borel- σ -algebras. Not surprisingly we will call such measurable mappings **Borel-measurable**. From the definition of continuity and Proposition 2.1.4 it is clear that all continuous functions between metric spaces are Borel-measurable.



Definition 3.1.11. Let (E, d) be a metric space, then

- $\mathcal{M}_f(E) := \mathcal{M}_f := \{\mu : \mu \text{ is a finite measures on } \mathcal{B}(E)\}$



- $\mathcal{M}_1(E) := \mathcal{M}_1 := \{\mu : \mu \text{ is a probability measures on } \mathcal{B}(E)\}$
- $\mathcal{M}_{\leq 1}(E) := \mathcal{M}_{\leq 1} := \{\mu : \mu \text{ is a sub-probability measures one } \mathcal{B}(E)\}$

After all these definition let us prove a first proposition on measures on Polish spaces. Before checking the proof have a quick thought how you would prove the statement on $\mathcal{B}(\mathbb{R})$ using continuity of measures with $(-n, n)$ and you will immediately appreciate a useful property of \mathbb{R} , namely, to be able to fill \mathbb{R} from the inside by increasing intervals.



Proposition 3.1.12. Suppose (E, d) is Polish and $\mu \in \mathcal{M}_f$. Then, for all $\varepsilon > 0$, there is a compact set K with $\mu(K^c) < \varepsilon$.

Proof. The trick is to replace the increasing intervals $[-n, n]$ in \mathbb{R} by the right substitute and then try to argue with continuity of measures. Let x_1, x_2, \dots the countable dense subset of E and $n \in \mathbb{N}$. Then $E = \bigcup_{k=1}^{\infty} B_{\frac{1}{n}}(x_k)$ for all $n \in \mathbb{N}$. Using continuity of measures (this needs $\mu \in \mathcal{M}_f$, compare Theorem 1.1.14) fix $N_n \in \mathbb{N}$ with

$$\mu\left(E \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(x_k)\right) < \frac{\varepsilon}{2^n}$$

Now define $A := \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(x_k) \in \mathcal{B}(E)$. A is totally bounded as $A \subseteq \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(x_k)$ for all $n \in \mathbb{N}$, hence, \bar{A} is compact as E is complete. If we choose $K := \bar{A}$, then

$$\mu(K^c) = \mu(E \setminus K) \stackrel{\text{mon.}}{\leq} \mu(E \setminus A) = \mu\left(\bigcup_{n=1}^{\infty} \bigcap_{k=1}^{N_n} B_{\frac{1}{n}}^c(x_k)\right).$$

Using sub-additivity we can continue the chain of inequalities with

$$\sum_{n=1}^{\infty} \mu\left(\bigcap_{k=1}^{N_n} B_{\frac{1}{n}}^c(x_k)\right) = \sum_{n=1}^{\infty} \mu\left(E \setminus \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(x_k)\right) \leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon$$

which finishes the proof. \square

We can now turn towards a crucial topic of this chapter. Is it possible to characterise measures using only integrals over certain functions? ²



Definition 3.1.13. Let (E, d) a metric space and $F \subseteq \mathcal{M}_f(E)$ a family of measures. A family C of measurable mappings $E \rightarrow \mathbb{R}$ is called **separating family for F** if for all $\mu, \nu \in F$

$$\int_E f \, d\mu = \int_E f \, d\nu \quad \forall f \in C \cap L^1(\mu) \cap L^1(\nu) \Rightarrow \mu = \nu$$

The most simplistic (and least useful) family is the family of all measurable indicator functions $C := \{\mathbf{1}_A : A \in \mathcal{B}(E)\}$ which trivially separates all families of measures as $f = \mathbf{1}_A$ yields

$$\nu(A) = \int_E f \, d\nu = \int_E f \, d\mu = \mu(A).$$

Since the set of all indicators on measurable set is equally big as the Borel- σ -algebra there is a big desire to finde more approachable sets such as all exponential functions or all polynomial

²brauchen wir wirklich \mathcal{M}_f ?

functions. To understand the background of separating families in probability let us recall Theorem ??, which in Stochastik 1 we stated without a proof:

$$\int_{\mathbb{R}} e^{tx} d\mathbb{P}_X(x) = M_X(t) = M_Y(t) = \int_{\mathbb{R}} e^{tx} d\mathbb{P}_Y(x), \quad \forall t \in [-\varepsilon, \varepsilon] \implies \mathbb{P}_X = \mathbb{P}_Y.$$

If we reformulate, Theorem ?? states that exponential functions are separating for probability measures with exponential moments. A more precise statement will be proved towards the end of this chapter.

During the course of this chapter we will get to know which families of measurable mappings are separating for different classes of measures. We start with a first smaller step and prove that Lipschitz continuous functions are separating for finite measures.



Definition 3.1.14. Let (E, d_E) and (F, d_F) be metric spaces.

- $f: E \rightarrow F$ is called Lipschitz continuous (with constant K) if

$$d_F(f(x), f(y)) \leq K \cdot d_E(x, y) \quad \forall x, y \in E$$

- $\text{Lip}_K(E, F) := \{f: E \rightarrow F \mid \text{Lipschitz with constant } K\}$
- $\text{Lip}(E, F) := \{f: E \rightarrow F \mid \text{Lipschitz}\}$
- $\text{Lip}_K(E) := \text{Lip}_K(E, \mathbb{R})$
- $\text{Lip}(E) := \text{Lip}(E, \mathbb{R})$

As promised the Lipschitz continuous functions are separating the finite measures on E . In fact, the same proof also shows that bounded and compactly supported continuous functions separate $\mathcal{M}_f(E)$.



Proposition 3.1.15. In metric spaces $\text{Lip}_1(E)$, $C_b(E)$, and $C_c(E)$ are separating for $\mathcal{M}_f(E)$.

Proof. First of all, note that $f \in \text{Lip}_K(E)$ implies $\frac{1}{K}f \in \text{Lip}_1(E)$. Hence, if we assume $\int_E f d\mu = \int_E f d\nu$ for all $f \in \text{Lip}_1(E)$ we can also use

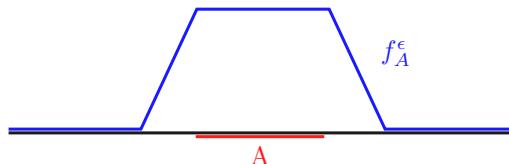
$$\int_E f d\mu = \int_E f d\nu, \quad \forall f \in \text{Lip}(E), \tag{3.1}$$

whenever the integrals are finite. We will show that (3.1) implies $\nu(A) = \mu(A)$ for all A closed. Since the closed sets form an \cap -stable generator of $\mathcal{B}(E)$ we can then deduce $\mu = \nu$ from Theorem 1.2.12 (this uses the finiteness of μ and ν).

We now fix $\mu, \nu \in \mathcal{M}_f(E)$ and $A \subseteq E$ closed. Then define the functions

$$f_A^\varepsilon(x) := \left(1 - \frac{1}{\varepsilon} \cdot d(x, A)\right)^+, \quad x \in E, \varepsilon > 0,$$

with $d(x, A) = \inf\{d(x, y) : y \in A\}$. Remember from basic analysis (or quickly check yourself by



using that \bar{A} are precisely the limits of sequences in A) that $d(x, A) = 0$ if and only if $x \in \bar{A}$. The functions f_A^ε have the following properties:

- $f_A^\varepsilon = 1$ on A
- $f_A^\varepsilon \rightarrow \mathbf{1}_A$, because for closed sets $d(x, A) = 0$ if and only if $x \in A$,
- $f_A^\varepsilon \leq 1$

Since the idea of using the functions f_A^ε is (almost) the entire point of the proof it is useful to check the following yourself:



$f_A^\varepsilon \in \text{Lip}_{\frac{1}{\varepsilon}}(E)$ and, obviously, $f_A^\varepsilon \in C_b(E)$, $f_A^\varepsilon \in C_c(E)$.

We can now use (3.1) to prove that μ and ν coincide on the closed set A :

$$\mu(A) = \int_E \mathbf{1}_A \, d\mu \stackrel{\text{DCT}}{=} \lim_{\varepsilon \rightarrow \infty} \int_E f_A^\varepsilon \, d\mu \stackrel{\text{ass.}}{=} \lim_{\varepsilon \rightarrow \infty} \int_E f_A^\varepsilon \, d\nu \stackrel{\text{DCT}}{=} \int_E \mathbf{1}_A \, d\nu = \nu(A).$$

As explained above we proved that (3.1) implies $\nu = \mu$, or, in other words, that $\text{Lip}_1(E)$ is separating for $\mathcal{M}_f(E)$. The claims about $C_b(E)$ and $C_c(E)$ follow in exactly the same way as $f_A^\varepsilon \in C_b(E)$ and $f_A^\varepsilon \in C_c(E)$. \square

More explicit classes of functions appear later in the course of this chapter.

3.2 Weak convergence of measures - the basics

As announced the ultimate goal is to prove the convergence of scaled random walks towards the Brownian motion, both seen as function-valued random variables. Let us recall from Definition ?? the notion of weak convergence of real-valued random variables

$$\int_{\mathbb{R}} f \, d\mathbb{P}_{X_n} = \mathbb{E}[f(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[f(X)] = \int_{\mathbb{R}} f \, d\mathbb{P}_X, \quad n \rightarrow \infty, \quad (3.2)$$

which is actually a notion of convergence for the sequence of probability measures $(\mathbb{P}_{X_n})_{n \in \mathbb{N}}$ on $\mathcal{B}(\mathbb{R})$. In order define a notion of convergence of stochastic processes we will introduce the law of a stochastic process (or a random variable with general state-space) and then introduce a general notion of weak convergence.



Definition 3.2.1. Let X be a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a metric space (E, d) , then **the law \mathbb{P}_X of X** is the probability measure

$$\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}), \quad A \in \mathcal{B}(E).$$

As for real-valued random variables we also use the notion $\mathbb{P}(X \in A)$ as this can be read more naturally as "probability of X in A ".

For the study of stochastic processes we will always keep in mind the example $E = C([0, 1])$ or $E = C([0, \infty))$ which are the state-spaces of stochastic processes indexed by $[0, 1)$ or $[0, \infty)$, respectively, reinterpreted as function-valued random variables.

In order to speak of convergence of processes we will thus have to define a notion of convergence on measures on more general state-spaces than \mathbb{R} . This leads us to the general notion of weak convergence of measures on metric spaces:



Definition 3.2.2. Let (E, d) a metric space and $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_f(E)$. We say $(\mu_n)_{n \in \mathbb{N}}$ converges weakly to μ if

$$\lim_{n \rightarrow \infty} \int_E f \, d\mu_n = \int_E f \, d\mu, \quad \text{for all } f \in C_b(E),$$



and write $\mu_n \xrightarrow{(w)} \mu$ or $\mu_n \Rightarrow \mu$ or $\mu = \text{w-lim}_{n \rightarrow \infty} \mu_n$.

Before returning to probability theory let us discuss important properties that help us to link the general weak convergence theory (a field of Functional Analysis) to properties from Section ?? in the particular case $E = \mathbb{R}$. There are many other ways of defining convergence of measures through distances. They are usually stronger (less sequences converge) which is one of the reasons to speak of weak convergence.



Let $(\delta_{x_n})_{n \in \mathbb{N}}$ a sequence of Dirac-measures on E such that $x_n \rightarrow x$ for $n \rightarrow \infty$.
Show that $(\delta_{x_n})_{n \in \mathbb{N}}$ converges weakly in $\mathcal{M}_1(E)$ to δ_x .

If you are familiar with Functional Analysis a bit of care is needed as the wording does not match. In the standard terminology of Functional Analysis this is not weak convergence but weak-*convergence on $\mathcal{M}_f(E)$ using that $\mathcal{M}_f(E)$ is the dual-space of $C_b(E)$.

Proposition 3.2.3. If (E, d) is separable then weak convergence in $\mathcal{M}_f(E)$ can be metrized:
Defining the **Prohorov metric** (here the definition for $\nu, \mu \in \mathcal{M}_1(E)$)

$$d_p(\mu, \nu) := \inf \{ \varepsilon > 0 : \mu(B) \leq \nu(B_\varepsilon(0)) + \varepsilon \ \forall B \in \mathcal{B}(E) \}$$

one can show that d_p is a metric on $\mathcal{M}_f(E)$ and

$$\mu_n \xrightarrow{(w)} \mu, n \rightarrow \infty \Leftrightarrow d_p(\mu_n, \mu) \rightarrow 0, n \rightarrow \infty.$$

This is important as all properties of convergence in metric spaces (such as uniqueness of limits) hold for weak convergence. Taking $f \equiv 1$ shows that also the total masses must converge. In particular, the weak limit of a sequence of probability measures is a probability measure, no mass gets lost or appears. In other words of topology, $\mathcal{M}_1(E)$ is a closed subset of $\mathcal{M}_f(E)$ with respect to the Prohorov metric.

Proof. to be written, not in the lecture □

Lecture 12

We approach weak convergence of measures in two steps. First, we will derive some equivalent conditions, usually referred to Portemantau theorem, via elementary (but tedious) manipulations with measures and integrals. In the next section we start to understand weak convergence from the point of view of convergence in the metric space (\mathcal{M}_f, d_P) with tools from metric space theory. The metric space approach is necessary in order to derive handy criteria that depend strongly on the corresponding underlying space (E, d) .

Here is the basic Portemantau theorem:



Theorem 3.2.4. (Portemantau theorem)

Let (E, d) a metric space and $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_1(E)$, then the following are equivalent:

- (i) $\mu_n \xrightarrow{(w)} \mu, n \rightarrow \infty$
- (ii) $\lim_{n \rightarrow \infty} \int_E f d\mu_n = \int_E f d\mu \quad \forall f \text{ bounded, Lipschitz continuous}$
- (iii) $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F) \quad \forall \text{ closed } F \subseteq E$
- (iv) $\liminf_{n \rightarrow \infty} \mu_n(G) \geq \mu(G) \quad \forall \text{ open } G \subseteq E$
- (v) $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A) \quad \forall A \text{ with } \mu(\partial A) = 0$

To understand similarities it is instructive to recall the statement and the proof of Theorem ?? for a sequence $\mu_n = \mathbb{P}_{X_n}$ of probability measures on $\mathcal{B}(\mathbb{R})$. In that simpler setting property (v) holds with the closed sets $A = (-\infty, t]$ because $\mathbb{P}_X(\{t\}) = 0$ is equivalent to t being a point of continuity of the distribution function F_X . In the next section we will derive a much more useful statement if (E, d) is even a Polish metric space:



- (vi) tightness + $\lim_{n \rightarrow \infty} \int_E f \, d\mu_n = \int_E f \, d\mu$ for some separating family $C \subseteq C_b(E)$
of $\mathcal{M}_1(E)$

More concrete criteria for special cases such as $E = [a, b]$, $E = [0, \infty)$, $E = \mathbb{R}^d$, or $E = C([0, 1])$ will follow below.

Proof. (i) \Rightarrow (ii): trivial (Lipschitz is continuous)

(ii) \Rightarrow (iii): Let F be closed and f_F^ε from the proof of Proposition 3.1.15. Then

$$\limsup_{n \rightarrow \infty} \mu_n(F) \stackrel{\mathbf{1}_F \leq f_F^\varepsilon}{\leq} \limsup_{n \rightarrow \infty} \int_E f_F^\varepsilon \, d\mu_n, \quad \forall \varepsilon > 0,$$

so that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mu_n(F) &\leq \inf_{\varepsilon > 0} \limsup_{n \rightarrow \infty} \int_E f_F^\varepsilon \, d\mu_n \\ &\stackrel{\text{Limit exists}}{=} \inf_{\varepsilon > 0} \lim_{n \rightarrow \infty} \int_E f_F^\varepsilon \, d\mu_n \\ &= \inf_{\varepsilon > 0} \int_E f_F^\varepsilon \, d\mu \stackrel{\text{DCT}}{=} \mu(F). \end{aligned}$$

(iii) \Leftrightarrow (iv): This follows by taking complements as F closed $\Leftrightarrow G = F^c$ open and $\mu_n(F) = 1 - \mu_n(F^c)$ and $\liminf_{n \rightarrow \infty} (-a_n) = -\limsup_{n \rightarrow \infty} (a_n)$.

(iii) + (iv) \Rightarrow (v): Let $A \in \mathcal{B}(E)$ with $\mu(\partial A) = 0$. First note that

$$\limsup_{n \rightarrow \infty} \mu_n(A) \stackrel{\text{monot.}}{\leq} \limsup_{n \rightarrow \infty} \mu_n(\bar{A}) \stackrel{(iii)}{\leq} \mu(\bar{A})$$

and

$$\liminf_{n \rightarrow \infty} \mu_n(A) \stackrel{\text{monot.}}{\geq} \liminf_{n \rightarrow \infty} (\dot{A}) \stackrel{(iv)}{\geq} \mu(\dot{A})$$

Since $\bar{A} = \dot{A} \cup \partial A$ we have $\mu(\dot{A}) + \mu(\partial A) \stackrel{\text{ass.}}{=} \mu(\bar{A})$ so that $\mu(A) = \mu(\bar{A}) = \mu(\dot{A})$ so that the above yields

$$\limsup_{n \rightarrow \infty} \mu_n(A) \leq \mu(A) \leq \liminf_{n \rightarrow \infty} \mu_n(A)$$

which implies that the limit exists and is equal to $\mu(A)$.

(v) \Rightarrow (iii): Let $F \subseteq E$ closed and enlarge F by δ : $F^\delta := \{x \in E : d(x, F) \leq \delta\}$. Since $\partial F^\delta \subseteq \{x \in E : d(x, F) = \delta\}$ the sets ∂F^δ are disjoint for different δ . Now we use that for a probability measure there cannot be uncountably many disjoint events with positive probability. So there must be a sequence $\delta_k \rightarrow 0$ along which $\mu(\partial F^{\delta_k}) = 0$. Hence, we can use (v) for F^{δ_k} for all $k \in \mathbb{N}$. Thus,

$$\limsup_{n \rightarrow \infty} \mu_n(F) \stackrel{\text{monot.}}{\leq} \limsup_{n \rightarrow \infty} \mu_n(F^{\delta_k}) \stackrel{(v)}{=} \lim_{n \rightarrow \infty} \mu_n(F^{\delta_k}) \stackrel{(v)}{=} \mu(F^{\delta_k}), \quad k \in \mathbb{N},$$

where we used that limit and limit superior coincide if a limit exists. Now we take limits in k on both sides. Since $F^{\delta_k} \downarrow F$ and the left hand side is independent of k , we obtain from monotonicity of measures the desired inequality $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$.

(iii) \Rightarrow (i): Let $f \in C_b(E)$. Without loss of generality it can be assumed that $0 < f < 1$. If not, we choose $a > 0$ and $b \in \mathbb{R}$ so that $\bar{f} := a \cdot f + b \in [0, 1]$ and the argument is continued using \bar{f} . Now define the closed sets $F_i^{(k)} := \{x \in E : f(x) \geq \frac{i}{k}\}$ for $k \in \mathbb{N}$, $0 \leq i \leq k$. For $\mu \in \mathcal{M}_1(E)$ monotonicity of integrals and the definition of the integral for simple functions yields

$$\sum_{i=1}^k \frac{i-1}{k} \left(\mu(F_{i-1}^{(k)}) - \mu(F_i^{(k)}) \right) \leq \int_E f \, d\mu \leq \sum_{i=1}^k \frac{i}{k} \left(\mu(F_{i-1}^{(k)}) - \mu(F_i^{(k)}) \right).$$

Warning: The inequalities look a bit strange. Typically one would partition the image space $[0, 1]$ into the disjoint sets $E_i^{(k)} = \{x \in E : \frac{i+1}{k} > f(x) \geq \frac{i}{k}\}$ and work with the inequalities $\sum \frac{i-1}{k} \mu(E_{i-1}^{(k)}) \leq \int_E f \, d\mu \leq \sum \frac{i}{k} \mu(E_{i-1}^{(k)})$, but those sets are not closed. If we use the closed sets $F_i^{(k)}$, then we double count and need to subtract the pieces counted twice. Since $F_k^{(k)} = \emptyset$ and writing out the summand this simplifies to

$$\frac{1}{k} \sum_{i=1}^{k-1} \mu(F_i^{(k)}) \leq \int_E f \, d\mu \leq \frac{1}{k} \sum_{i=1}^k \mu(F_{i-1}^{(k)}).$$

This looks more complicated than it is and only comes from looking closely to see the simplifications $\frac{i}{k} \mu(F_{i-1}^{(k)}) - \frac{i-1}{k} \mu(F_{i-1}^{(k)}) = \frac{1}{k} \mu(F_{i-1}^{(k)})$ which cancel half of the summands. Note that the estimates also hold for integrals with μ_n instead of μ , the argument was only based on splitting the function f . Using (iii) for the closed sets then gives

$$\begin{aligned} \limsup_{n \rightarrow \infty} \int_E f \, d\mu_n &\leq \limsup_{n \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \mu_n(F_{i-1}^{(k)}) \\ &\leq \frac{1}{k} \sum_{i=1}^k \limsup_{n \rightarrow \infty} \mu_n(F_{i-1}^{(k)}) \\ &\stackrel{(iii)}{\leq} \frac{1}{k} \sum_{i=1}^k \mu(F_{i-1}^{(k)}) \\ &\stackrel{\mu \leq 1}{\leq} \frac{1}{k} + \frac{1}{k} \sum_{i=1}^{k-1} \mu(F_{i-1}^{(k)}) \\ &\leq \frac{1}{k} + \int_E f \, d\mu. \end{aligned}$$

Since k was arbitrary, we obtain $\limsup_{n \rightarrow \infty} \int_E f \, d\mu_n \leq \int_E f \, d\mu$. The same reasoning for $-f$ gives $\liminf_{n \rightarrow \infty} \int_E f \, d\mu_n \geq \int_E f \, d\mu$. Both inequalities prove (i). \square

For the next theorem recall from Section 2.2 the notion of the push-forward of a measure under a measurable map. If $g : \Omega \rightarrow \Omega'$ is $(\mathcal{A}, \mathcal{A}')$ -measurable and μ is a measures on \mathcal{A} , then we defined

$$\mu_g(A) := \mu(g^{-1}(A)), \quad A \in \mathcal{A}',$$

also written $\mu \circ g^{-1}$ (more useful if there is a further index), is a measure on \mathcal{A}' . The measure is called the push-forward of μ under g or the image measure. If we think of converging sequences of measures it is natural to also ask for convergence of the sequence of push-forwards for a given mapping. If g is continuous, this can be proved easily:



Theorem 3.2.5. (continuous mapping theorem)

Let (E_1, d_1) , (E_2, d_2) be metric spaces and $g : E_1 \rightarrow E_2$ continuous. If μ, μ_1, μ_2, \dots



are finite measures on $\mathcal{B}(E_1)$, then

$$\mu_n \xrightarrow{(w)} \mu, n \rightarrow \infty \implies \mu_n \circ g^{-1} \xrightarrow{(w)} \mu \circ g^{-1}, n \rightarrow \infty.$$

In fact, continuity is not really needed, the assumption $\mu(\{x \in E_1 \mid g \text{ not continuous in } x\}) = 0$ is enough for the theorem to hold, but the proof is a bit lengthy.

Proof. Let $f \in C_b(E_2)$, then $f \circ g \in C_b(E_1)$. Using the transformation formula for integrals twice (see Theorem ??) then gives

$$\int_{E_2} f d\mu_n \circ g^{-1} = \int_{E_1} f \circ g d\mu_n \xrightarrow{n \rightarrow \infty} \int_{E_1} f \circ g d\mu = \int_{E_2} f d\mu \circ g^{-1}.$$

Hence, $\mu_n \circ g^{-1} \xrightarrow{(w)} \mu \circ g^{-1}$. □

We finish the section by writing down the generalised notion of convergence that was announced at the beginning of this section:



Definition 3.2.6. Let X, X_1, X_2, \dots be random variables with values in a metric space (E, d) . Then we say that $(X_n)_{n \in \mathbb{N}}$ **converges to X in distribution** if $\mathbb{P}_{X_n} \xrightarrow{(w)} \mathbb{P}_X, n \rightarrow \infty$. One usually writes $X_n \xrightarrow{(d)} X, X_n \Rightarrow X$ or $X_n \Rightarrow \mathbb{P}_X$ for $n \rightarrow \infty$ and also says $(X_n)_{n \in \mathbb{N}}$ converges weakly to X .

Realising that $f(X_n)$ is a real-valued random variable we find that convergence in distribution can be rewritten in the more probabilistic language

$$X_n \xrightarrow{(d)} X, n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)], \quad \forall f \in C_b(E).$$

In the case of real-valued random variables this is nothing else but the convergence in distribution from (3.2) that was used to formulate the central limit theorem. This also explains why sometimes convergence in distribution of random variables is called weak convergence. For random variables the continuous mapping theorem states

$$X_n \xrightarrow{(d)} X, n \rightarrow \infty \implies g(X_n) \xrightarrow{(d)} g(X), n \rightarrow \infty$$

for all continuous g , a statement that is used a lot!

We will now turn towards the study of weak convergence of measures seen as metric convergence in $(\mathcal{M}_f(E), d_P)$ which is based on the following characterisation of convergence:



Proposition 3.2.7. Let (E, d) a metric space and $(x_n)_{n \in \mathbb{N}}$ a sequence in E . Then the following are equivalent:

- $x_n \rightarrow x, n \rightarrow \infty$,
- (i) $A := \{x_n : n \in \mathbb{N}\}$ is relatively sequentially compact.
- (ii) $x_{n_k} \rightarrow x, k \rightarrow \infty$, for all converging subsequences.

Proof. " \Rightarrow ": Any subsequences of converging sequences converge to the same limit, hence, the second limit follows trivially and also the relative sequential compactness follows.

" \Leftarrow ": If (x_n) does not converge towards x then there is some $\varepsilon > 0$ and a subsequence with $d(x_{n_k}, x) > \varepsilon$. Since also the subsequence $(x_{n_k}) \subseteq (x_n)$ is relatively sequentially compact there is another subsequence $(x_{n'_k})$ of (x_{n_k}) that converges. Since the limit must be x by assumption we find $d(x_{n'_k}, x) < \varepsilon$ for k large enough. But this is a contradiction. □

In order to use the proposition for weak convergence of measures we will proceed in three steps. In Section 3.3 a characterisation of relative sequential compactness will be proved. The so-called tightness gives a condition that can be checked if compact sets of E are sufficiently well understood. Section 3.4 addresses the problem of identifying limits of subsequences. Finally, combining both sections we can derive good conditions for convergence of measures on different Polish spaces (E, d) .

Lecture 13

3.3 Relative sequential compactness of measures

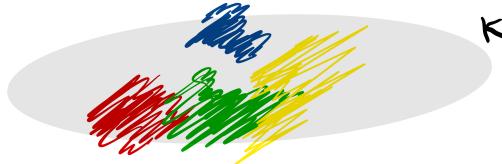
In the light of Proposition 3.2.7. The aim of this section is to derive a well accessible equivalent notion to relative sequential compactness in this particular metric space.



Definition 3.3.1. A family $\mathcal{F} \subseteq \mathcal{M}_f(E)$ is called **tight** if for every $\varepsilon > 0$ there is a compact set $K \subseteq E$ with

$$\sup_{\mu \in \mathcal{F}} \mu(K^c) < \varepsilon$$

To get a visual idea of tightness recall the interpretation of a measure that describes how mass is distributed over the state-space E . Tightness means that no mass gets lost towards infinity, the mass of all measures is (uniformly) concentrated in compact sets. Before we go into Prohorov's



Tightness of four measures visualised in terms of distribution of mass

theorem let us check some examples to get some feeling of tight families.

Example 3.3.2. • By Proposition 3.1.12 every family consisting of a single measure on a Polish space is tight

- $(\delta_{x_n})_{n \in \mathbb{N}}$ is a tight sequence for a Polish space (E, d) if and only if $(x_n)_{n \in \mathbb{N}}$ is totally bounded in E . For instance if $(x_n)_{n \in \mathbb{N}}$ is bounded for $E = \mathbb{R}$.
- If a family $(X_\alpha)_{\alpha \in I}$ of real-valued random variables is bounded in L^1 , then the family of the laws $\{\mathbb{P}_{X_\alpha} : \alpha \in I\}$ is tight. This follows from our beloved Markov inequality as follows. If C is the L^1 -bound and $\varepsilon > 0$, then $K := [-C/\varepsilon, C/\varepsilon]$ does the job:

$$\mathbb{P}_{X_\alpha}(K^c) = \mathbb{P}(|X_\alpha| > C/\varepsilon) \stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}[|X_\alpha|]}{C/\varepsilon} \leq \varepsilon$$

- $\{U([-n, n]) : n \in \mathbb{N}\}$ is not tight in $\mathcal{M}_f(\mathbb{R})$ as mass disappears towards infinity.
- If K is a compact set and $(\mu_n)_{n \in \mathbb{N}}$ is a sequence in $\mathcal{M}_1(E)$ with $\mu_n(K) = 1$ for all $n \in \mathbb{N}$, then $(\mu_n)_{n \in \mathbb{N}}$ is tight.

The final example shows that a good understanding of compact sets of E can be extremely useful to understand tightness. We will return to this observation once we discuss in more detail $E = C([0, 1])$ in which for instance sets of bounded Hölder continuous functions are compact.

The importance of tightness becomes clear with Prohorov's famous theorem in combination with Proposition 3.2.7.

**Theorem 3.3.3. (Prohorov)**

Let (E, d) be a Polish metric space and $\mathcal{F} \subseteq \mathcal{M}_1(E)$, then

$$\mathcal{F} \text{ is weakly relatively sequentially compact} \Leftrightarrow \mathcal{F} \text{ is tight.}$$

The formulation of Prohorov's theorem is frightening on first sight, weakly relatively sequentially compact means that the subset \mathcal{F} of $\mathcal{M}_1(E)$ is relatively compact (the closure is compact) with respect to the topology induced by the Prohorov metric. Using properties of closures and the sequential compactness this means that every sequence (μ_n) in \mathcal{F} has a subsequence that converges to some limit in \mathcal{M}_f but the limit does not necessarily belong to \mathcal{F} . The entire point is that Prohorov's theorem connects a measure property (tightness) with a topological property (compactness).

Proof. There is a relatively simple and a very hard direction. We start with the simpler one and discuss the hard direction only for $E = \mathbb{R}$. A sketch for the general case is given after the proof.

" \Leftarrow ": We start with arguments similar to the ones from the proof of Proposition 3.1.12. Let x_1, x_2, \dots be dense in E and

$$A_{N,n} := \bigcup_{k=1}^N B_{\frac{1}{n}}(x_k) \uparrow E, \quad N \rightarrow \infty.$$

We first prove that

$$\lim_{N \rightarrow \infty} \inf_{\mu \in \mathcal{F}} \mu(A_{n,N}) = 1, \quad \forall n \in \mathbb{N}. \quad (3.3)$$

Suppose (3.3) does not hold for some $n \in \mathbb{N}$. Then there is some $c < 1$, an increasing sequence (N_j) of natural numbers and a sequence of measures $\mu_j \in \mathcal{F}$ with

$$\lim_{j \rightarrow \infty} \mu_j(A_{n,N_j}) \leq c < 1.$$

Since the sequence is weakly relatively sequentially compact there is a subsequence (μ_{j_k}) that converges to some $\mu \in \mathcal{M}_1(E)$ (not necessarily in \mathcal{F} , μ could be in $\bar{\mathcal{F}}$). Since the $A_{n,N}$ are open, Portemanteau gives for all N

$$\mu(A_{n,N}) \leq \limsup_{k \rightarrow \infty} \mu_{j_k}(A_{n,N}) \stackrel{\text{mon.}}{\leq} \limsup_{k \rightarrow \infty} \mu_{j_k}(A_{n,N_{j_k}}) \leq c < 1.$$

But this gives a contradiction: $1 = \mu(E) \stackrel{\text{cont.}}{=} \lim_{k \rightarrow \infty} \mu(A_{n,N_{j_k}}) \leq c < 1$.

Now we use (3.3) to deduce the tightness. There are N, n_N with $\mu(A_{n_N, N}) \geq 1 - \frac{\varepsilon}{2^N}$ for all $\mu \in \mathcal{F}$. Defining $K := \bigcap_{N=1}^{\infty} \bar{A}_{n_N, N}$, K is closed and totally bounded, hence, K is compact as E is complete. Additionally,

$$\begin{aligned} \mu(K^c) &= \mu\left(\bigcup_{N=1}^{\infty} \bar{A}_{n_N, N}^c\right) \\ &\stackrel{\text{sub. add.}}{\leq} \sum_{N=1}^{\infty} \mu(\bar{A}_{n_N, N}^c) \\ &= \sum_{N=1}^{\infty} (1 - \mu(\bar{A}_{n_N, N})) \\ &\stackrel{\text{mon.}}{\leq} \sum_{N=1}^{\infty} (1 - \mu(A_{n_N, N})) \leq \varepsilon \end{aligned}$$

for all $\mu \in \mathcal{F}$. Hence, \mathcal{F} is tight.

" \Rightarrow ": Considering only $E = \mathbb{R}$ we have the big advantage that one can argue using the cumulative distribution function since

$$\mu_n \xrightarrow{(w)} \mu, n \rightarrow \infty \Leftrightarrow F_{\mu_n}(t) \rightarrow F_\mu(t), n \rightarrow \infty$$

for all points of continuity of F . Recall Theorem ?? and keep in mind that convergence in distribution of a sequence of random variables $X_n \sim F_n$ is equivalent to weak convergence of the corresponding sequence of probability measures \mathbb{P}_{F_n} . Here we use the measures μ_n and denote the corresponding CDFs by $F_n := \mu_n((-\infty, t])$, $t \in \mathbb{R}$. Reformulated this way, Prohorov's theorem is only a theorem on CDFs: Every tight sequence of probability measures has a subsequence for which the CDFs converge to a limiting CDF at all points of continuity.

Let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence in \mathcal{F} and $(F_n)_{n \in \mathbb{N}}$ the corresponding sequence of CDFs. Since $(F_n(t))_{n \in \mathbb{N}}$ is bounded for all $t \in \mathbb{R}$ there is a convergent subsequence. Using diagonalisation³ there is a subsequence $(n_k)_{k \in \mathbb{N}}$ so that $F_{n_k}(q) \rightarrow \tilde{F}(q)$, $q \in \mathbb{Q}$, for some function \tilde{F} . We will check that

- (i) $F(t) := \inf \{\tilde{F}(q) : q > t, q \in \mathbb{Q}\}$, $t \in \mathbb{R}$, is a cumulative distribution function.
- (ii) $\mu_n \xrightarrow{(w)} \mu$, $n \rightarrow \infty$, where $\mu \sim F$.

Both claims are mostly technical and not too surprising, the interesting point is how the tightness comes in. The tightness is needed to prove that F does not loose mass at infinity, i.e. $\lim_{t \rightarrow +\infty} F(t) = 1$.

- (i) Let us check the defining properties of a cumulative distribution function. We are a bit sloppy for the claims that obviously hold even though writing the details is a bit tedious (playing with ε - N and the definition of the infimum), those details do not lead to any deeper understanding. We actually skipped the same arguments before in the proof of Theorem ??.
 - $F: \mathbb{R} \rightarrow [0, 1]$, since $F_n: \mathbb{R} \rightarrow [0, 1]$.
 - \tilde{F} is increasing on \mathbb{Q} as all F_n are increasing, then F inherits the property construction.
 - F is right-continuous by definition (here one should work careful with the definition of the infimum).
 - $\lim_{t \rightarrow +\infty} F(t) = 1$, $\lim_{t \rightarrow -\infty} F(t) = 0$ is the interesting part. This is tightness, no mass gets lost. The implications look complicated but the argument is simple:

$$\begin{aligned} \text{Tightness} &\Rightarrow \forall \varepsilon > 0 \exists M \in \mathbb{Q}: \mu([-M, +M]^c) < \varepsilon \quad \forall \mu \in \mathcal{F} \\ &\Rightarrow \forall \varepsilon > 0 \exists M \in \mathbb{Q}: \mu_n([-M, +M]^c) < \varepsilon \quad \forall n \in \mathbb{N} \\ &\Rightarrow \forall \varepsilon > 0 \exists M \in \mathbb{Q}: 1 - F_n(M) + F_n(-M) < \varepsilon \quad \forall n \in \mathbb{N} \\ &\Rightarrow \forall \varepsilon > 0 \exists M \in \mathbb{Q}: F_n(M) \geq 1 - \varepsilon, F_n(-M) \leq \varepsilon \quad \forall n \in \mathbb{N} \\ &\Rightarrow \forall \varepsilon > 0 \exists M \in \mathbb{Q}: F(M) \geq 1 - \varepsilon, F(-M) \leq \varepsilon \\ &\Rightarrow \lim_{t \rightarrow +\infty} F(t) = 1, \lim_{t \rightarrow -\infty} F(t) = 0. \end{aligned}$$

- (ii) As explained above we can apply Theorem ?? and prove the pointwise convergence of the CDFs F_n at all points of continuity of F . Let t be a point of continuity and $\varepsilon > 0$.⁴ Then there are $q_i \in \mathbb{Q}$ with $q_1 < q_2 < t < q_3$ with $F(q_3) - F(q_1) < \varepsilon$. Using monotonicity we have

$$F_{n_k}(q_2) \leq F_{n_k}(t) \leq F_{n_k}(q_3),$$

³explain

⁴Bild

which gives

$$\tilde{F}(q_2) = \liminf F_{n_k}(q_2) \leq \liminf F_{n_k}(t) \leq \limsup F_{n_k}(t) \leq \limsup F_{n_k}(q_3) = \tilde{F}(q_3)$$

using the convergence on \mathbb{Q} . Hence,

$$F(q_1) \leq \tilde{F}(q_2) \leq \liminf F_{n_k}(t) \leq \limsup F_{n_k}(t) \leq \tilde{F}(q_3) \leq F(q_3).$$

But this implies that

$$\liminf F_{n_k}(t), \limsup F_{n_k}(t) \in [F(t) - \varepsilon, F(t) + \varepsilon].$$

Since ε is arbitrary we proved that $\lim F_{n_k}(t)$ exists and is equal to $F(x)$.

□

Here is a sketch on how the complicated " \Rightarrow " direction of Prohorov's theorem is proved for general Polish spaces:

- (i) Since E is separable there is a countable base $\mathcal{U} \subseteq \tau$ for the topology, i.e. $O = \bigcup_{U \in \mathcal{U}, U \subseteq O} U$ for all O open.
- (ii) Define a possible limit measure on \mathcal{U} : $(\mu_n(U))_{n \in \mathbb{N}}$ is a bounded sequence, hence, has a converging subsequence. Since there are countably many $U_1, U_2, \dots \in \mathcal{U}$ the same diagonalisation argument gives a subsequence such that $(\mu_{n_k}(U))_{k \in \mathbb{N}}$ converges for all U .
- (iii) Define $\mu(U) := \lim_{k \rightarrow \infty} \mu_{n_k}(U)$ for all $U \in \mathcal{U}$.
- (iv) Main step: Carathéodory extension style construction of a probability measure $\bar{\mu}$ on $\mathcal{B}(E)$ with $\bar{\mu}(U) = \mu(U)$ for all $U \in \mathcal{U}$.
- (v) Show $\bar{\mu}(O) \leq \liminf_{k \rightarrow \infty} \mu_{n_k}(O)$ for all O open.
- (vi) Portemanteau implies $\mu_{n_k} \xrightarrow{(w)} \mu$, $k \rightarrow \infty$.

Hence, there is a weakly converging subsequence (not necessarily to a limit in \mathcal{F}) and this is precisely the relative compactness with respect to weak convergence.

A first simple consequence of Prohorov's characterisation is the following reformulation of Proposition 3.2.7 for weak convergence of probability measures:



Proposition 3.3.4. (A useful characterization of weak convergence)

Let (E, d) be a Polish metric space and $\mu, \mu_1, \mu_2, \dots \in \mathcal{M}_1(E)$. Then weak convergence of μ_n to μ is equivalent to

- (i) $\{\mu_n : n \in \mathbb{N}\}$ is tight,
- (ii) $\lim_{n \rightarrow \infty} \int_E f \, d\mu_n = \int_E f \, d\mu$ for some separating family $C \subseteq C_b(E)$ of $\mathcal{M}_1(E)$.

Proof. " \Rightarrow ": Converging sequences are relatively compact sets (in metric spaces all subsequences have the same limit), hence, tight by Prohorov's theorem. The convergence of the integrals holds by definition of weak convergence (even for all $f \in C_b(E)$).

" \Leftarrow ": We use Proposition 3.2.7. Tightness implies the sequential relative compactness and we only need to identify μ as limit of all converging subsequences. If (μ_{n_k}) is a subsequence of (μ_n) with limit ν , then

$$\int_E f \, d\mu = \lim_{n \rightarrow \infty} \int_E f \, d\mu_n = \lim_{k \rightarrow \infty} \int_E f \, d\mu_{n'_k} = \int_E f \, d\nu, \quad f \in C.$$

The first equality holds for all $f \in C$ by assumption, the second as (μ_{n_k}) is a subsequence, and the third even for all $f \in C_b(E)$. Since C was assumed to be separating we proved $\nu = \mu$. □

In order to prove weak convergence we will often refer to Proposition 3.3.4, but there are two drawbacks that depend crucially on the underlying space E . One needs a good understanding of tightness (i.e. of compact sets of E) and one needs a good separating family. For instance for $E = C([0, 1])$ it will turn out to be more useful to circumvent the formulation of Proposition 3.3.4 and argue a bit more directly.

Lecture 14

3.4 Identification of probability laws on \mathbb{R}^d

The previous section showed how to reformulate relative sequential compactness in terms of tightness. In this section we will deal more closely with the second property from Proposition 3.2.7, the identification of limits of converging subsequences. We will give precise examples that help to apply Proposition 3.3.4 for particularly simple E . As a side product we derive ways of determining the law of a random variable through "enough" expectations. Before doing so let us dive a bit into approximation theory.


Theorem 3.4.1. (Weierstraß approximation theorem)

Let $f: [0, 1] \rightarrow \mathbb{R}$ be continuous, then there is a sequence $(f_n)_{n \in \mathbb{N}}$ of polynomials on $[0, 1]$ with $\|f_n - f\|_\infty \rightarrow 0$, $n \rightarrow \infty$.

In words: The polynomials are dense in $(C([0, 1]), \|\cdot\|_\infty)$.

The proof we give is beautiful, a probabilistic proof for an analytic theorem. The sequence of polynomials is actually given explicitly, the so-called Bernstein polynomials.

Proof. Let X_1, \dots, X_n be iid $\text{Ber}(p)$ -distributed random variables, hence, $S_n = \sum_{k=1}^n X_k$ is $\text{Bin}(n, p)$ -distributed. Now recall the proof of the weak law of large numbers (??) - Tschebycheff and Bienaymé:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \leq \frac{\mathbb{V}(S_n)}{n^2 \varepsilon^2} = \frac{n\mathbb{V}[X_1]}{n^2 \varepsilon^2} = \frac{p(1-p)}{n \varepsilon^2}$$

Next, computing the discrete expectation for $\text{Bin}(n, p)$ yields

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k} = f_n(p)$$

with the so-called Bernstein polynomial

$$f_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}, \quad x \in [0, 1].$$

Now fix $\varepsilon > 0$. Since f is uniformly continuous ($[0, 1]$ is compact) there is some $\delta > 0$ with

$$|x - x'| < \delta \Rightarrow |f(x) - f(x')| < \varepsilon,$$

so that we can estimate

$$\left|f\left(\frac{S_n}{n}\right) - f(p)\right| \leq \varepsilon + \mathbf{1}_{|\frac{S_n}{n} - p| \geq \delta} 2 \|f\|_\infty.$$

This gives

$$\begin{aligned} |f_n(p) - f(p)| &= \left| \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] - \mathbb{E}[f(p)] \right| \\ &\leq \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(p)\right|\right] \\ &= \varepsilon + 2 \cdot \|f\|_\infty \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right) \\ &\leq \varepsilon + 2 \cdot \|f\|_\infty \cdot \frac{p(1-p)}{n \cdot \delta}. \end{aligned}$$

Putting together what we have so far we obtain

$$\|f_n - f\|_\infty = \sup_{p \in [0,1]} |f_n(p) - f(p)| \leq \varepsilon + 2 \cdot \|f\|_\infty \cdot \frac{1}{n\delta^2} \rightarrow \varepsilon, \quad n \rightarrow \infty.$$

Since ε was arbitrary, we proved that $\|f_n - f\|_\infty \rightarrow 0$, $n \rightarrow \infty$. \square

Additionally, every polynomial can be approximated uniformly by a sequence of polynomials with rational coefficients. To see this claim we only need to approximate all coefficients by a sequence of rational coefficients to get

$$\|p - p_n\|_\infty \leq \sup_{x \in [0,1]} \sum_{k=1}^n |(a_k - a_{k,n})| |x^k| \leq \sum_{k=1}^n |a_k - a_{k,n}| \rightarrow 0.$$

Hence, Stone-Weierstraß combined with a diagonal argument shows that $(C([0,1]), \|\cdot\|_\infty)$ is separable. From basic analysis it should also be known that $(C([0,1]), \|\cdot\|_\infty)$ is complete, hence, it is a Polish space. In Section 4 we will show that also $C([0,\infty))$ can be turned into a Polish space, an important fact to prove Donsker's theorem.

For later purposes we now turn towards complex-valued functions that sometimes can be more useful in function approximation than real-valued functions. For that sake let us recall some definitions and facts for the complex numbers:



As a set the complex numbers are identical to \mathbb{R}^2 with a different notation for the vectors:

$$\mathbb{C} := \{z = u + iv \mid u, v \in \mathbb{R}\}.$$

The following properties will be used:

- $|z| = \sqrt{u^2 + v^2}$
- $\operatorname{Re}(z) = u, \operatorname{Im}(z) = v$
- Seen as a metric space $(\mathbb{C}, |\cdot|)$ is a Polish metric space and identical to $(\mathbb{R}^2, |\cdot|)$. In particular $\mathcal{B}(\mathbb{C}) = \mathcal{B}(\mathbb{R}^2)$ and in terms of measure theory all results from Section ?? apply. As an example, using Proposition ??, a \mathbb{C} -valued random variable $X: \Omega \rightarrow \mathbb{C}$ is measurable if and only if $\operatorname{Re}(X)$ and $\operatorname{Im}(X)$ are measurable.
- Complex conjugation: $\bar{z} = u - i \cdot v$
- Polar coordinates: $z = |z| \cdot e^{i\varphi}$ for an angle $\varphi \in [0, 2\pi)$.
- Multiplication $z_1 \cdot z_2 := (u_1 u_2 + u_2 v_2) + i(u_1 v_2 + v_2 u_1)$ so that $i^2 = -1$, or in polar coordinates: $z_1 \cdot z_2 = |z_1| |z_2| e^{i(\varphi_1 + \varphi_2)}$, which corresponds to rotation and expanding.
- Addition: $z_1 + z_2 = (u_1 + u_2) + i(v_1 + v_2)$
- Field with $+, \cdot, 0 = (0,0), 1 = (1,0)$
- $\operatorname{Re}(z) = \frac{z + \bar{z}}{2}, \operatorname{Im}(z) = \frac{z - \bar{z}}{2i}$
- Exponential function $\exp: \mathbb{C} \rightarrow \mathbb{C}$

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} = \exp(z) = \exp(u + iv) = \exp(u) \exp(iv)$$

with

$$\exp(z_1 + z_2) = \exp(z_1) \exp(z_2).$$



and Euler formula

$$\exp(iv) = \cos(u) + i \sin(v)$$

that implies $|e^{iv}| = 1$ for all $v \in \mathbb{R}$.

- $\cos(u) = \frac{e^{iu} + e^{-iu}}{2}$, $\sin(u) = \frac{e^{iu} - e^{-iu}}{2i}$

Complex integration of functions $f : \mathbb{C} \rightarrow \mathbb{C}$ is covered in complex analysis lectures with all magic tricks to compute such integrals. We will not touch upon this topic and only define the Lebesgue integral for measurable $f : \Omega \rightarrow \mathbb{C}$ by integrating separately real- and imaginary part:

$$\int_{\Omega} f \, d\mu := \int_{\Omega} \mathcal{R}e(f) \, d\mu + i \int_{\Omega} \mathcal{I}m(f) \, d\mu \in \mathbb{C}$$

if both (real-valued) integrals exist. Rules for the integral can be deduced from rules for both (real-valued) integrals. Expectations of complex-valued random variables are defined as follows:

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega) = \int_{\Omega} \mathcal{R}e(X)(\omega) \, d\mathbb{P}(\omega) + i \int_{\Omega} \mathcal{I}m(X)(\omega) \, d\mathbb{P}(\omega) \in \mathbb{C}$$

which can be computed with typical tools for real-valued random variables since

$$\mathbb{E}[X] = \mathbb{E}[\mathcal{R}e(X)] + i \mathbb{E}[\mathcal{I}m(X)].$$

An important special case appears when X is a real-random variable and $g : \mathbb{R} \rightarrow \mathbb{C}$. Expectations can be calculated in the usual way as

$$\mathbb{E}[g(X)] = \begin{cases} \int_{\mathbb{R}} g(x)f(x) \, dx & : X \text{ is absolutely continuous with density } f \\ \sum_{k=1}^N g(a_k)p_k & : X \text{ is discrete with values } a_k \text{ and probabilities } p_k \end{cases}. \quad (3.4)$$

To see why just split the expectations into two real-valued expectations, use the standard formulas and put them together. The most important example that we will discuss below is $\mathbb{E}[e^{itX}]$ for real-valued random variables.



Definition 3.4.2. Let (E, d) be a metric space and $K = \mathbb{R}$ or $K = \mathbb{C}$. A subset $C \subseteq C_b(E, K)$ is called an **algebra** of functions if

- $1 \in C$,
- $f, g \in C \Rightarrow f \cdot g, f + h \in C$,
- $f \in C, \alpha \in K \Rightarrow \alpha \cdot f \in C$,

If $K = \mathbb{C}$ we always assume C is closed under complex conjugation. We say the **algebra C separates points** if for all $y, x \in E$ there is some $f \in C$ with $f(x) \neq f(y)$.

There are many examples of algebras, for instance the set of polynomials or all exponential functions. We will discuss the examples in the upcoming sections but first deal with an important generalisation of the Weierstraß approximation theorem. The Stone-Weierstraß approximation theorem allows to generalise the $[0, 1]$ to some compact metric space and the polynomials to any algebra of functions:



Theorem 3.4.3. (Stone-Weierstraß approximation theorem)

Let (E, d) a compact metric space, $K = \mathbb{R}$ or $K = \mathbb{C}$, and $C \subseteq C_b(E, K)$ an algebra of functions that separates points. Then C is dense in $(C_b(E, K), \|\cdot\|_{\infty})$.

To see why the algebra should separate points take as an example the set of all constant functions. This forms an algebra, does not separate points and is clearly not large enough to approximate all bounded continuous functions.

Proof. Let us first consider the case $K = \mathbb{R}$.

- (i) First note that also \bar{C} is an algebra, as the defining properties rely on continuous operations that transfer through taking limits (recall that \bar{C} consists of all limits of sequences from C). By the Weierstraß approximation theorem there is a sequence $(p_n)_{n \in \mathbb{N}}$ of polynomials with $p_n \rightarrow p, n \rightarrow \infty$, uniformly on $[0, 1]$, where $p(x) = \sqrt{x}$. If $f \in \bar{C}$, then $|f| \in \bar{C}$ because

$$|f| = \|f\|_\infty \cdot \lim_{n \rightarrow \infty} p_n \underbrace{\left(\frac{f^2}{\|f\|_\infty^2} \right)}_{\in [0, 1]} \in \bar{C},$$

$\underbrace{\quad}_{\in \bar{C} \text{ as } \bar{C} \text{ is an algebra}}$

as \bar{C} is closed. Using $f \vee g = \frac{1}{2}(f + g + |f - g|)$ and $f \wedge g = \frac{1}{2}(f + g - |f - g|)$ we see that the algebra \bar{C} (operations transfer to limit) is also closed under taking pointwise maxima and minima.

- (ii) Now fix $\varepsilon > 0$. Let $f \in C_b(E, \mathbb{R})$ and $x \in E$. Then there is $g_x \in \bar{C}$ with

- $g_x(x) = f(x),$
- $g_x(y) \leq f(y) + \varepsilon, \forall y \in E.$

To see why note that C separates points, hence, for all $z \in E \setminus \{x\}$ there is a function $H_z \in C$ with $H_z(z) \neq H_z(x)$. By adding a constant we can assume that $H_z(x) = 0$. Next, define $h_x = f$ and, for $z \neq x$,

$$h_z(y) := f(z) + \frac{f(x) - f(z)}{H_z(x)} \cdot H_z(y), \quad y \in E,$$

which is a function in C that coincides with f in x and z . Since f and h_z are continuous, for all $z \in E$ there is a neighbourhood U_z of z with $h_z \leq f + \varepsilon$ on U_z . Using compactness there is a finite covering U_{z_1}, \dots, U_{z_n} of E of such neighbourhoods. If finally we define $g_x := \min\{h_{z_1}, \dots, h_{z_n}\}$, then $g \in \bar{C}$ by (i) and g satisfies the two claimed properties by the construction.

- (iii) Since f and g_x are continuous and $f(x) = g_x(x)$ there are neighborhoods U_x of x with $g_x \geq f - \varepsilon$ on U_x . By compactness finitely many V_{x_1}, \dots, V_{x_k} cover E . Then define $g := \max\{g_{x_1}, \dots, g_{x_k}\} \in \bar{C}$. By construction this gives $f + \varepsilon \geq g \geq f - \varepsilon$ or $\|f - g\|_\infty < \varepsilon$. Since ε is arbitrary we can find a sequence in \bar{C} that converges uniformly to f . In other words, $\bar{C} = C_b(E, \mathbb{R})$.

It remains to consider the case $K = \mathbb{C}$. If $f = \mathcal{Re}(f) + i\mathcal{Im}(f) \in C$, then real- and imaginary-part are in C . This follows from the assumption on C by writing $\mathcal{Re}(f) = \frac{f + \bar{f}}{2}$ and $\mathcal{Im}(f) = \frac{f - \bar{f}}{2i}$. Hence,

$$C_{\mathcal{R}} := \{\mathcal{Re}(f) : f \in C\} \subseteq C \quad \text{and} \quad C_{\mathcal{I}} := \{\mathcal{Im}(f) : f \in C\} \subseteq C$$

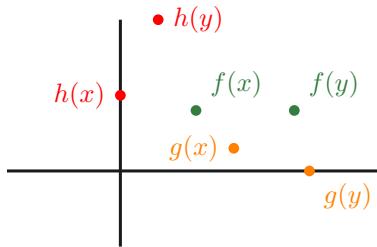
and, thus, both form algebras of real functions. Both sets are also separating according to the following trick. Suppose x and y are separated by $f \in C$, that is $f(x) \neq f(y)$. Since C is closed under adding 1 and multiplication with constants (which is rotation and stretching) there are functions $h, g \in C$ such that

$$h(x) = ia_1, h(y) = a_2 + ia_3 \quad \text{and} \quad g(x) = b_1 + ib_2, g(y) = b_3$$

for some $a_i, b_i \neq 0$. Thus,

$$\mathcal{Re}(h)(x) = 0 \neq b_1 = \mathcal{Re}(g)(x) \quad \text{and} \quad \mathcal{Im}(h)(y) = a_3 \neq 0 = \mathcal{Im}(g)(y).$$

Hence, $\mathcal{Re}(h) \in C_{\mathcal{R}}$ and $\mathcal{Im}(g) \in C_{\mathcal{I}}$ both separate x and y . This proves that $C_{\mathcal{R}}$ and $C_{\mathcal{I}}$ are separating algebras of real functions. Why did we need to involve this little trick? It is possible that the imaginary parts and/or real parts of $f(x)$ and $f(y)$ coincide. If they do, then the imaginary parts and/or real parts of f do not separate x and y . To avoid this problem we can multiply f by a constant z' to rotate (and stretch) the complex numbers $f(x)$ and $f(y)$ to ensure that the real and imaginary parts of $h(z) := z'f(z) \in C$ differ and thus separate points. The trick is best understood in a picture:



The rotation trick to separate points

Therefore, from the above, $\bar{C}_{\mathcal{R}} = \bar{C}_{\mathcal{I}} = C_b(E, \mathbb{R})$. Since $C_b(E, \mathbb{C}) = C_b(E, \mathbb{R}) + i \cdot C_b(E, \mathbb{R})$ we find that $C = C_{\mathcal{R}} + i \cdot C_{\mathcal{I}}$ is dense in $C_b(E, \mathbb{C})$. \square

What is the magic of Stone-Weierstraß? Proving that families of functions are dense in other families is typically hard, a purely analytic ε - N story. Stone-Weierstraß allows us to prove the density in a completely different way, only checking very simple algebraic properties! If we think about polynomials or exponential functions the algebraic properties are easily checked.

The situation becomes even simpler if integrals are involved as linearity of integrals fits nicely to the properties of algebras.

Lecture 15



Corollary 3.4.4. Let (E, d) be a compact metric space and $C \subseteq C_b(E, K)$ a family that separates points, is closed under multiplication and contains 1 (constant 1 function). Then C is separating for $\mathcal{M}_1(E)$.

The examples that will appear in the sequel are typically polynomials or exponential functions on compact subsets of \mathbb{R} .

Proof. We start with a little trick on separating functions and algebras that relies on the linearity of integrals. Let $\mu, \nu \in \mathcal{M}_1(E)$ with $\int_E g \, d\mu = \int_E g \, d\nu$ for all $g \in C$. Taking all linear combinations of functions in C and calling them C' then by linearity of integrals the equality also holds for all functions in C' and C' is an algebra of functions. Using the Stone-Weierstraß theorem, the equality of integrals holds for a dense subset of $C_b(E)$.

Fix $\varepsilon > 0$. Then for all $f \in C_b(E)$ there exists $g \in C'$ with $\|f - g\|_{\infty} < \varepsilon$ so that

$$\begin{aligned} & \left| \int_E f \, d\mu - \int_E f \, d\nu \right| \\ & \leq \left| \int_E f \, d\mu - \int_E g \, d\mu \right| + \underbrace{\left| \int_E g \, d\mu - \int_E g \, d\nu \right|}_{=0, \text{ by assumption}} + \left| \int_E g \, d\nu - \int_E f \, d\nu \right| \\ & \leq \|f - g\|_{\infty} \cdot \mu(E) + 0 + \|f - g\|_{\infty} \cdot \nu(E) = 2\varepsilon. \end{aligned}$$

Since this works for all $\varepsilon > 0$ it follows that

$$\int_E f \, d\mu = \int_E f \, d\nu, \quad \forall f \in C_b(E).$$

Recalling from Proposition 3.1.15 that $C_b(E)$ is separating for $\mathcal{M}_1(E)$ we proved $\mu = \nu$. Hence, C is separating for $\mathcal{M}_1(E)$. \square



Theorem 3.4.5. (i) Every measure $\mu \in \mathcal{M}_1([a, b])$ is uniquely determined by all integrals

$$\int_{[a,b]} x^m \, d\mu(x), \quad m \in \mathbb{N}.$$

(ii) The law of a bounded real-valued random variable is determined by all its moments, i.e. if $\mathbb{E}[X^m] = \mathbb{E}[Y^m]$ for all $m \in \mathbb{N}$, then $X \sim Y$.

Proof. All we need to do is to apply Corollary 3.4.4.

First note that $([a, b], |\cdot|)$ is a compact metric space. Define C to be the family of monomials on $[a, b]$, that is $C = \{x^m : m \in \mathbb{N}_0\}$. Then C is closed under multiplication, separates points in $[a, b]$ and contains the constant 1 function. Hence, C is a separating family for $\mathcal{M}_1([a, b])$. But then, recalling the definition of a separating family,

$$\int_{[a,b]} x^m \, d\mu(x) = \int_{[a,b]} x^m \, d\nu(x), \quad \forall m \in \mathbb{N}_0,$$

implies $\nu = \mu$.

(ii) follows from the first claim applied to the laws \mathbb{P}_X as $\mathbb{E}[X^m] = \int_{[a,b]} x^m \, d\mathbb{P}_X(x)$. \square

The most simple examples of bounded random variables are uniformly distributed random variables:



A little induction shows that all moments of $U \sim \mathcal{U}([a, b])$ are given by

$$\mathbb{E}[U^n] = \frac{1}{n+1} \sum_{k=0}^n a^k b^{n-k} = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)} \quad (3.5)$$

Hence, if for some random variable X with values in $[a, b]$ one can show that all moments satisfy the formula from (3.5) then the random variable is uniformly distributed on $[a, b]$. If for some reason one can show that a random variable X with values in $[a, b]$ has the moments from (??), then X is uniformly distributed.

It is important to have an example in mind that shows that all moments do not uniquely determine the law of unbounded random variables. Here is nice counterexample that appears in the Black-Scholes theory of time-continuous financial markets. A random variable X is called **log-normal** distributed if $X = \exp(Z)$, with $Z \sim \mathcal{N}(0, 1)$. Using substitution in $\mathbb{P}(X \leq t) = \mathbb{P}(Z \leq \log(t)) = \int_{-\infty}^{\log(t)} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ shows that X has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-\frac{1}{2} \log(x)^2} \mathbf{1}_{[0,\infty)}(x), \quad x \in \mathbb{R}.$$

From the moment generating function of the standard Gaussian distribution (see ??) we obtain a formula for the moments: $\mathbf{E}[X^n] = \mathbf{E}[e^{nZ}] = e^{\frac{1}{2}n^2}$. Next, we will write down an entire family of density functions that give the same moments but are log-normal densities. For $\alpha \in [-1, 1]$ define

$$f_\alpha(x) := f_X(x) \cdot (1 + \alpha \cdot \sin(2\pi \log(x))) \mathbf{1}_{[0,\infty)}(x), \quad x \in \mathbb{R}.$$

Those functions are clearly non-negative and the computation below for $n = 0$ shows they integrate to 1, hence, they are densities. To prove the moments are identical to those of the log-normal distribution it suffices to show that

$$m(n) := \int_0^\infty x^n \cdot f_X(x) \cdot \sin(2\pi \log(x)) dx = 0, \quad \forall n \in \mathbb{N}_0,$$

and combine with $\int_{\mathbb{R}} x^n \cdot f_X(x) dx = \mathbb{E}[X^n] = e^{\frac{1}{2}n^2}$. Here is the computation:

$$\begin{aligned} m(n) &= \int_{-\infty}^\infty e^{y \cdot n} \cdot f_X(e^y) \cdot \sin(2\pi y) \cdot e^y dy \\ &\stackrel{\text{subst. } y=z+n}{=} \int_{-\infty}^\infty e^{zn+n^2} f_X(e^{z+n}) \sin(2\pi z + 2\pi n) e^{z+n} dz \\ &= \frac{1}{\sqrt{2\pi}} e^{n^2} \int_{-\infty}^\infty e^{-\frac{1}{2}z^2 - \frac{1}{2}n^2} \sin(2\pi z) dz \\ &= \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}n^2} \underbrace{\int_{-\infty}^\infty e^{-\frac{1}{2}z^2} \sin(2\pi z) dz}_{=0, \text{ odd function and integrable}} = 0. \end{aligned}$$

That's it.

Log-normal distributions are not determined by all moments but they are determined by all negative exponential moments as the following theorem shows:



Theorem 3.4.6. (i) Every measure $\mu \in \mathcal{M}_1([0, \infty])$ is uniquely determined by all integrals

$$\int_{[0, \infty)} e^{-\lambda x} d\mu(x), \quad \lambda > 0.$$

(ii) The law of a non-negative random variable is determined by its Laplace transformation $L_X(\lambda) := \mathbb{E}[e^{-\lambda X}]$, $\lambda \geq 0$, i.e. if $L_X = L_Y$ then $X \sim Y$.

Of course part (ii) of the theorem also holds for non-positive random variables replacing the Laplace transformation by the moment generating function $M_X(t) = \mathbb{E}[e^{tX}]$ for all $t > 0$.

Proof. As in the previous proof we would like to apply Corollary 3.4.4 to the family of exponential functions, but there is a problem as $[0, \infty)$ is not compact. We apply a trick from Functional Analysis and use $E = [0, \infty]$ instead (the so-called one-point compactification).



$([0, \infty], d)$ is a compact metric space with $d(x, y) := |e^{-x} - e^{-y}|$. The metric is compatible with the usual convergence in $[0, \infty]$.

Note that convergence towards ∞ is divergence from basic analysis. Continuity of functions $f : [0, \infty] \rightarrow \mathbb{R}$ is best understood through sequences, $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ and this takes in particular sequences that diverge to ∞ . Hence, we can extend continuous functions on $[0, \infty)$ to continuous functions on $[0, \infty]$ if and only if $\lim_{x \rightarrow \infty} f(x)$ exists and in that case we must set $f(\infty) = \lim_{x \rightarrow \infty} f(x)$. For the exponential functions we thus define

$$f_\lambda(x) := \begin{cases} e^{-\lambda x} & : x \in [0, +\infty) \\ \lim_{x \rightarrow +\infty} e^{-\lambda x} & : x = +\infty \end{cases}.$$

Then $C := \{f_\lambda : \lambda \geq 0\} \subseteq C_b([0, \infty])$ separates points, $f_0 \equiv 1 \in C$, and $f_\mu \cdot f_\lambda = f_{\mu+\lambda}$. By Corollary 3.4.4 C separates $\mathcal{M}_1([0, \infty])$. By extending measures as

$$\bar{\mu}(A) := \mu(A \cap [0, \infty)), \quad A \in \mathcal{B}([0, \infty]).$$

we find that the exponential functions also separate $\mathcal{M}_1([0, \infty))$:

$$\begin{aligned} \int_{[0, \infty)} e^{-\lambda x} d\mu(x) &= \int_{[0, \infty)} e^{-\lambda x} d\nu(x), \quad \forall \lambda \geq 0 \\ \Rightarrow \quad \int_{[0, \infty]} f d\bar{\mu} &= \int_{[0, \infty]} f d\bar{\nu}, \quad \forall f \in C \\ \Rightarrow \quad \bar{\mu} &= \bar{\nu} \\ \Rightarrow \quad \mu &= \nu. \end{aligned}$$

(ii) follows from the first claim applied to the laws \mathbb{P}_X as $L_X(\lambda) = \int_{[0, \infty]} e^{-\lambda x} d\mathbb{P}_X(x)$. \square

Here is another version of the theorem which can help you to understand the proof better.



Prove that a random variable $X \geq 1$ is determined by all its negative moments $\mathbb{E}[X^{-m}]$, $m \in \mathbb{N}$.

Could we extend the same argument to measures on \mathbb{R} (i.e. all real-valued random variables)? No! To do so we needed a family that is closed under multiplication (that leads to polynomials or exponentials) that converge at $+\infty$ and $-\infty$ in order to be extandable to continuous function on $[-\infty, +\infty]$. Trying to find such a family is a good exercise to better understand the previous proofs. If you try you will realise that all algebras you come up with do not separate the points $+\infty$ and $-\infty$!

The trick that we get to know below is to replace the exponential functions by complex exponential functions.



Definition 3.4.7. For $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ the function $\varphi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$ defined by

$$\varphi_\mu(t) := \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mu(x), \quad t \in \mathbb{R}^d,$$

is called the **characteristic function** (or Fourier transform) of μ .

The wording Fourier transformation comes from Fourier Analysis where integrable functions are studied through their Fourier transform $\hat{f}(x) := \int e^{i\langle x, y \rangle} f(y) dy$. The name characteristic functions comes from the fact that the function φ_μ uniquely characterised μ as we prove below.

As always we can either define objects for probability measures or random variables with the corresponding law. Since the characteristic function is such an important object let us fix the other notation as well:



Definition 3.4.8. For a random vector X the function $\varphi_X : \mathbb{R}^d \rightarrow \mathbb{C}$ defined by

$$\varphi_X(t) := \mathbb{E}[e^{i\langle t, X \rangle}] = \varphi_{\mathbb{P}_X}(t), \quad t \in \mathbb{R}^d,$$

is called the **characteristic function** of X .

The characteristic function plays exactly the same role as moment generating function \mathcal{M}_X , or the distribution function F_X . It is just a function with certain properties (see below) that can be used to study more complicated mathematical objects such as measures or random variables. The major advantage of the characteristic function is that φ is always well-defined because φ_X is always defined as $|e^{ix}| = 1$. The magic about the complex exponential $x \mapsto e^{ix}$ is that it is much more useful to work with (bounded) but carries as much information as the unbounded real exponential function $x \mapsto e^x$.

Here are some properties that indicate that characteristic functions are useful to work with:



Lemma 3.4.9. Let X be a random vector and $\varphi_X: \mathbb{R}^d \rightarrow \mathbb{C}$ the characteristic function.

- (i) $\varphi_X(0) = 1$ and $|\varphi_X(t)| \leq 1$ for all $t \in \mathbb{R}^d$.
- (ii) $\varphi_{aX+b}(t) = \varphi_X(at)e^{i\langle t, b \rangle}$ for all $a \in \mathbb{R}$ and $t, b \in \mathbb{R}^d$.
- (iii) φ is real-valued if X is symmetric, i.e. $\mathbb{P}_X = \mathbb{P}_{-X}$.
- (iv) $\varphi_{X+Y} = \varphi_X \cdot \varphi_Y$ if X and Y are independent.

Proof. If you are not used to integrals for complex-valued functions the arguments are a bit more complicated than one might think.



To get started please check the linearity $\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$ for $\alpha, \beta \in \mathbb{C}$ of complex integrals by using the definition and properties of real integrals.

- (i) The first claim is clear as $e^0 = 1$. Let us first assume the usual triangle inequality also for complex integrals. Then we obtain

$$|\varphi_X(t)| = |\mathbb{E}[e^{i\langle t, X \rangle}]| = \left| \int_{\Omega} e^{i\langle t, X(\omega) \rangle} d\mathbb{P}(\omega) \right| \stackrel{\Delta}{\leq} \int_{\Omega} |e^{i\langle t, X(\omega) \rangle}| d\mathbb{P}(\omega) = \int_{\Omega} 1 d\mathbb{P}(\omega) = 1.$$

The triangle inequality $|\int_{\Omega} f d\mu| \leq \int_{\Omega} |f| d\mu$ is more complicated than one might think. Suppose φ is the angle from the polar coordinate representation of $\int_{\Omega} f d\mu$. Then

$$\left| \int_{\Omega} f d\mu \right| = e^{-i\varphi} \int_{\Omega} f d\mu = \int_{\Omega} e^{-i\varphi} f d\mu = \int_{\Omega} \Re(e^{-i\varphi} f) d\mu \leq \int_{\Omega} |e^{-i\varphi} f| d\mu = \int_{\Omega} |f| d\mu.$$

The first equality holds as multiplying with $e^{-i\varphi}$ is a rotation, the second is linearity, the third holds as $|\cdot| \in \mathbb{R}$ so that the imaginary part of the integral vanishes, and the fourth is monotonicity for real-valued integrals. Finally, the fifth equality is $|e^{-i\varphi} f| = |e^{-i\varphi}| |f| = |f|$.

- (ii) Write it down yourself!
- (iii) Recall that $e^{-i\theta} = \cos(-\theta) + i \sin(-\theta) = \cos(\theta) - i \sin(\theta) = \overline{e^{i\theta}}$, hence,

$$\varphi_{-X}(t) = \mathbb{E}[e^{-i\langle t, X \rangle}] = \mathbb{E}[\overline{e^{i\langle t, X \rangle}}] = \overline{\mathbb{E}[e^{i\langle t, X \rangle}]} = \overline{\varphi_X(t)},$$

where we used

$$\overline{\int f d\mu} = \overline{\int \Re(f) d\mu + i \int \Im(f) d\mu} = \int \Re(f) d\mu - i \int \Im(f) d\mu = \int \overline{f} d\mu.$$

The claim now follows immediately.

- (iv) The proof is essentially the same that we have seen for moment generating functions in Proposition ???. All we need to do is to extend Theorem ?? to complex-valued functions. But this is simple using the linearity from above:



Check that $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ holds for independent random variables X and Y .

But then

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{i\langle t, X+Y \rangle}] = \mathbb{E}[e^{i\langle t, X \rangle}] \cdot \mathbb{E}[e^{i\langle t, Y \rangle}] = \varphi_X(t) \cdot \varphi_Y(t),$$

for all $t \in \mathbb{R}$.

□ Lecture 16

To get a feeling let us check some examples. Most computations are almost identical to the computations with moment generating functions from Section ??.

Example 3.4.10. Using the discrete computation rule from (3.4) yields

$$\varphi_{\text{Poi}(\lambda)}(t) = \mathbb{E}[e^{itX}] = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{\lambda(e^{it}-1)}, \quad t \in \mathbb{R},$$

and

$$\begin{aligned} \varphi_{\text{Bin}(n,p)}(t) &= \mathbb{E}[e^{itX}] \\ &= \sum_{k=0}^n e^{itk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^{it})^k (1-p)^{n-k} = (1-p+pe^{it})^n, \quad t \in \mathbb{R}. \end{aligned}$$

The second example can also be computed from (iv) of Lemma 3.4.9 using

$$\varphi_{\text{Ber}(p)}(t) = \mathbb{E}[e^{itX}] = e^{it}p + (1-p)e^{it0}, \quad t \in \mathbb{R},$$

and that a binomial random variable is a sum of n independent Bernoulli random variables.

Example 3.4.11. Using the computation rule from (3.4) yields

$$\varphi_{\mathcal{U}([0,a])}(t) = \frac{e^{iat} - 1}{iat}, \quad t \in \mathbb{R}.$$

The computation is straight forward using the definition of the complex integral:

$$\begin{aligned} \mathbb{E}[e^{itX}] &= \int_0^a e^{itx} \frac{1}{a} dx \\ &= \frac{1}{a} \int_0^a \cos(tx) dx + \frac{1}{a} i \int_0^a \sin(tx) dx \\ &= \frac{1}{at} ([\sin(tx)]_0^a - i[\cos(tx)]_0^a) \\ &= \frac{1}{at} (\sin(at) - i \cos(at) + i) \\ &\stackrel{i^2=-1}{=} \frac{1}{at} \left(\cos(at) + i \sin(at) + i^2 \right) = \frac{e^{iat} - 1}{ait} \end{aligned}$$

Unfortunately, it is not always the case that the complex integrals can be computed easily using real-valued integration. In many instances contour integrals $\int_{\gamma} f(z) dz$ need to be used in order to compute $\int_{\mathbb{R}} f(x) dx$. If you know about contour integration this is a good exercise, otherwise it is a good reason to learn about contour integration!



Example 3.4.12.

$$\varphi_{\mathcal{N}(\mu,\sigma^2)}(t) = e^{i\mu t} e^{-\frac{\sigma^2}{2}t^2}, \quad t \in \mathbb{R},$$

and

$$\varphi_{\text{Exp}(\lambda)}(t) = \frac{\lambda}{\lambda - it}, \quad t \in \mathbb{R}.$$

Apart from being nice to compute with there must be a deeper reason to study characteristic functions. The fundamental theorem on characteristic functions states that the complex exponential functions are separating for $\mathcal{M}_1(\mathbb{R}^d)$ or, formulated in the language of random variables, the law of a random variable is uniquely determined by the characteristic function. Indeed, this is the justification for the name, the law is characterised by the characteristic function.



Theorem 3.4.13. (i) Every measure $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ is uniquely determined by all integrals

$$\int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mu(x), \quad t \in \mathbb{R}^d.$$

(ii) The law of a random vector X is determined by its characteristic function, i.e. if $\varphi_X = \varphi_Y$ then $X \sim Y$.

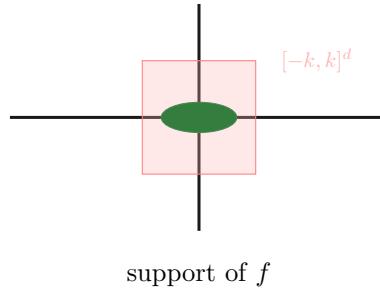
Of course one might ask why one might be interested in computing moments (for bounded random variables) or Laplace transformations (for non-negative random variables) if there is one general theorem that states that characteristic functions uniquely determine all real random variables. This is because it is easier to compute moments or Laplace transformations these are the preferred tools in the special situations when they are sufficiently powerful.

Proof. Since $C_c(\mathbb{R}^d)$ is separating by Proposition 3.1.15 it suffices to prove that the complex exponentials are dense in $C_c(\mathbb{R}^d)$. Then we can proceed similarly (but with an extra trick) to the proof of Corollary 3.4.4.

Let us fix two measures $\mu_1, \mu_2 \in \mathcal{M}_1(\mathbb{R}^d)$, $f \in C_c(\mathbb{R}^d)$, $\varepsilon > 0$, and fix $k \in \mathbb{N}$ such that

- f vanishes outside of the box $[-k, k]^d$,
- $\mu_i(\mathbb{R}^d \setminus [-k, k]^d) < \varepsilon$ or $i = 1, 2$.

The first can be achieved as f vanishes outside of a compact set, the second can be achieved using continuity of measures or Proposition 3.1.12.⁵ Next we define



$$g_m: \mathbb{R}^d \rightarrow \mathbb{C}, \quad g_m(x) = e^{i\langle \frac{4\pi m}{3k}, x \rangle}, \quad m, x \in \mathbb{Z}^d,$$

and C^k as the algebra of finite linear combinations of the g_m . To apply Stone-Weierstraß we restrict C^k to the compact set $[-k, k]^d$.



Check that the restriction $C^k_{|[-k,k]^d} \subseteq C_b([-k, k]^d, \mathbb{C})$ is an algebra that separates points of $[-k, k]^d$ (use the unit vectors $m = e_i$, the Euler formula and look at the sine/cosine functions with period $\frac{3}{2}k$ on $[-k, k]$). The strange choice of constants in the exponentials ensure that all points are separated.

Using the Stone-Weierstraß theorem we see that the restriction is dense in $C_b([-k, k]^d, \mathbb{C})$ and, in particular, also dense in the subset $C_b([-k, k]^d, \mathbb{R})$. Hence, there is some $g \in C^k$ with

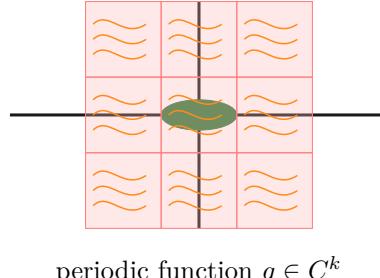
$$\sup_{x \in [-k, k]^d} |f(x) - g(x)| < \varepsilon.$$

⁵Koordinatensystem etwas kürzer

In order to bound the difference also outside of $[-k, k]^d$ it is crucial to realize that all g_m are periodic (compare the picture for what this means) with period $\frac{3k}{2}$. For $n \in \mathbb{Z}^d$ they satisfy

$$g_m\left(x + \frac{3k}{2}n\right) = e^{i\langle \frac{4\pi m}{3k}, x \rangle} \cdot e^{i\langle \frac{4\pi m}{3k}, \frac{3k}{2}n \rangle} = e^{i\langle \frac{4\pi m}{3k}, x \rangle} \cdot \underbrace{e^{i2\pi\langle m, n \rangle}}_{=1} = g_m(x),$$

as $e^{i2\pi l} = 1$ for all $l \in \mathbb{Z}$. Hence, the entire family C^k is periodic! Functions are schematically represented in the picture.⁶ Using the periodicity allows us to bound g on \mathbb{R}^d by comparing



with the values in $[-\frac{3k}{4}, \frac{3k}{4}]^d$, this is one full periodic block, where g is close to f :

$$\sup_{x \in \mathbb{R}^d} |g(x)| = \sup_{x \in [-\frac{3k}{4}, \frac{3k}{4}]^d} |g(x)| \leq \sup_{x \in [-\frac{3k}{4}, \frac{3k}{4}]^d} |g(x) - f(x)| + \sup_{x \in [-\frac{3k}{4}, \frac{3k}{4}]^d} |f(x)| \leq \varepsilon + \|f\|_\infty.$$

Note that this argument would not work if we approximated f on $[-k, k]^d$ for instance through polynomials (they grow to infinity at infinity) or any other dense set of continuous functions that is not bounded on \mathbb{R}^d , periodicity is key to control g outside of $[-k, k]^d$! Putting everything together, the above thoughts yield, please compare the proof of Corollary 3.4.4,

$$\begin{aligned} & \left| \int_{\mathbb{R}^d} f d\mu_1 - \int_{\mathbb{R}^d} f d\mu_2 \right| \\ & \leq \left| \int_{\mathbb{R}^d} f d\mu_1 - \int_{\mathbb{R}^d} g d\mu_1 \right| + \underbrace{\left| \int_{\mathbb{R}^d} g d\mu_1 - \int_{\mathbb{R}^d} g d\mu_2 \right|}_{=0, \text{ by assumption}} + \left| \int_{\mathbb{R}^d} g d\mu_2 - \int_{\mathbb{R}^d} f d\mu_2 \right| \\ & \leq \int_{[-k, k]^d} |f - g| d\mu_1 + \int_{([-k, k]^d)^c} |f - g| d\mu_1 + 0 + \int_{[-k, k]^d} |f - g| d\mu_2 + \int_{([-k, k]^d)^c} |f - g| d\mu_2 \\ & \leq \varepsilon \mu_1(\mathbb{R}) + \mu_1(([-k, k]^d)^c) \cdot (\varepsilon + 2 \|f\|_\infty) + \varepsilon \mu_2(\mathbb{R}) + \mu_2(([-k, k]^d)^c) \cdot (\varepsilon + 2 \|f\|_\infty) \\ & \leq 2\varepsilon + 2\varepsilon(\varepsilon + 2 \|f\|_\infty). \end{aligned}$$

For the inequalities we used that μ_i are probability measures, $|f - g| \leq |f| + |g| \leq \|f\|_\infty + \varepsilon + \|f\|_\infty$, and that μ_i only have mass at most ε outside of the chosen box. Since ε was arbitrary we proved $\int_{\mathbb{R}^d} f d\mu_1 = \int_{\mathbb{R}^d} f d\mu_2$ for all $f \in C_c(\mathbb{R}^d)$. But since $C_c(\mathbb{R}^d)$ separates measures we get $\mu_1 = \mu_2$. (ii) follows from the first claim applied to the laws \mathbb{P}_X as $\varphi_X(\lambda) = \varphi_{\mathbb{P}_X}(\lambda)$. \square

Here is a little application of the theorem that is useful to get a feeling of how to use it.



Use the uniqueness theorem to prove the converse of (iii) in Lemma 3.4.9. If a characteristic function only takes real values, then the random variable is symmetric.

Before discussing further applications and weak convergence in \mathbb{R}^d let us quickly discuss the method. The main point of the argument was to approximate continuous functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$

⁶Koordinatensystem zu kurz, gruen etwas groesser machen

with compact support. It looks strange that for the approximation we use function $g : \mathbb{R}^d \rightarrow \mathbb{C}$ and not functions $g : \mathbb{R}^d \rightarrow \mathbb{R}$. Since f is real-valued the definition of the complex norm yields

$$|f - g| = \sqrt{|\mathcal{R}e(f) - \mathcal{R}e(g)|^2 + |\mathcal{I}m(f) - \mathcal{I}m(g)|^2} = \sqrt{|f - \mathcal{R}e(g)|^2 + |\mathcal{I}m(g)|^2} \geq |f - \mathcal{R}e(g)|.$$

Hence, since we can approximate with complex g , we could even better approximate with the real-valued functions $\mathcal{R}e(g)$. Keeping in mind the Euler formula $e^{ix} = \cos(x) + i \sin(x)$ it becomes clear that real-parts of all complex linear combinations of the g_m are so-called real-valued trigonometric polynomials⁷

$$\sum_{n=1}^N (\alpha_n \cos(\langle m_n, x \rangle) + \beta_n \sin(\langle m_n, x \rangle)), \quad \alpha_n, \beta_n \in \mathbb{R}, m_n \in \mathbb{R}^d.$$

In fact, it is a well-known fact that the real-valued trigonometric polynomials are dense in $C_b([-k, k]^d, \mathbb{R})$! Hence, we could completely skip the complex numbers and instead work with cosine and sine moments. There are three main reasons why it is preferable to use the complex numbers even though it seems more complicated:

- Since $\mathbb{E}[e^{i\langle t, X \rangle}] = \mathbb{E}[\cos(\langle t, X \rangle)] + i\mathbb{E}[\sin(\langle t, X \rangle)]$ there is absolutely no practical advantage of working with cosine and sine moments instead of the characteristic function!
- There are wonderful tricks from complex analysis to compute the complex integral $\mathbb{E}[e^{i\langle t, X \rangle}]$. Just try to compute $\mathbb{E}[\cos(tX)]$ for $X \sim \mathcal{N}(0, 1)$ and you will realize that working with the exponential is a good idea!
- How do we prove that the trigonometric polynomials are dense? Exactly the way we proceeded in the proof, by checking that complex exponentials are dense via Stone-Weierstraß and then taking the real parts. A direct proof is more complicated as one needs to use the additivity theorems of cosine and sine to derive the algebra properties. This is simpler for the complex exponential since both additivity theorems together give the simple multiplication property of the exponential.

In many examples the uniqueness theorem is applied as follows. In order to identify the distribution of a random variable (such as a sum of other random variables) one tries to compute the characteristic function. If that characteristic function is already known then the law can be identified using the theorem. We already know this trick for moment generating functions (if they exist) from Proposition ?? and the example below but always used the trick without knowing a proof of the uniqueness theorem for moment generating function (see Corollary 3.6.2 below). Here is another exercise to play with simple examples, Lemma 3.4.9 and the uniqueness theorem. Please use characteristic functions and the uniqueness theorem to check the following examples:



- If $Y \sim \mathcal{N}(0, 1)$, then $Y := \sigma X + Y \sim \mathcal{N}(\mu, \sigma^2)$.
- If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then
$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$
- If $X \sim \text{Poi}(\lambda)$ and $Y \sim \text{Poi}(\beta)$ are independent, then $X + Y \sim \text{Poi}(\lambda + \beta)$.

3.5 Weak convergence on $\mathcal{B}(\mathbb{R}^d)$

We can finally return to the question of weak convergence. Recall from Proposition 3.3.4 that a sequence of probability measures (μ_n) converges weakly to μ (equivalently a sequence of random variables (X_n) converges in distribution to X) if and only if

⁷The wording "polynomial" is motivated by writing $z^n = |z|e^{i\varphi n}$ in terms of cosine and sine functions (Euler formula) so that complex polynomials can be written as trigonometric sums with complex coefficients.

- (i) $\{\mu_n : n \in \mathbb{N}\}$ is tight,
- (ii) $\lim_{n \rightarrow \infty} \int_E f d\mu_n = \int_E f d\mu$ for some separating family $C \subseteq C_b(E)$ of $\mathcal{M}_1(E)$.

We have identified separating families in the previous section. It now remains to identify situations in which the tightness trivially holds (bounded and non-negative) and give a handable criterion to check the tightness (convergence of characteristic functions). We keep the order of the previous section and first deal with the simple cases:



Theorem 3.5.1. (i) Suppose that $\mu, \mu_1, \dots \in \mathcal{M}_1([a, b])$, then

$$\mu_n \xrightarrow{(w)} \mu, n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} \int_{[a,b]} x^m d\mu_n(x) = \int_{[a,b]} x^m d\mu(x), \quad \forall m \in \mathbb{N}_0.$$

(ii) Suppose that X, X_1, \dots are random variables with values in $[a, b]$, then

$$X_n \xrightarrow{(d)} X, n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} \mathbb{E}[X_n^m] = \mathbb{E}[X^m], \quad \forall m \in \mathbb{N}_0.$$

Proof. (i) The necessity of convergence of the integrals follows as x^m is bounded and continuous on $[a, b]$. The sufficiency is also simple as $([a, b], |\cdot|)$ is compact. Hence, every sequence of measures is automatically tight (choose $K = [a, b]$ in the definition of tightness). According to Proposition 3.3.4 only convergence of integrals needs to be checked for a separating family. Since the family of monomials $C = \{x^m : m \in \mathbb{N}_0\}$ is separating according to Theorem 3.4.5 the proof is complete.
(ii) follows from (i) by the definition of convergence in distribution with the measures $\mu_n := \mathbb{P}_{X_n}$ using that $\mathbb{E}[X_n^m] = \int x^m d\mathbb{P}_{X_n}(x)$. \square

The situation is very similar for non-negative (or non-positive) sequences of random variables. In contrast to bounded sequences we now need convergence of the exponential moments (Laplace transformation):



Theorem 3.5.2. (i) Suppose that $\mu, \mu_1, \dots \in \mathcal{M}_1([0, \infty))$, then

$$\mu_n \xrightarrow{(w)} \mu, n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} \int_{[0,\infty)} e^{-\lambda x} d\mu_n(x) = \int_{[0,\infty)} e^{-\lambda x} d\mu(x), \quad \forall \lambda > 0.$$

(ii) Suppose that X, X_1, \dots are non-negative random variables, then

$$X_n \xrightarrow{(d)} X, n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} L_{X_n}(\lambda) = L_X(\lambda), \quad \forall \lambda > 0.$$

Proof. (i) We argue as in the proof of Theorem 3.4.6. The necessity is clear as the exponentials are bounded and continuous on $[0, \infty)$. For the sufficiency we argue with compactification. Compactifying $[0, \infty]$ and extending the measures trivially to measures $\bar{\mu}_n$ yields a sequence of measures in a compact space. As in the previous proof the sequence is automatically tight (choose $K = [0, \infty]$). According to Proposition 3.3.4 the weak convergence is equivalent to convergence of all integrals against a separating family. Since the family of exponentials $C := \{f_\lambda : \lambda > 0\}$ is separating according to Theorem 3.4.6 and

$$\int_{[0,\infty]} f_\lambda d\bar{\mu}_n(x) = \int_{[0,\infty)} f_\lambda d\mu_n(x)$$

the proof is complete.

(ii) follows from (i) by the definition of convergence in distribution with the measures $\mu_n := \mathbb{P}_{X_n}$ using that $L_{X_n}(\lambda) = \int e^{-\lambda x} d\mathbb{P}_{X_n}(x)$. \square

The most general theorem is **Lévy's continuity theorem**. Without any further assumption weak convergences is equivalent to convergence of the characteristic functions.

Lecture 17



Theorem 3.5.3. (i) Suppose that $\mu, \mu_1, \dots \in \mathcal{M}_1(\mathbb{R}^d)$, then

$$\mu_n \xrightarrow{(w)} \mu, n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} \varphi_{\mu_n} = \varphi_\mu(t), \quad \forall t \in \mathbb{R}^d.$$

(ii) Suppose that X, X_1, \dots are \mathbb{R}^d -valued random variables, then

$$X_n \xrightarrow{(d)} X, n \rightarrow \infty \Leftrightarrow \lim_{n \rightarrow \infty} \varphi_X(t) = \varphi_X(t), \quad \forall t \in \mathbb{R}^d.$$

Proof. As earlier we only prove (i) as (ii) follows directly using $\mu_n = \mathbb{P}_{X_n}$.

" \Rightarrow ": Since $g = e^{i\langle t, \cdot \rangle} = \cos(\langle t, \cdot \rangle) + i \sin(\langle t, \cdot \rangle)$ and $\cos, \sin \in C_b(\mathbb{R}^d)$ the pointwise convergence follows from the definition of weak convergence.

" \Leftarrow ": We will actually prove a stronger statement than claimed, the so-called **general continuity theorem**:



Suppose $\lim_{n \rightarrow \infty} \varphi_{\mu_n}(t) = f(t)$, for all $t \in \mathbb{R}^d$, where $f: \mathbb{R}^d \rightarrow \mathbb{C}$ is continuous at 0.

Then there is a probability measure Q on $\mathcal{B}(\mathbb{R}^d)$ with $f = \varphi_Q$ and $\mu_n \xrightarrow{(w)} Q, n \rightarrow \infty$.

Once we proved the general continuity theorem we immediately get the special continuity theorem by choosing $f = \varphi_\mu$ which is continuous at 0 by dominated convergence ($|e^{i\langle t, X \rangle}| = 1$).

Comparing with the proofs of the previous two theorems the strategy is as follows:

- deduce tightness of (μ_n) ,
- find a separating sequence for which the integrals converge,

because then the claim follows from Proposition 3.3.4. Choosing $C := \{e^{i\langle t, x \rangle} : t \in \mathbb{R}\}$ we already proved in Theorem 3.4.13 that C is separating. Hence, we need to prove that the assumed pointwise convergence of φ_{μ_n} implies tightness of (μ_n) .

To do so let us first check that without loss of generality we may assume $d = 1$. If $\pi_j(x) = x_j$ denotes the projection on the j th coordinate, then we define $\mu_n^j := \mu_n \circ \pi_j^{-1}$, the push-forward of the measures on the coordinates. Their characteristic functions $\varphi_{\mu_n^j}$ can be expressed through the characteristic functions of μ_n :

$$\varphi_{\mu_n^j}(t) = \int_{\mathbb{R}} e^{ixt} d\mu_n^j(x) = \int_{\mathbb{R}^d} e^{i\langle x, te_j \rangle} d\mu_n(x) = \varphi_{\mu_n}(te_j), \quad t \in \mathbb{R}.$$

Hence, the assumed pointwise convergence of φ_{μ_n} implies the pointwise convergence of all $\varphi_{\mu_n^j}$. If now we can prove that this implies tightness for all $(\mu_n^j)_{n \in \mathbb{N}}$, then the tightness of $(\mu_n)_{n \in \mathbb{N}}$ follows. The last claim can for instance be shown as follows: If K_1, \dots, K_d are compact sets such that $\sup_{n \in \mathbb{N}} \mu_n^j(K_j^c) < \frac{\varepsilon}{d}$ then the compact set $K_1 \times \dots \times K_d$ does the job for (μ_n) . This follows from subadditivity as $(K_1 \times \dots \times K_d)^c \subseteq \bigcup_{j=1}^d \pi_j^{-1}(K_j^c)$. From now on we assume $d = 1$ and that $(\varphi_{\mu_n})_{n \in \mathbb{N}}$ converges pointwise to a function f that is continuous at 0.

First recall that

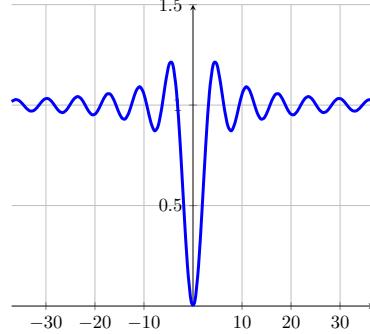
$$\varphi_\mu(t) = \int e^{itx} d\mu(x) = \int \cos(tx) d\mu(x) + i \int \sin(tx) d\mu(x)$$

We want to relate tightness (probabilities) to integrals (characteristic functions). Usually we took closed sets A and approximated $\mathbf{1}_A$ with f_A^ε . Here we use a different trick by estimating suitable

indicators from above by the sine/cosine functions. Define $h: \mathbb{R} \rightarrow [0, \infty)$ as

$$h(x) = \begin{cases} 1 - \frac{\sin(x)}{x} & : x \neq 0 \\ 0 & : x = 0 \end{cases},$$

which is non-negative and continuous on \mathbb{R} (Analysis):



plot of h

Now define

$$\alpha := \inf\{h(x): |x| \geq 1\} = 1 - \sin(1) > 0$$

so that $\frac{h(x)}{\alpha} \geq 1$ for $|x| > 1$. Then, for $k \geq 1$,

$$\begin{aligned} \mu_n([-k, k]^c) &\stackrel{\text{mon.}}{\leq} \int_{[-k, k]^c} \frac{1}{\alpha} h\left(\frac{x}{k}\right) d\mu_n(x) \\ &\leq \frac{1}{\alpha} \int_{\mathbb{R}} h\left(\frac{x}{k}\right) d\mu_n(x) \\ &= \frac{1}{\alpha} \int_{\mathbb{R}} \underbrace{\left(\int_0^1 \left(1 - \cos\left(\frac{tx}{k}\right)\right) dt \right)}_{=1-\sin\left(\frac{x}{k}\right)} d\mu_n(x) \\ &\stackrel{\text{Fubini}}{=} \frac{1}{\alpha} \int_0^1 \int_{\mathbb{R}} \left(1 - \cos\left(\frac{tx}{k}\right)\right) d\mu_n(x) dt \\ &= \frac{1}{\alpha} \int_0^1 \left(1 - \mathcal{Re}\left(\varphi_{\mu_n}\left(\frac{t}{k}\right)\right)\right) dt. \end{aligned}$$

Now we use dominated convergence to obtain

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mu_n([-k, k]^c) &\leq \frac{1}{\alpha} \lim_{n \rightarrow \infty} \int_0^1 \left(1 - \mathcal{Re}\left(\varphi_{\mu_n}\left(\frac{t}{k}\right)\right)\right) dt \\ &\stackrel{\text{DCT}}{=} \frac{1}{\alpha} \int_0^1 \lim_{n \rightarrow \infty} \left(1 - \mathcal{Re}\left(\varphi_{\mu_n}\left(\frac{t}{k}\right)\right)\right) dt \\ &= \frac{1}{\alpha} \int_0^1 \left(1 - \mathcal{Re}\left(f\left(\frac{t}{k}\right)\right)\right) dt. \end{aligned}$$

In the last step we have used that convergence of a sequence of complex numbers implies convergence of real- and imaginary-parts. To deduce the tightness we use the continuity of f . The continuity of f implies continuity of $\mathcal{Re}(f)$ at 0. In $\varepsilon - \delta$ formalism this means that for every $\varepsilon > 0$ there is some $\delta > 0$ such that

$$\left|0 - \frac{t}{k}\right| < \delta \quad \Rightarrow \quad \left|1 - \mathcal{Re}\left(f\left(\frac{t}{k}\right)\right)\right| = \left|\mathcal{Re}(f(0)) - \mathcal{Re}\left(f\left(\frac{t}{k}\right)\right)\right| < \alpha\varepsilon.$$

Hence, if we chose $k > \frac{1}{\delta}$, then the integrand is bounded by $\alpha\varepsilon$ for all $t \in [0, 1]$. With this choice of k we obtain

$$\limsup_{n \rightarrow \infty} \mu_n([-k, k]^c) < \varepsilon.$$

The tightness can be deduced as follows. For fixed $\varepsilon > 0$ we can find a compact set $[-K, K]$ and $N \in \mathbb{N}$ such that $\sup_{n \geq N} \mu_n([-K, K]^c) < \varepsilon$. Since single measures on $\mathcal{B}(\mathbb{R})$ are always tight we can also find compact sets $[-K_i, K_i]$ such that $\mu_i([-K_i, K_i]^c) < \varepsilon$ for $i = 1, \dots, N - 1$. If now we chose $M := \max\{K_1, \dots, K_{N-1}, K\}$, then $\sup_{n \in \mathbb{N}} \mu_n([-M, M]^c) < \varepsilon$. Hence, the sequence (μ_n) is tight.

Now we can summarize: Under the assumption of the general continuity theorem we proved that (μ_n) is tight. Using Prohorov, there is a weakly converging subsequences (μ_{n_k}) with some limit Q . But then the characteristic functions $\varphi_{\mu_{n_k}}$ converge to φ_Q . Since we also assumed that the characteristic functions of (μ_n) converge towards f it follows that $f = \varphi_Q$. \square

We finish the section with a short discussion of the usefulness of the generalised version of the continuity theorem. Let us recall the different ways of characterising random variables:

$$\text{random variable } X \Leftrightarrow \text{law } \mathbb{P}_X \Leftrightarrow \text{CDF } F_X.$$

In fact, just as CDFs are functions with a set of properties one can ask if also characteristic functions are functions with a set of axiomatic properties that uniquely characterise all random variables:

$$\text{random variable } X \Leftrightarrow \text{law } \mathbb{P}_X \Leftrightarrow \text{CDF } F_X \Leftrightarrow \text{characteristic function } \varphi_X$$

Indeed, this is the case but unfortunately the appearing property of positive definiteness is hard to check:



Theorem 3.5.4. (Bochner)

A function $\varphi: \mathbb{R}^d \rightarrow \mathbb{C}$ is the characteristic function of a random vector (or a probability measure on $\mathcal{B}(\mathbb{R}^d)$) if and only if

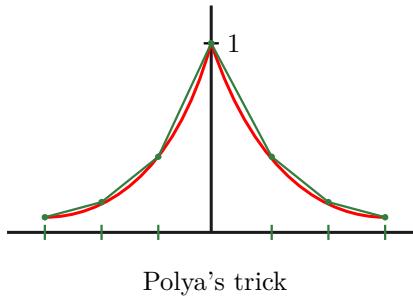
- $\varphi(0) = 1$
- φ is continuous at 0
- φ is **positive semidefinite**, i.e.

$$\sum_{k,l=1}^n y_k \bar{y}_l \varphi(t_k - t_l) \geq 0, \quad \forall n \in \mathbb{N}, t_i \in \mathbb{R}^d, y_i \in \mathbb{C}.$$

Proof. " \Rightarrow ": Suppose $\varphi(t) = \mathbb{E}[e^{itX}]$ for some random variable X . Then $\varphi(0) = 1$ is clear, continuity at 0 follows from dominated convergence, and

$$\begin{aligned} \sum_{k,l=1}^n y_k \bar{y}_l \varphi(t_k - t_l) &= \sum_{k,l=1}^n y_k \bar{y}_l \int_{\mathbb{R}^d} e^{i\langle x, t_k - t_l \rangle} d\mu(x) \\ &= \int_{\mathbb{R}^d} \sum_{k,l=1}^n y_k e^{i\langle x, t_k \rangle} \overline{y_l e^{i\langle x, t_l \rangle}} d\mu(x) \\ &= \int_{\mathbb{R}^d} \left| \sum_{k=1}^n y_k e^{i\langle x, t_k \rangle} \right|^2 d\mu(x) \geq 0. \end{aligned}$$

" \Leftarrow ": Too hard \square



There are many classes of characteristic functions that can be understood more directly using the general version of the continuity theorem. If for some reason we have a sequence of characteristic functions φ_n that converges pointwise to a function φ that is continuous at 0, then φ is a characteristic function. For all even continuous functions $\varphi: \mathbb{R} \rightarrow [0, 1]$ with $\varphi(0) = 1$ that are convex on $[0, \infty)$ this can be done (Pólya's theorem). One can in fact construct simple explicit random variables which characteristic functions φ_n that are approximations of such such functions and converge pointwise to φ .

The most prominent class of distributions that is defined using Polya's theorem are the stable laws:

Example 3.5.5. For $\alpha \in (0, 1]$ and $\lambda > 0$ the function $\varphi_\alpha(t) = e^{-\lambda|t|^\alpha}$ is even and convex on $[0, \infty)$. By Pólya's theorem this is a characteristic function of a random variable X . The random variable is called α -stable. In fact, the assumption $\alpha \in (0, 1]$ is only needed to apply Pólya's theorem, actually φ_α is a characteristic function if and only if $\alpha \in [0, 2]$. All such random variables are called α -stable and generalise the Gaussian distribution which appears for $\alpha = 2$.

The interesting point about the stable laws is the simple form of their characteristic functions whereas the densities (they exist!) are not simple at all.

Lecture 18

3.6 Applications

Recall from Theorem ?? the relation $\mathbb{E}[X^n] = \frac{\mathcal{M}_X^{(n)}(0)}{n!}$ between derivatives of the moment generating function $\mathcal{M}_X(t) = \mathbb{E}[e^{tX}]$ and the moments. The relation is useful as it allows to compute moments easily for a couple of examples (such as the Gaussian). In principle the idea is very simple, here for the first moment:

$$\mathcal{M}'_X(t) = \frac{d}{dt} \mathbb{E}[e^{tX}] = \mathbb{E}\left[\frac{d}{dt} e^{tX}\right] = \mathbb{E}[X e^{tX}]$$

and then plugging-in $t = 0$. What makes the proof tricky is the interchange of differentiation and expectation for which dominated convergence needs to be applied in the right way and this forced us to assume finiteness of \mathcal{M}_X in some interval $(-\varepsilon, \varepsilon)$. Since finiteness of $\mathcal{M}_X(t)$ means existence of exponential moments the theorem looks much better than it is, the assumption is extremely strong! We will now repeat the same story using the complex exponential. Before doing so we should first say a word on differentiation for functions $f: \mathbb{R} \rightarrow \mathbb{C}$. This is either defined by interpreting \mathbb{C} as \mathbb{R}^2 and then the rewriting the derivative (a vector) in terms of polar coordinates or directly by taking limits in \mathbb{C} : $\frac{d}{dt} f(t) = f'(t) = \lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h}$. Writing

$$\lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h} = \lim_{h \rightarrow 0} \frac{\Re f(t+h) - \Re f(t)}{h} + i \lim_{h \rightarrow 0} \frac{\Im f(t+h) - \Im f(t)}{h}$$

shows that both approaches give exactly the same. As always it is important to keep in mind that we can freely ignore the formal definition and just compute in \mathbb{C} as we are used to in \mathbb{R} with the convention $i^2 = -1$. Most importantly, it holds that $\frac{d}{dt} e^{itx} = ixe^{itx}$. Higher order

derivatives are defined recursively and denoted as usually by $f^{(n)}$. Now suppose, as for moment generating functions, the differentiation can be switched into the expectation, then we should get $\varphi_X^{(n)}(t) = \mathbb{E}[i^n X^n e^{itX}]$ so that plugging-in $t = 0$ gives again a moment formula $\mathbb{E}[X^n] = \frac{\varphi_X^{(n)}(0)}{i^n}$. Here again the magic of the complex exponential occurs. It is equally powerful in terms of what one can get from it but it is much more friendly because it is bounded. In essence the following theorem shows how to replace the real-exponential (moment generating function) by the complex-exponential (characteristic function) to obtain the same kind of results with minimal assumptions on the random variable.


Theorem 3.6.1. (Moments of X and differentiability of φ_X)

Let X be a real-valued random variable with characteristic function φ_X .

- (i) If $\mathbb{E}[|X^n|] < \infty$, then φ_X is n -times continuously differentiable with

$$\varphi_X^{(n)}(t) = \mathbb{E}[e^{itX} i^n X^n], \quad t \in \mathbb{R}.$$

In particular, $\mathbb{E}[X^n] = \frac{\varphi_X^{(n)}(0)}{i^n}$.

- (ii) If $\mathbb{E}[|X^n|] < \infty$ for some $n \in \mathbb{N}$, then φ_X satisfies the Taylor approximation

$$\varphi_X(t) = \sum_{k=0}^n \frac{i^k \mathbb{E}[X^k]}{k!} t^k + h_n(t) t^n, \quad t \in \mathbb{R}, \quad (3.6)$$

with a residual term satisfying $\lim_{t \rightarrow 0} h_n(t) = 0$.

- (iii) If all moments are finite and $\lim_{k \rightarrow \infty} \frac{t^k \cdot \mathbb{E}[|X^k|]}{k!} = 0$ for some $t \in \mathbb{R}$, then $\lim_{n \rightarrow \infty} h_n(t) = 0$. In particular, $\varphi_X(t)$ has the power series representation

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{i^k \mathbb{E}[X^k]}{k!} t^k.$$

The power series representation of φ_X is not surprising at all. Writing the complex exponential as a power series and exchanging freely expectation and infinite sum this nothing but

$$\varphi_X(t) = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{i^k t^k X^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{i^k \mathbb{E}[X^k]}{k!} t^k.$$

Of course, the interchange is non-trivial and there are essentially two ways to go. Either, arguing as in the proof of Theorem ?? one takes the limit of the partial sums, justifies dominated convergence, and then gets the moment formula by differentiating the power series, or, as we argue below, one first identifies the derivatives and then refers to Taylor's theorem to derive the power series.

Proof. In order to switch differentiation and expectation we will use the differential quotient and use dominated convergence to justify the change of expectation and limit in ε . We can do this since the differences of complex exponentials have useful bounds. Let us first check the basic estimate $|e^{itx} - e^{isx}| \leq |t - s| \cdot |x|$. If $s < t$ and $x > 0$, then expanding the exponential and using the Euler formula yields

$$\begin{aligned} |e^{itx} - e^{isx}| &= |e^{is\frac{x}{2}}| \cdot |e^{it\frac{x}{2}}| \cdot |e^{i(t-s)\frac{x}{2}} - e^{-i(t-s)\frac{x}{2}}| \\ &= 2|\sin((t-s)x/2)| \\ &= 2 \left| \int_0^{(t-s)\frac{x}{2}} \cos(u) du \right|^{\cos| \leq 1} \leq |(t-s)x| \end{aligned}$$

and the other cases are treated similarly.

(i) The first derivative is now simple:

$$\lim_{h \rightarrow 0} \frac{\varphi_X(t) - \varphi_X(t+h)}{h} = \lim_{h \rightarrow 0} \mathbb{E}\left[\frac{e^{itx} - e^{i(t+h)x}}{h}\right] \stackrel{\text{DCT}}{=} \mathbb{E}\left[\lim_{h \rightarrow 0} \frac{e^{itx} - e^{i(t+h)x}}{h}\right] = \mathbb{E}[iX e^{itX}]$$

Changing limits and expectation was justified by the integrable (assumption) upper bound

$$\left| \frac{e^{itx} - e^{i(t+h)x}}{h} \right| \leq \frac{|h \cdot x|}{|h|} = |x|$$

which was justified above. Hence, $\varphi'_X(t) = \mathbb{E}[e^{itX} iX]$ exists. Inductively we proceed in exactly the same way to differentiate $\varphi_X^{(k)}(t)$:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\varphi_X^{(k)}(t) - \varphi_X^{(k)}(t+h)}{h} &= \lim_{h \rightarrow 0} \frac{\mathbb{E}[e^{itX} i^k X^k] - \mathbb{E}[e^{i(t+h)X} i^k X^k]}{h} \\ &= \lim_{h \rightarrow 0} \mathbb{E}\left[\frac{(e^{itX} - e^{i(t+h)X}) i^k X^k}{h}\right] \\ &\stackrel{\text{DCT}}{=} \mathbb{E}\left[\lim_{h \rightarrow 0} \frac{(e^{itX} - e^{i(t+h)X}) i^k X^k}{h}\right] = \mathbb{E}[i^{k+1} X^{k+1} e^{itX}] \end{aligned}$$

The interchange of limit and expectation is again justified by the upper bound

$$\frac{|(e^{itx} - e^{i(t+h)x}) i^k x^k|}{h} \leq |x| \cdot |x^k| = |x^{k+1}|.$$

The proof shows very clearly that φ_X can be differentiated as long as there are enough finite moments of X . But this is no additional assumption as otherwise the formula does not make sense anyways!

(ii) Since all derivatives at 0 are known from (i) the complex version of Taylor's theorem for $x_0 = 0$ gives

$$\varphi_X(t) = 1 + it\mathbb{E}[X] - \frac{1}{2}t^2\mathbb{E}[X^2] + \dots + \frac{i^n t^n \mathbb{E}[X^n]}{n!} + h_n(t)t^n,$$

with a remainder term satisfying $\lim_{t \rightarrow 0} h_n(t) = 0$.

(iii) We need to show that, for fixed t , the residual $h_n(t)$ vanishes as n tends to infinity. It is most practical to use the integral representation for $R_n(t) := h_n(t)t^n$:

$$|R_n(t)| = \left| \int_0^t \frac{\varphi_X^{(n+1)}(s)}{(n+1)!} (t-s)^n ds \right| \leq \int_0^t \frac{\mathbb{E}[|X^{n+1}|]}{(n+1)!} (t-s)^n ds = \frac{t^{n+1}}{(n+1)} \frac{\mathbb{E}[|X^{n+1}|]}{(n+1)!}$$

Here we used the formula from (i) and that $|e^{itX} i^n| = 1$. But then $h_n(t) \rightarrow 0$ for $n \rightarrow \infty$. \square

Just as we used the moment generating functions to compute moments for random variables with exponential moments we can also use the characteristic functions:



Compute the first few moments of $\mathcal{N}(0, 1)$, $\text{Poi}(\lambda)$, and $\text{Exp}(\lambda)$.

As an application we prove what is sometimes called the **method of moments** and, as a special case, we can finally give a proof of Theorem ??.



Corollary 3.6.2. (i) Let X be a real-valued random variable such that there



is a constant C with $\frac{1}{n}(\mathbb{E}[|X|^n])^{\frac{1}{n}} < C$ for all $n \in \mathbb{N}$. Then the law of X is uniquely determined by all its moments.

- (ii) In particular, if $\mathcal{M}_X(t) < \infty$ for $t \in (-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$, then \mathcal{M}_X uniquely determines the law of X .

The corollary is a significant extension of Theorem 3.4.5. If $|X|$ is bounded by some C , then $\mathbb{E}[|X^n|]$ is bounded by C^n so that the assumption of the corollary is clearly satisfied. The corollary states that also for other (rather special) random variables which moments increase slowly enough the same statement holds.

Proof. (i) For $|t| < \frac{1}{3C}$, using the Sterling formula, we have

$$\limsup_{n \rightarrow \infty} \frac{\mathbb{E}[|X|^n] \cdot |t|^n}{n!} \stackrel{??}{=} \limsup_{n \rightarrow \infty} \left(\mathbb{E}[|X|^n]^{\frac{1}{n}} |t| \frac{e}{n} \right)^n \sqrt{2\pi n} \leq \limsup_{n \rightarrow \infty} \left(\frac{e}{3} \right)^n \sqrt{2\pi n} = 0$$

Hence, for all $t \in (-\frac{1}{3C}, \frac{1}{3C})$ we can use Theorem 3.6.1 (iii) to express φ_X as a power series. Since the coefficients are the moments, φ_X is determined on $(-\frac{1}{3C}, \frac{1}{3C})$ through the moments. Since a power series is uniquely determined by the values on some interval φ_X is also uniquely determined on \mathbb{R} by all the moments. Finally, since the law of X is uniquely determined by φ_X according to Theorem 3.4.13 the proof is complete.

(ii) We argue as in the proof of Theorem ??:

$$\sum_{k=0}^{\infty} \frac{t^k \mathbb{E}[|X|^k]}{k!} \stackrel{\text{MCT}}{=} \mathbb{E}[e^{t|X|}] \leq \mathbb{E}[e^{tX}] + \mathbb{E}[e^{-tX}] = \mathcal{M}_X(t) + \mathcal{M}_X(-t) < \infty, \quad t \in (-\varepsilon, \varepsilon).$$

Now we can use the proof of (i) to finish the proof. \square

To finish the chapter with a spectacular theorem we give a proof of the central limit theorem. The proof we have given in Section ?? imposed the additional assumption $\mathbb{E}[|X_1|^3] < \infty$, hence, this is our first complete proof of the most important theorem of probability theory and statistics.



Theorem 3.6.3. (Central Limit Theorem)

Let X_1, X_2, \dots be iid with $\mu := \mathbb{E}[X_1]$ and $\sigma^2 := \mathbb{V}[X_1] < \infty$. Then

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{\sigma^2 n}} \xrightarrow{(d)} \mathcal{N}(0, 1), \quad n \rightarrow \infty.$$

Proof. Without loss of generality we may assume $\mu = 0$ and $\sigma^2 = 1$ as otherwise we can consider the standardised random variables $\tilde{X} := \frac{X-\mu}{\sigma}$. Writing $S_n = \sum_{k=1}^n X_k$, Lévy's continuity theorem we only need to prove that

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) \longrightarrow e^{-\frac{1}{2}t^2} = \varphi_{\mathcal{N}(0,1)}(t), \quad \forall t \in \mathbb{R}.$$

Using Lemma 3.4.9 and the iid assumption shows that we only need to prove

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \left(\varphi_{X_1} \left(\frac{t}{\sqrt{n}} \right) \right)^n \longrightarrow e^{-\frac{1}{2}t^2}, \quad n \rightarrow \infty.$$

The trick is to replace the exponential by the first two summands of its Taylor expansion (3.6) and then to use

$$\lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n} \right)^n = e^{-\frac{1}{2}t^2}, \quad t \in \mathbb{R}.$$

Do do this rigorously first check by a quick induction that

$$|u^n - v^n| \leq |u - v| \cdot n \cdot \max(|u|, |v|)^{n-1}, \quad \forall u, v \in \mathbb{C}.$$

Using that $|\varphi_{X_1}|, |1 - \frac{t^2}{2n}| \leq 1$ for n large enough then yields

$$\begin{aligned} \left| \left(1 - \frac{t^2}{2n}\right)^n - \left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n \right| &\leq \left| \left(1 - \frac{t^2}{2n}\right)^n - \left(1 - \frac{1}{2} \frac{t^2}{n} + h_2\left(\frac{t}{\sqrt{n}}\right) \frac{t^2}{n}\right)^n \right| \\ &\leq n \cdot \left| h_2\left(\frac{t}{\sqrt{n}}\right) \frac{t^2}{n} \right| \\ &= t^2 \cdot \left| h_2\left(\frac{t}{\sqrt{n}}\right) \right| \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

□

In many textbooks one can see the formulation

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sum_{k=0}^n X_k - n \cdot \mu}{\sqrt{n\sigma^2}} \in [a, b]\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx, \quad \forall a < b,$$

which holds due to Portemanteau because $\mathbb{P}_{\mathcal{N}(0,1)}(\partial[a, b]) = 0$.

There are plenty of generalisations of the central limit theorem. Here is a multi-dimensional version that will be used for the proof of the Donsker theorem in the next section.



Definition 3.6.4. If Y is a d -dimensional random vector consisting of independent $\mathcal{N}(0, 1)$ random variables, $\mu \in \mathbb{R}^p$, and $B \in \mathbb{R}^{d \times p}$, then $X := \mu + BY$ is called a p -dimensional **Gaussian random vector**. In analogie to $\mathcal{N}(\mu, \sigma^2)$ we write $X \sim \mathcal{N}(\mu, \Sigma)$, where $\Sigma := BB^T$ is the so-called covariance matrix.

There are several equivalent definitions of Gaussian random vectors. Most prominently through the characteristic function involving a mean vector m and a symmetric positive semidefinite matrix Σ or by saying that all linear combinations of the entries are univariate Gaussian. All approaches lead to the same characteristic functions (see the next exercise) and therefore are equivalent.



If $X \sim \mathcal{N}(\mu, \Sigma)$, then

$$\varphi_X(t) = \exp\left(i\langle t, \mu \rangle - \frac{1}{2}\langle t, \Sigma t \rangle\right), \quad t \in \mathbb{R}^d \tag{3.7}$$

and

$$\mathbb{E}[X_k] = \mu_k, \quad \text{Cov}(X_i, X_j) = \Sigma_{i,j}.$$

If furthermore Σ is invertible, then X has the multivariate density

$$f(x) = \frac{1}{(2\pi)^{d/2} \det(\Sigma)} \exp\left(-\frac{1}{2}\langle x - \mu, \Sigma(x - \mu) \rangle\right), \quad x \in \mathbb{R}^d.$$

The exercise shows that the key quantities for Gaussian random vectors are the mean vector μ and the covariance matrix Σ . If a random vector is known to be Gaussian it is uniquely determined through the mean vector and all covariances of two-coordinates. In particular, coordinates of a Gaussian random vector are independent if and only if they are uncorrelated, a fact, which is wrong for non-Gaussian vectors.



Show the following multivariate version of the central limit theorem. If X_1, \dots is an iid sequence of random vectors with mean vector $\mu_k = \mathbb{E}[X_{1,k}]$ and covariance matrix $\Sigma_{i,j} = \text{Cov}(X_{1,i}, X_{1,j})$, then

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \mu) \xrightarrow{(d)} \mathcal{N}(\mu, \Sigma), \quad n \rightarrow \infty.$$

Note that differences, sums, and scalar multiplications are in the vector sense.

Kapitel 4

Brownian Motion

bla bla¹

4.1 Introduction to stochastic processes in continuous time

Stochastic processes in discrete time have been studied in Chapter ?? mostly from the point of view of sequences of random variables. In continuous time things escalate quickly if the index set is uncountable. In this section we lay the foundations of stochastic processes seen as path-valued random variables. This interpretation forces us to discuss quite a bit of measure theory. We will lay the foundations to study the Brownian motion and convergence of sequences of stochastic processes in the general frame-work of weak convergence in Polish spaces from the previous chapter.

Let us first recall the definition of a stochastic process. For a fixed probability space $(\Omega, \mathcal{A}, \mathbb{P})$ a collection $X = (X_t)_{t \in I}$ of random variables with values in a measurable space (E, \mathcal{E}) is called a stochastic process with state-space E . Most of the time we can assume that E is a Polish metric space with Borel σ -algebra $\mathcal{E} = \mathcal{B}(E)$ but always keep in mind the most important example $E = \mathbb{R}$. The index set I was assumed to be discrete in the discussion of martingales but can also be continuous here. The (random) mapping $t \mapsto X_t(\omega)$ is called a trajectory (or sample path) of X . While the study of discrete time stochastic processes was rather elementary (playing with stopping times and the martingale property) the study of continuous time processes will be more involved caused by the measure theory on path-space where the continuum appears twice (in space and time). An illustrative task is to answer (or properly define) the question if sample paths of stochastic processes are almost surely continuous or differentiable.

In order to develop the full theory of stochastic processes we will first extend the triology of measure theory behind random vectors, i.e. stochastic processes with $|I| < \infty$, from Section ??:

- define a σ -algebra $\mathcal{E}^{\otimes I}$ on the generic path space E^I ,
- understand all probability measures on $\mathcal{E}^{\otimes I}$
- reinterpret stochastic processes as path-valued random variables.

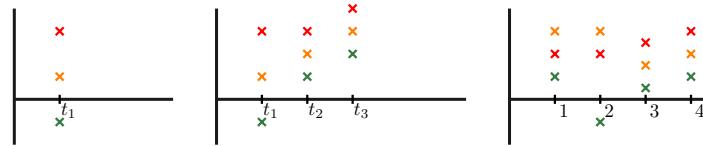
There is already quite a bit of pain involved in generalizations from finite I to uncountable I and even worse if trajectories are supposed to be continuous, which we study next.

4.1.1 Stochastic processes and the generic path space

The biggest hurdle towards a thorough understanding of continuous-time stochastic processes is the measure theory on path space. Even though the main interest of this course is processes with continuous sample paths (such as the Brownian motion) it is useful to first discuss measure theory

¹was intelligenteres hinschreiben

][[h]

A few elements of $\mathbb{R}^{\{t_1\}}$, $\mathbb{R}^{\{t_1, t_2, t_3\}}$, and $\mathbb{R}^{\mathbb{N}}$.

on generic functions $f : I \rightarrow E$ with absolutely no assumptions on f . In the previous chapter we enlarged $[0, \infty)$ to the one-point compactification $[0, \infty]$ in order to gain compactness. It is much less simple to see what is gained if continuous functions $C([0, \infty))$ are enlarged to all generic functions $\mathbb{R}^{[0, \infty)}$. There is only one very non-trivial reason, the Kolmogorov extension theorem that we prove below. Kolmogorov's extension theorem is a theorem that allows to construct measures on the natural σ -algebra on paths $\mathbb{R}^{\otimes [0, \infty)}$ which combined with the Kolmogorov-Chentsov theorem gives a possibility to construct measures on $C([0, \infty))$ and continuous-times stochastic processes. In order to pass that hurdle we carefully discuss all appearing objects and new ideas, and then pass on to the Brownian motion as an application of the strategy.

Before getting started it is useful to visualize and reinterpret the generic path space \mathbb{R}^I of all functions from I to \mathbb{R} for the simplest index sets:

- \mathbb{R}^I corresponds to \mathbb{R} if $|I| = 1$,
- \mathbb{R}^I corresponds to \mathbb{R}^d if $|I| = d$, by reinterpreting a vector as a mapping from d arbitrary time-points (the indices of the vector) to \mathbb{R} ,
- \mathbb{R}^I corresponds to the set of real-valued sequences for $I = \mathbb{N}$,

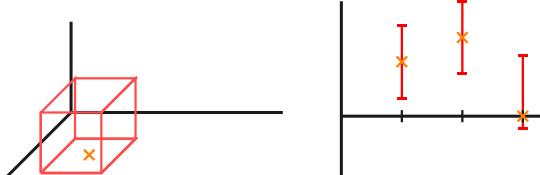
Hence, it is little surprising that we will always try to generalize ideas from \mathbb{R}^d to \mathbb{R}^I . Before starting the measure theory on path space the reader might want to have a quick look at the beginning of Section ?? to recall the steps σ -algebra, measures, random variables for \mathbb{R}^d as the next sections are similar but much harder in terms of notation.

(A) The path σ -algebra

Recall from Section ?? the notion of a product σ -algebra. If \mathcal{E} is a σ -algebra, then the d -fold product σ -algebra on E^d is defined as

$$\mathcal{E}^{\otimes d} := \mathcal{E} \otimes \cdots \otimes \mathcal{E} := \sigma(\{B_1 \times \cdots \times B_d : B_i \in \mathcal{E}\}).$$

In the sequel the appearing sets $B_1 \times \cdots \times B_d$ are called **boxes** and interpreted using the reinterpretation of E^d as mappings from d arbitrary time-points to E .



Reinterpretation of a three-dimensional box for arbitrary three time-pointes

If $\mathcal{E} = \sigma(\mathcal{A})$, then $\mathcal{E}^{\otimes d}$ can also be expressed as

$$\mathcal{E}^{\otimes d} = \sigma(\{B_1 \times \cdots \times B_d : B_i \in \mathcal{A}\}).$$

Most importantly, if $E = \mathbb{R}$ and \mathcal{A} is chosen as the set of intervals than the product σ -algebra is generated by rectangular sets. The aim is to introduce a generalization to infinite products, but what is a good notion for a (possibly uncountable) infinite product of sets? As usually in Mathematics the trick is to reduce as far as possible to the finite case:



Definition 4.1.1. A subset C of E^I is called **finitely generated cylinder set** if

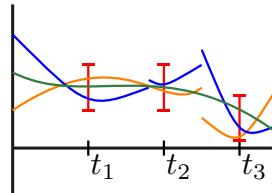
$$\begin{aligned} C &= \pi_J^{-1}(B_1 \times \cdots \times B_n) \\ &= \{f : I \rightarrow E \mid f(t_1) \in B_1, \dots, f(t_n) \in B_n\}, \end{aligned}$$

where $J = \{t_1, \dots, t_n\} \subseteq I$ and $B_1, \dots, B_n \in \mathcal{E}$. The projections

$$\pi_J(f) = (f(t_1), \dots, f(t_n))$$

evaluate a path at finitely many given time-points. We say C is spanned by the time-points t_1, \dots, t_d and the box $B_1 \times \cdots \times B_d \in \mathcal{E}^{\otimes n}$.

Before proceeding it is extremely important to gain a visual understanding of finitely generated cylinder sets. The picture to keep in mind is that of a box tied to the time-points with all possible functions $I \rightarrow E$ passing through the box. The set we are speaking about is not the box but that of all such functions!



The cylinder set is the set of all functions passing through the box tied at times from J

There are two quantities that have to be distinguished very carefully. The spanning box, which is a finite-dimensional measurable set, and the cylinder set which is a set of functions.



Definition 4.1.2. The **path σ -algebra** (or **product σ -algebra**) $\mathcal{E}^{\otimes I}$ on E^I is the smallest σ -algebra that contains all finitely generated cylinder sets. In the special case $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ one typically writes $\mathcal{B}(\mathbb{R}^d)$, $\mathcal{B}(\mathbb{R}^\infty)$, $\mathcal{B}(\mathbb{R}^{[0,T]})$ to abbreviate the notation and also speaks of the product σ -algebra.

To get an idea of the definition it is useful to compare the cases $|I| = d$ and $I = \mathbb{N}$ for $E = \mathbb{R}$. For the first the cylinder sets are just boxes. Some of the B_i will typically be equal to E , not imposing a restriction. The path σ -algebra is nothing but the well-known d -fold product $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}$ from Section ???. This motivates why sometimes we call path σ -algebras product σ -algebra also for infinite index sets I . The infinite product structure becomes more visible for $I = \mathbb{N}$, where the finitely generated cylinder sets are actually infinite products of the type $\mathbb{R} \times B_1 \times \cdots \times B_2 \times \mathbb{R} \times \cdots$ as the factors \mathbb{R} impose no restriction. While the infinite cartesian product makes sense for sequences the good way of writing an uncountable product with restrictions at finitely many time-points is the formalism using projections on finitely many time-points.

Here are two simple exercise that we will later need and that is very useful to get acquainted with the new definitions.



Show that the set of all finitely generated cylinder sets form a semiring. First draw some pictures and then try to formalize your findings using projections.

As always there are many different generators of the σ -algebra, not all are equally useful.



Suppose that $\mathcal{E} = \sigma(\mathcal{A})$. Check that

$$\begin{aligned}\mathcal{E}^{\otimes I} &= \sigma(\{\pi_{t_1, \dots, t_n}^{-1}(B_1 \times \dots \times B_n) : n \in \mathbb{N}, t_i \in J, B_i \in \mathcal{A}\}) \\ \mathcal{E}^{\otimes I} &= \sigma(\{\pi_i^{-1}(B) : i \in I, B \in \mathcal{E}\}), \\ \mathcal{E}^{\otimes I} &= \sigma(\{\pi_i^{-1}(B) : i \in I, B \in \mathcal{A}\}),\end{aligned}$$

and draw some examples of sets of paths in the generators of $\mathcal{E}^{\otimes I}$. Which of the generators of $\mathcal{E}^{\otimes I}$ is intersection stable?

If $|I| < \infty$ or I is countable not much more needs to be said, there are no bad surprises, everything works more or less similarly to $\mathcal{B}(\mathbb{R}^d)$. The only trouble is the usual difference between $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{B}(\mathbb{R}^d)$ that arises from the fact that countable set operations in a σ -algebra do not allow to approximate arbitrary subsets from \mathbb{R}^d using intervals/open sets/closed sets/etc. Most interesting sets of interest belong to $\mathcal{B}(\mathbb{R}^d)$, the generator is rich enough. The story is very different for uncountable index sets I because a second and more severe uncountability issue appears. The generator of $\mathcal{E}^{\otimes I}$ only uses single (or finitely many) points in time which is by far not enough to capture a lot of information about functions in time. It is instructive to compare with Example 1.2.5 where it was shown that the smallest σ -algebra containing singletons can only cover the countable sets. Hence, the finitely generated cylinder sets can at most cover countably generated cylinder with the meaning that a set of functions C is countably generated if there is a countable set of time-points $J \subseteq I$ such that ²

$$f \in C \Leftrightarrow f_{|J} \in B_1 \times \dots \times B_{|J|}$$

for some $B_i \in \mathcal{E}$. Equivalently, the sets C can be written as $C = \pi_J^{-1}(B_1 \times \dots \times B_{|J|})$ with the projection π_J on countably many time-points.



Proposition 4.1.3. The set \mathcal{V} of all countably generated cylinder sets is a σ -algebra on E^I that contains $\mathcal{E}^{\otimes I}$.

Proof. To check the three properties of a σ -algebra is straight forward.

- (i) Choosing J to be an arbitrary singleton and $B = E$ proves that $\Omega = E^I$ is countably generated.
- (ii) If C is generated at some countable set of times J , then C^c (the path that do not pass through the given sets at some time $t \in J$) can be written as union of the countably many cylinder sets that pass through B_t^c at time $t \in J$.

then also C^c is countably generated by the same time-points J using B^c .³

- (iii) Suppose C_1, \dots are generated by countable sets J_1, J_2, \dots and corresponding sets B_1, B_2, \dots . Then also the union $J := \cup_{k=1}^{\infty} J_k$ is countable and $\cup_{k=1}^{\infty} C_k$ is generated by J and⁴ $B := B_1 \times B_2 \times \dots \in \mathcal{E}^{\otimes J}$.

Since all finitely generated sets are also infinitely generated the generator of $\mathcal{E}^{\otimes I}$ is a subset of \mathcal{V} . Hence, $\mathcal{E}^{\otimes I} \subseteq \sigma(\mathcal{V}) = \mathcal{V}$ which is the claim. \square

The proposition has dramatic consequences! Almost all sets of paths that we might be interested in to study the behavior of paths $t \mapsto X_t(\omega)$ are **not** measurable in the path σ -algebra!

Example 4.1.4. Suppose $E = \mathbb{R}$ and $I = \mathbb{R}$, $I = [0, 1]$, or $I = [0, \infty)$.

- Singleton sets $\{f\}$ containing only a single path are not measurable.

²generator, letzter beweis!

³schnitt!

⁴noch sauber hinschreiben

- The set of all constant functions $\{f \equiv c \mid c \in E\}$ is not measurable.
- The set of non-negative functions $\{f : I \rightarrow E \mid f \geq 0\}$ is not measurable.
- The set of all continuous functions $\{f : I \rightarrow E \mid f \text{ continuous}\}$ is not measurable.
- The set of all differentiable functions $\{f : I \rightarrow E \mid f \text{ differentiable}\}$ is not measurable.

The reason is simple. If any of those sets was measurable in the σ -algebra $\mathbb{R}^{\otimes I}$ of paths, then there should be a countable set J of time-points that spans the set. If I is uncountable there is an index $i \notin J$. Now take a function f from the sets and add some value at i that destroys the defining property. Then the new function \tilde{f} is still in the set as the values on J are not changed. But then the set contains a function that violates the defining property which should not be the case. As a consequence we should be very careful with statements of the kind "P-almost all paths are continuous" with measures on the uncountable product σ -algebras on paths.

Since uncountable product σ -algebras are delicate one should ask if the decision to work with $\mathcal{E}^{\otimes I}$ was clever or non-sense. The surprising answer is clever as there is a natural way to construct measures on $\mathcal{E}^{\otimes I}$ and most delicate issues about zero sets and path properties can be resolved.

(B) Probability measures on paths

The aim of this section is to identify all probability measures on the product σ -algebra $\mathcal{E}^{\otimes I}$. For $E = \mathbb{R}$ and $|I| < \infty$ this has already been done in Section 1.4, we now generalize to arbitrary I and E as long as E is a Polish space. The strategy is not simple but can be compared to the theory of distribution functions for probability measures on $\mathcal{B}(\mathbb{R})$. Recall the approach from Section 1.4:

- (i) Downsize the information of a measure into something simple that uniquely characterizes the measure. In that case the simplest \cap -stable generator of $\mathcal{B}(\mathbb{R})$ was the set of all intervals and we defined $F(t) = \mathbb{P}((-\infty, t])$, $t \in \mathbb{R}$.
- (ii) Identify natural properties of the simpler object. We used continuity of measures to deduce monotonicity, limits at infinity, and right-continuity for F . Such functions were then called distribution functions.
- (iii) Reverse the game: Show that for all distribution function there is a corresponding probability measure (Carathéodory extension theorem)

The approach for measures on $\mathcal{E}^{\otimes I}$ is exactly the same.

- (i) Downsized \mathbb{P} by only considering events from the finite cylinder sets (which are \cap -stable). This leads to so-called finite dimensional distributions \mathbb{P}_J on $\mathcal{E}^{\otimes J}$ for all finite subsets $J \subseteq I$.
- (ii) Check that the family $\{\mathbb{P}_J : |J| < \infty\}$ of all finite dimensional distributions satisfies a natural property, consistency.
- (iii) Reverse the game: Show that for all consistent families of finite dimensional distributions there is a corresponding probability measures (Carathéodory consistency theorem).

The only trouble of the approach comes from the notation. The notation with cylinder sets and finite dimensional distributions is either very long (writing out all appearing cylinder sets) or very compact (abbreviating all cylinder sets). We will stick to the compact way but add figures with illustrations.

Here is the first step, the restrictions to the \cap -stable generator of finitely generated cylinder sets.



Definition 4.1.5. If \mathbb{P} is a probability measure on $\mathcal{E}^{\otimes I}$ and $J \subseteq I$ is finite, then

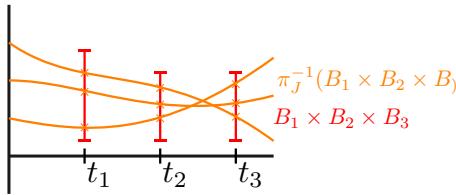


the unique measure on $\mathcal{E}^{\otimes J}$ defined by

$$\mathbb{P}_J(B_1 \times \cdots \times B_{|J|}) := \mathbb{P}(\pi_J^{-1}((B_1 \times \cdots \times B_{|J|})), \quad B_i \in \mathcal{E},$$

is called a **finite dimensional marginal distribution** of \mathbb{P} .

Here is a very important point to understand. On the one hand we have boxes (measurable sets of a finite-dimensional space) on the other hand we have cylinder sets of functions passing through the boxes at time-points J (measurable sets of an infinite-dimensional space). The finite dimensional marginals are the measures that we get by crossing out the functions and only keep the marginals at time J .



restricting the functions at $J = \{t_1, t_2, t_3\}$ to $B_1 \times B_2 \times B_3$

The set of finite dimensional marginals \mathbb{P}_J is much easier than \mathbb{P} as these are only measures on finite product σ -algebras $\mathcal{E}^{\otimes J}$. This reflects the analogy that distribution functions are much simpler than probability measures on $\mathcal{B}(\mathbb{R})$. Similar to distribution functions of measures on $\mathcal{B}(\mathbb{R})$ the finite dimensional distributions uniquely define \mathbb{P} :



Proposition 4.1.6. Every probability measure \mathbb{P} on $\mathcal{E}^{\otimes I}$ is uniquely defined through the family of all finite dimensional distributions $\{\mathbb{P}_J : |J| < \infty\}$.

Proof. By the uniqueness theorem 1.2.12, \mathbb{P} is uniquely defined on any \cap -stable generator of the product σ -algebra $\mathcal{E}^{\otimes I}$. We chose the set of all finitely generated cylinder sets

$$\mathcal{S} := \{\pi_J^{-1}(B_1 \times \cdots \times B_{|J|}) : J \subseteq I, |J| < \infty, B_i \in \mathcal{E}\},$$

which is a generator of $\mathcal{E}^{\otimes I}$ and clearly intersection stable. Here it is crucial to use the finitely generated sets and not the sets generated by all single time-points as the second generator of $\mathcal{E}^{\otimes I}$ is not \cap -stable! \square

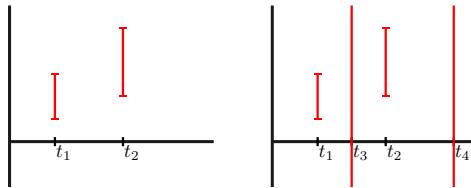
The next step in the analogy to distribution functions is to derive what will later turn out to be the right characterizing property of families of finite-dimensional distributions to be related to measures on $\mathcal{E}^{\otimes I}$.



Proposition 4.1.7. If \mathbb{P} is a probability measures on $\mathcal{E}^{\otimes I}$, then the family $\{\mathbb{P}_J : |J| < \infty\}$ of all finite dimensional marginal distributions fulfills the following property. If $J \subset J'$ are both finite subsets of I and $\pi_{J,J'}$ denotes the projection from E^J to $E^{J'}$, then

$$\mathbb{P}_J \circ \pi_{J,J'}^{-1} = \mathbb{P}_{J'}.$$

The proof is simple, only the notation is troubling. The projection $\pi_{J,J'}$ has the following meaning. Take a function indexed by J and ignore all values from $J \setminus J'$. More interestingly, the preimage under $\pi_{J,J'}$ of a box indexed by J' is the box indexed by J with E placed at the missing time-points from $J \setminus J'$. Here is a drawing:



$B_1 \times B_2$ and $\pi_{J,J'}^{-1}(B_1 \times B_2)$ for $J = \{t_1, \dots, t_4\}$ and $J' = \{t_1, t_2\}$

A family of consistent measures has the property that adding E s to a product does not change the measures corresponding to the increased time-set. We already know this situation from random vectors where we frequently used that

$$\mathbb{P}_{X_1}(1) = \mathbb{P}(X_1 \in A) = \mathbb{P}(X_1 \in A, X_2 \in \mathbb{R}) = \mathbb{P}_{X_1, X_2}(A \times \mathbb{R}),$$

which is nothing else but saying that the 1-dimensional and 2-dimensional marginals \mathbb{P}_{X_1} and \mathbb{P}_{X_1, X_2} are consistent. This is exactly what lies behind the lemma. If the finite dimensional distributions are defined through a "mother measures" \mathbb{P} then adding time-points with no restriction is doing nothing.

Proof. Suppose that J' has n elements. We check the equality on a generator of the n -fold product σ -algebra, which are all products (boxes) made up from n Borel-sets:

$$\begin{aligned} \mathbb{P}_{J'}(B_1 \times \dots \times B_n) &\stackrel{\text{Def.}}{=} \mathbb{P}(\{f : I \rightarrow E \mid f(t_i) \in B_i \forall t_i \in J'\}) \\ &\stackrel{\text{do nothing}}{=} \mathbb{P}(\{f : I \rightarrow E \mid f(t_i) \in B_i \forall t_i \in J', f(t_j) \in E \forall t_j \in J \setminus J'\}) \\ &\stackrel{\text{Def.}}{=} \mathbb{P}_J \circ \pi_{J,J'}^{-1}(B_1 \times \dots \times B_n). \end{aligned}$$

□

Just as for distribution functions let us give a name to the characteristic property that holds for all finite dimensional martingals of a measures \mathbb{P} on $\mathcal{E}^{\otimes I}$:



Definition 4.1.8. A set of finite dimensional distributions $\{\mathbb{P}_J : |J| < \infty\}$ on I is called **consistent** if $\mathbb{P}_J \circ \pi_{J,J'}^{-1} = \mathbb{P}_{J'}$ for all $J' \subseteq J$.

By the lemma consistency is a necessary property that must hold if there is a "mother" measure \mathbb{P} behind. The amazing truth is that consistency alone is also sufficient for the existence of a "mother" measures as long as measurable space (E, \mathcal{E}) is nice enough.



Theorem 4.1.9. (Kolmogorov extension theorem)

Let E a Polish metric space, $\mathcal{E} = \mathcal{B}(E)$, and I is an index set. If $\{\mathbb{P}^J : |J| < \infty\}$ is a family of consistent finite dimensional distributions on I , then there is a unique probability measures \mathbb{P} on $\mathcal{E}^{\otimes I}$ such that $\mathbb{P} \circ \pi_J^{-1} = \mathbb{P}_J$ for all finite $J \subseteq I$.

The importance of the theorem lies in the fact that finite dimensional distributions are easy (in many cases they are given by multivariate distribution functions or characteristic functions) whereas measures on infinite product spaces are not easy at all!

Since the proof is lengthy let us first discuss one of the most important examples, the infinite product measures that was already claimed to exist in the proof of Theorem ???. To use the Kolmogorov extension theorem one only needs to write down a family of finite dimensional distributions and check they are consistent. But how should we do that without knowing the measures \mathbb{P} ? This is very much in parallel to the construction of probability measures on $\mathcal{B}(\mathbb{R})$. As an example, in order to prove the existence of uniformly distributed random variables we

could already guess the distribution function $F_{\mathcal{U}([0,1])}$ and then check that it satisfies the defining properties needed. The approach is similar here:



If a priori we have an idea on how (at least) the finite dimensional distributions of a measure \mathbb{P} on $\mathcal{E}^{\otimes I}$ should look like then we can use that insight as an Ansatz to construct \mathbb{P} through the Kolmogorov extension theorem.

Here is an example of a situation in which the form of all finite dimensional distributions is a priori known.



Definition 4.1.10. Suppose (E, \mathcal{E}) is a measurable space and \mathbb{P} is a probability measure on \mathcal{E} . Then a measures μ on $\mathcal{E}^{\otimes \mathbb{N}}$ is called an **infinite product measure** if the product formula

$$\mu(B_1 \times \cdots \times B_n \times E \times E \cdots) = \mathbb{P}(B_1) \cdot \dots \cdot \mathbb{P}(B_n) \quad (4.1)$$

holds for all $n \in \mathbb{N}$ and $B_i \in \mathcal{E}$. An infinite product measure is typically written as $\mathbb{P}^{\otimes \infty}$ or $\mathbb{P}^{\otimes \mathbb{N}}$.

Since $\mathbb{P}(E) = 1$ the definition of the infinite product measures is already in terms of all finite dimensional distributions that are needed for the construction.



Theorem 4.1.11. If E is a Polish metric space and \mathbb{P} is a probability measure on $\mathcal{B}(E)$, then the infinite product measure $\mathbb{P}^{\otimes \infty}$ exists and is unique.

Proof. Uniqueness: We already know the argument from finite product measures (compare the proof of Theorem ??). Since the righthand side of (4.1) defines the measure on all finitely generated cylinder sets for $I = \mathbb{N}$ the measure is defined on a \cap -stable generator of $\mathcal{B}(E)^{\otimes \mathbb{N}}$. Hence, the uniqueness theorem for measures implies the uniqueness.

Existence: We already know from the definition that all finite dimensional marginals of \mathbb{P} must be finite product measures of \mathbb{P} . Hence, we should apply the Kolmogorov extension theorem to the family $\{\mathbb{P}_J : |J| < \infty\}$, where $\mathbb{P}_J = \mathbb{P}^{\otimes |J|}$ for all finite subsets J of \mathbb{N} . Of course those are consistent as all factors with $\mathbb{P}(E) = 1$ can be taken out from the preimages $\pi_{J,J'}^{-1}(B_1 \times \cdots \times B_{|J'|})$, for instance,

$$\begin{aligned} \mathbb{P}_J(B_1 \times E \times B_2 \times \cdots \times B_3) &= \mathbb{P}(B_1) \cdot \mathbb{P}(E) \cdot \mathbb{P}(B_2) \cdot \mathbb{P}(E) \cdots \mathbb{P}(E) \cdot \mathbb{P}(B_3) \\ &= \mathbb{P}(B_1) \cdot \mathbb{P}(B_2) \cdot \mathbb{P}(B_3) \\ &= \mathbb{P}_{J'}(B_1 \times B_2 \times B_3). \end{aligned}$$

Hence, there is a measure on $\mathcal{B}(E)^{\otimes \mathbb{N}}$ with the prescribed product measures as finite dimensional distributions. But this matches exactly the definition of the infinite product measures. \square

The proof of the existence of infinite product measures finally completes the proof of Theorem ?? . Now we really know that infinite sequences of independent random variables exist!

Lecture 20

Proof of the Kolmogorov extension theorem. The uniqueness statement follows directly from Proposition 4.1.1.

The basic idea of the proof is to apply the Carathéodory extension theorem. We will chose all finitely generated cylinder sets as the semiring \mathcal{S} that generates the path σ -algebra and define a σ -additive set-function on \mathcal{S} through the family of consistent finite dimensional distributions. The hard part will be to check the σ -additivity for which (as always) a compactness argument is needed. To carry out the compactness argument we essentially need to replace all appearing measurable sets by compacts sets while controlling the error that appears in the approximation. To carry this out two preliminary steps are needed.



Every probability measure μ on the Borel- σ -algebra of a Polish metric space (E, d) is inner regular, i.e. for all $B \in \mathcal{B}(E)$ there is a sequence of compact sets $K_n \subseteq B$ such that $\lim_{n \rightarrow \infty} \mu(B \setminus K_n) = 0$.

Recall from the proof of Proposition 3.1.12 the set $A := \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{N_n} B_{\frac{1}{n}}(x_k) \in \mathcal{B}(E)$. Since A is totally bounded it's closure (which is also totally bounded) is compact. Defining K to be the closure of A we have already proved that $\mu(E \setminus K) \leq \varepsilon$. For C closed we only need to define $K_C := K \cap C$. Then K_C is compact (closed and totally bounded) and $\mu(C \setminus K_C) = \mu((E \setminus K) \cap C) \leq \mu(E \setminus K) \leq \varepsilon$. Hence, all closed sets are inner regular, their measure can be approximated by measures of compact subsets. Now we can use the trick of good sets. Define

$$\mathcal{M} = \{B \in \mathcal{B}(E) : B \text{ is inner regular}\}.$$

It is not hard to see that \mathcal{M} is a Dynkin system. If B_1, \dots are disjoint, then chose $K_n \subseteq B_n$ compact such that $\mu(B_n \setminus K_n) \leq \frac{\varepsilon}{2^n}$. Then the⁵

⁵. Since the closed sets \mathcal{C} generate $\mathcal{B}(E)$ we can finish the argument as always:

$$\mathcal{B}(E) = \sigma(\mathcal{C}) \stackrel{1.2.11}{=} d(\mathcal{C}) \subseteq d(\mathcal{M}) = \mathcal{M} \subseteq \mathcal{B}(E).$$

Hence, all measurable sets are inner regular. The inner regularity of probability measures can now be used in a special situation where we replace "growing" sequences of Borel sets by growing sequences of compact sets:



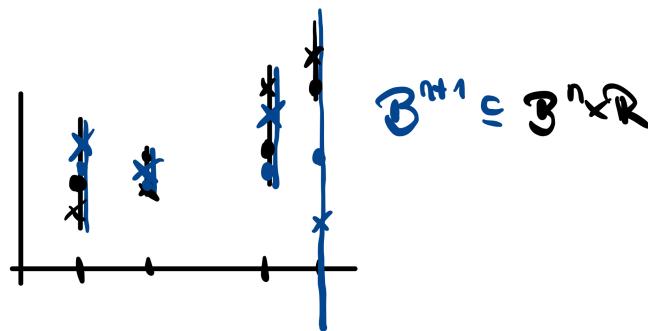
Let $B_n \subseteq E^n$ a sequence of Borel sets such that $B_{n+1} \subseteq B_n \times E$, and μ_n a consistent sequence of measures on $\mathcal{B}(E)^{\otimes n}$, i.e.

$$\mu_n(A_1 \times \cdots \times A_{n-1} \times \mathbb{R}) = \mu_{n-1}(A_1 \times \cdots \times A_{n-1}).$$

If $\mu_n(B_n) > \varepsilon$ for all $n \in \mathbb{N}$, then there is a sequence of compact sets $K_n \subseteq B_n$ so that

- $K_{n+1} \subseteq K_n \times \mathbb{R}$,
- $\mu_n(K_n) \geq \frac{\varepsilon}{2}$ for all $n \in \mathbb{N}$.

The claim essentially states that without loss of generality we will later be allowed to assume that all appearing measurable sets are compact sets. In principle the idea is simple. Approximate



More sequences through blue than black restrictions

all B_n with a smaller compact K_n and that's it. We just need to be careful since each K_n gives an approximation error of ε so the error of their combination will add up in n . That's why we

⁵An einem regnerischen tag aus-x0-en. heute regnets, dann der naechste kalte regnerische Tag.

approximate finer and finer so that the sum of the approximation errors is ε . It is not surprising how to proceed: Choose compact $K_n^* \subseteq \mathbb{R}^n$ such that $K_n^* \subseteq B_n$ and $\mu_n(B_n \setminus K_n^*) \leq \frac{\varepsilon}{2^{n+1}}$ and then define

$$K_n = (K_1^* \times \mathbb{R}^{n-1}) \cap \dots \cap (K_{n-1}^* \times \mathbb{R}) \cap K_n^*.$$

We only need to check that K_n does the job. It follows directly that $K_n \subseteq B_n$ and $K_{n+1} \subseteq K_n \times \mathbb{R}$ and also that

$$\begin{aligned} \mu_n(K_n) &\stackrel{\sigma\text{-add.}}{=} \mu_n(B_n) - \mu_n(B_n \setminus K_n) \\ &= \mu_n(B_n) - \mu_n(B_n \setminus ((K_1^* \times \mathbb{R}^{n-1}) \cap \dots \cap (K_{n-1}^* \times \mathbb{R}) \cap K_n^*)) \\ &\stackrel{\sigma\text{-add.}}{\geq} \mu_n(B_n) - \mu_n(B_n \setminus (K_1^* \times \mathbb{R}^{n-1})) - \dots - \mu_n(B_n \setminus K_n^*) \\ &\stackrel{B_k \subseteq B_1 \times \mathbb{R}^{k-1}}{\geq} \mu_n(B_n) - \mu_n(B_1 \times \mathbb{R}^{n-1} \setminus K_1^* \times \mathbb{R}^{n-1}) - \dots - \mu_n(B_n \setminus K_n^*) \\ &\stackrel{\text{consistent}}{\geq} \mu_n(B_n) - \mu_1(B_1 \setminus K_1^*) - \dots - \mu_n(B_n \setminus K_n^*) \\ &\geq \varepsilon - \frac{\varepsilon}{4} - \dots - \frac{\varepsilon}{2^{n+1}} \geq \frac{\varepsilon}{2}. \end{aligned}$$

That's it. Before we proceed let us quickly recall Carathéodory's extension Theorem 1.3.7:



If μ is a σ -additive set-function on a semiring \mathcal{S} , then there is a unique measures $\bar{\mu}$ on $\sigma(\mathcal{S})$ with $\mu(A) = \bar{\mu}(A)$ for all $A \in \mathcal{S}$.

We proceed in two main steps. First, a semiring \mathcal{S} and a set-function μ are defined which is shown to be well-defined. Secondly, the properties of a σ -additive set function are checked. We define \mathcal{S} as the set of all finitely generated cylinder sets, so all sets C that can be written as $\pi_J^{-1}(B_1 \times \dots \times B_n)$ for some finite subset $J = \{t_1, \dots, t_n\}$ of I and Borel sets B_i . The set of finitely generated cylinder sets is clearly a semiring. The set function μ on \mathcal{S} is defined as

$$\mu(\pi_J^{-1}(B_1 \times \dots \times B_n)) := \mathbb{P}_J(B_1 \times \dots \times B_n) \quad (4.2)$$

for all cylinder sets from \mathcal{S} . We now come to the main step of the argument:



μ is a well-defined σ -additive set-function on \mathcal{S} .

(i) **μ is well-defined:** Suppose a cylinder set C from \mathcal{S} has two representations

$$\pi_J^{-1}(B_1 \times \dots \times B_n) = C = \pi_{\tilde{J}}^{-1}(\tilde{B}_1 \times \dots \times \tilde{B}_m) \quad (4.3)$$

which can happen if some of the B_i or \tilde{B}_i are equal to E . It is precisely the assumed consistency property (we can cross out all E and delete the time-points) that ensures that the measures of both representations is equal.

(ii) **μ is a σ -additive set-function:** There are two properties to check. We start with the empty set. Since μ is well-defined we are allowed to write \emptyset in some arbitrary way as a finite cylinder set: $\emptyset = \pi_t^{-1}(\emptyset)$ for arbitrary $t \in I$ yields $\mu(\emptyset) = \mathbb{P}_{\{t\}}(\emptyset) = 0$.

The difficulty is to prove the σ -additivity of μ on \mathcal{S} . Assume C_n is a sequence of disjoint cylinder sets from \mathcal{S} such that $C := \bigcup_{k=1}^{\infty} C_k \in \mathcal{S}$. We have to prove that $\mu(C) = \sum_{k=1}^{\infty} \mu(C_k)$. If we write

$$\mu(C) = \mu(C \setminus \bigcup_{k=1}^N C_k) + \mu(\bigcup_{k=1}^N C_k)$$

and apply monotonicity of measures it is clear that it suffices to prove

$$\lim_{N \rightarrow \infty} \mu(D_N) = 0, \quad (4.4)$$

where $D_N = C \setminus \cup_{k=1}^N C_k \in \mathcal{S}$. To prove (4.4) we use a compactness argument based on the steps above. The sequence $\mu(D_N)$ is decreasing (monotonicity of measures) to some number ε . Let us assume $\varepsilon > 0$ and construct a contradiction. The contradiction will be that

$$\cap_{N=1}^{\infty} D_N \neq \emptyset \quad (4.5)$$

which is absurd because as it would contradict $C = \cup_{k=1}^{\infty} C_k$. Since all D_N are (finite) cylinder sets, also $D := \cup_{N=1}^{\infty} D_N$ is a countable cylinder set for some times $t_1 < t_2 < \dots$. Since the D_N decrease, without loss of generality (otherwise add artificially extra points with corresponding Borel-set E that have no effect) we can assume that D_N is finitely generated by more and more time-points:

$$D_N = \{f : (f(t_1), \dots, f(t_N)) \in B_N\}$$

for a sequence of Borel sets $B_N \subseteq \mathcal{E}^N$ satisfying $B_{N+1} \subseteq B_N \times E$. Now we apply the regularity from the step above and replace without loss of generality the sets B_N by compact sets $K_N \subseteq B_N$ (changing ε into $\varepsilon/2$). Since the K_N are non-empty (they have positive measures), we can pick vectors $x^N = (x_1^N, \dots, x_N^N) \in K_N$. The decreasing property of the K_N implies that the sequences (x_1^N) lies in the compact set K_1 . Hence, there is a convergent subsequence $(x_1^{N_k})$ that converges to some $x_1 \in K_1$. Next, the sequence $(x_1^{N_k}, x_2^{N_k})$ is contained in K_2 , thus, there is another converging subsequence $(x_1^{N'_k}, x_2^{N'_k})$ that converges to some $(x_1, x_2) \in K_2$. Continuing like that we obtain a sequence (x_n) in E such that $(x_1, x_2, \dots, x_n) \in K_n \subseteq B_n$ for all $n \in \mathbb{N}$. But then the set of functions passing through x_k at times t_k are in all D_N . Hence, (4.5) is matched and we have our contradiction. \square

(C) Path valued random variables and stochastic processes

The third step in the triology of measure theory is to understand the connection of stochastic processes and path-valued random variables. We start with the natural generalization of Proposition ??:



Proposition 4.1.12. $X = (X_t)_{t \in I}$ is an E -valued stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$ if and only if the mapping

$$X : \omega \mapsto (t \mapsto X_t(\omega))$$

is a path-valued random variable.

A bit of care is needed to explain the statement. If (X_t) is a family of random variables it is clear how the path-valued mapping has to be interpreted. Fix ω and then consider all values $X_t(\omega)$ as a function in t . The converse is a bit more tricky. To switch back from the path-valued valued mapping to a family of E -valued mappings one needs to evaluate X at all times T , formalized using the projections: $X_t := \pi_t \circ X$.

Proof. " \Leftarrow ": For $B \in \mathcal{E}$ we have

$$X_t^{-1}(B) = X^{-1}(\pi_t^{-1}(B)) \in \mathcal{A},$$

because $\pi_t^{-1}(B)$ is a finitely generated cylinder set and X is $(\mathcal{A}, \mathcal{E}^{\otimes I})$ -measurable.

" \Rightarrow ": Using Proposition 2.1.4 the measurability property only needs to be checked on an arbitrary generator. We use the cylinder sets $\pi_t^{-1}(B)$ generated by single time-points:

$$X^{-1}(\pi_t^{-1}(B)) = \{\omega \in \Omega : X_t(\omega) \in B\} = X_t^{-1}(B) \in \mathcal{A}.$$

Hence, seen as a path-valued random variable X is measurable. \square

Next, the notion of the law from random variables and random vectors is generalized to stochastic processes:



Definition 4.1.13. Suppose X is an E -valued stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$.

- (i) The image measure

$$\mathbb{P}_X(A) := \mathbb{P}(X \in A), \quad A \in \mathcal{E}^{\otimes I},$$

is called the **law of X** .

- (ii) For a finite subset $J = \{t_1, \dots, t_n\}$ of I the law evaluated on the cylinder sets spanned by J

$$\begin{aligned}\mathbb{P}_X^J(B_1 \times \dots \times B_n) &:= \mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) \\ &= \mathbb{P}_X(\pi_{t_1, \dots, t_n}^{-1}(B_1 \times \dots \times B_n)), \quad t_i \in I, B_i \in \mathcal{E},\end{aligned}$$

is called a **finite dimensional distribution of X** .

Warning! The law of a stochastic process is a delicate object, a measure on the path space. Usually the law does not play a big role and we are mostly interested in the finite dimensional distributions. Both concepts are essentially equivalent but finite dimensional distributions are laws on much nicer σ -algebras.



Proposition 4.1.14. The law of a stochastic process is uniquely determined by the family of all finite dimensional distributions.

Proof. This follows from the abstract theory of the previous section. Since all finitely generated cylinder sets generate the path σ -algebra $\mathcal{E}^{\otimes I}$ and are intersection stable the measure \mathbb{P}_X is uniquely determined on all sets $\pi_{t_1, \dots, t_n}^{-1}(B_1 \times \dots \times B_n)$. \square

It is crucial to assume that **all** finite dimensional distributions are equal. If X is a symmetric random variable, then $(X, -X)$ and (X, X) have different laws but the same one-dimensional distributions.



Definition 4.1.15. Two stochastic processes X and Y with index set I have the same law if $\mathbb{P}_X = \mathbb{P}_Y$. Equivalently, we say they have the same finite-dimensional distributions if $\mathbb{P}_X^J = \mathbb{P}_Y^J$ for all finite $J \subseteq I$.

Just as for random variables, the case $|I| = 1$, two stochastic processes with same law do not need to be defined on the same probability space. The notion of equality in law is the weakest equality notation available, stronger notions will be discussed in the next section.

If you remember well the previous 300 pages of these lecture notes you will know what is up next. The existence of stochastic processes using the canonical constructions. Earlier versions covered the cases $|I| = 1$, $|I| < \infty$, and $I = \mathbb{N}$. Here is the general version for arbitrary I only assuming E is Polish.



Theorem 4.1.16. (Canonical construction on E^I)

Suppose E is a Polish metric space, I an index set, and $\{\mathbb{P}_J : |J| < \infty\}$ a consistent family of finite-dimensional distributions. Then there is a stochastic process X on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with marginals $\mathbb{P}_X^J = \mathbb{P}_J$ for all finite sets of times.

Proof. In the usual way we construct a probability space and a (path-valued) random variable:

- $\Omega = E^I$
- $\mathcal{A} = \mathcal{E}^{\otimes I}$

- $\mathbb{P} = \mathbb{P}_X$, the measure on $\mathcal{E}^{\otimes I}$ obtained from the consistent family using Theorem 4.1.9.
- X is the identity mapping from E^I to E^I .
- $X_t = \pi_t(X)$ the evaluation of X at time t .

The measurability of X is clear (preimages of measurable sets are the same measurable sets) and the law is $\mathbb{P}_X = \mathbb{P}$ which has the good finite dimensional distributions by construction. \square

People tend to like the canonical construction as it always works (justifying the name canonical) but there are very good reasons to avoid the canonical construction if possible. For instance the statement " X is \mathbb{P} -almost surely continuous/differentiable/constant/etc." is problematic for a real-valued process on the canonical probability space because such events are not measurable in $\mathcal{B}(\mathbb{R}^{[0,\infty)})$, hence, no probabilities can be defined. In fact, the problem can be avoided using a better definition of nullsets than the one we gave in Definition ??, compare also the footnote after the definition.



From now on a set $A \subseteq \Omega$ is called a **nullset** if there is a measurable set B with $A \subseteq B$ and $\mathbb{P}(B) = 0$. A property E holds \mathbb{P} -almost surely, if $\{\omega : E(\omega) \text{ does not hold}\}$ is contained in a measurable set with measure 0.

There is no problem for the earlier discussion as all previous nullsets of interest $\{X = Y\}$, $\{X \geq 0\}$, $\{X_n \text{ converges}\}$, etc. are measurable themselves. We will return a few times to the measurability problems on the path space but now want to get towards examples. Here is a first important class of stochastic processes that can be constructed using the theory developed above.



Definition 4.1.17. A stochastic process $(X_t)_{t \in I}$ is called a **Gaussian process** if all finite dimensional distributions are multi-variate Gaussian. We call

$$m(t) := \mathbb{E}[X_t], \quad t \in I,$$

the mean-function of X and

$$K(t, s) := \mathbb{E}[(X_t - m(t))(X_s - m(s))], \quad t, s \in I,$$

the covariance function of X . A Gaussian process is called **centered** if $m \equiv 0$.

Choosing $I = \{1, \dots, d\}$ shows that Gaussian processes are direct generalizations of Gaussian vectors and in this case m is the mean-vector and K the covariance matrix. Covariance functions are symmetric and positive semi-definite, i.e.

$$\begin{aligned} \sum_{i,j=1}^n a_i a_j K(t_i, t_j) &= \sum_{i,j=1}^n a_i a_j \mathbb{E}[(X_{t_i} - m(t_i))(X_{t_j} - m(t_j))] \\ &= \mathbb{E}\left[\left(\sum_{i=1}^n a_i (X_{t_i} - m(t_i))\right)^2\right] \geq 0, \end{aligned}$$

for all $n \in \mathbb{N}$, $t_i, t_j \in I$ and $a_i \in \mathbb{R}$. Note that in the case $I = \{1, \dots, d\}$ this is exactly the definition of a positive semi-definite matrix. Here are some examples for symmetric positive semi-definite functions on $I = [0, \infty)$:

$$K(t, s) = t \wedge s, \quad K(t, s) = \exp(|t - s|), \quad K(s, t) = (1 + |t - s|^2)^{-\alpha}, \quad \alpha \geq 0.$$

Gaussian processes are particularly useful as they are easy to construct through Kolmogorov's extension theorem:



Let $m : I \rightarrow \mathbb{R}$ and $K : I \times I \rightarrow \mathbb{R}$ be symmetric and positive semi-definite. Then there is a Gaussian process $(X_t)_{t \in I}$ with mean-vector m and covariance function K .

Hint: In order to use Theorem 4.1.16 one has to write down all finite dimensional distributions. Just as in the construction of infinite product measures we have the advantage that those are already known, as they are supposed to be multivariate Gaussian. Go back to the end of Section 3.6 to see why only m and K are needed to guess the form of all finite dimensional distributions and then show that they are consistent.

A further nice application of the construction theorem is to construct Markov chains on a infinite time-horizon:



Use Theorem 4.1.16 to construct a Markov chain $(X_t)_{t \in \mathbb{N}}$ on a finite state-space $\{1, \dots, n\}$ with transition matrix P .

4.1.2 Stochastic processes with continuous sample paths

Lecture 21

So far we developed a tool to construct stochastic processes with given finite-dimensional distributions using the canonical construction based on the Kolmogorov extension theorem. The story is not over, yet, if we are interested in stochastic processes with regular (e.g. continuous) sample paths. One of the motivations to keep the weak convergence chapter more general than just \mathbb{R}^d was to introduce a notion of convergence of stochastic processes. In order to do so we must be able to interpret a stochastic process as a random variable with values in a Polish space. So far, it is unclear what that Polish space could be used because the generic path space \mathbb{R}^I has no topological structure what so ever. The aim of this section is two-fold. First, we first explain the path-space of continuous real-valued functions and stochastic processes with continuous paths, secondly, we discuss ways of constructing such processes. Unfortunately, it's not possible to prove a generic construction theorem like the Kolmogorov extension theorem, the proof of the Kolmogorov extension theorem breaks down if used on continuous functions. Instead, we will introduce a method to modify stochastic processes without path regularity into processes with (Hölder) continuous processes.

The path space of continuous functions

The space of continuous functions on a compact set (here: compact interval) has already been studied in the previous chapter. It is well-known from basic analysis that $C([0, T])$ is complete, the Stone-Weierstraß theorem applied to polynomials with rational coefficients also shows that $C([0, T])$ is separable. Hence, $(C([0, T]), \|\cdot\|_\infty)$ is a Polish space and as such suitable for the theory developed in the previous chapter. As we are mostly interested in stochastic processes indexed by $[0, \infty)$ it is important to see that also continuous functions indexed by $[0, \infty)$ can be made into a Polish space:



Proposition 4.1.18. (i) If we define

$$d(f, g) := \sum_{k=1}^{\infty} \frac{d_k(f, g)}{2^k}, \quad d_k(f, g) = \sup_{x \in [0, k]} |f(x) - g(x)| \wedge 1,$$

then d is a metric on $C([0, \infty))$.

(ii) d metrizes convergence on compacts, that is

$$\lim_{n \rightarrow \infty} d(f_n, g) = 0 \Leftrightarrow \lim_{n \rightarrow \infty} \sup_{x \in K} |f_n(x) - g(x)| = 0 \text{ for all } K \text{ compact.}$$

(iii) $(C([0, \infty)), d)$ is a Polish metric space.

The truncation by 1 is important for two reasons. First, it ensures the metric is always finite, and secondly, it allows to use dominated convergence in the proof below.

Proof. (i) We know from basic analysis that d_k is a metric on $C([0, k])$, the properties of a metric space then follow immediately for d .

(ii) It follows from the Heine-Borel Theorem that all compact subsets of $[0, \infty)$ are bounded.

" \Rightarrow ": Suppose $\lim_{n \rightarrow \infty} d(f_n, g) = 0$. Choose K compact and some k such that $K \subseteq [0, k]$. Since $\frac{d_k(f_n, g)}{2^k} \leq d(f_n, g)$ we find $\lim_{n \rightarrow \infty} \sup_{x \in K} |f_n(x) - g(x)| \leq \lim_{n \rightarrow \infty} d_k(f_n, g) = 0$.

" \Leftarrow ": The truncation by 1 allows us to apply the dominated convergence theorem to get

$$\lim_{n \rightarrow \infty} d(f_n, g) \stackrel{\text{DCT}}{=} \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} d_k(f_n, g) \frac{1}{2^k} = 0,$$

since all $[0, k]$ are compact.

(iii) Separability: The set of polynomials restricted to $[0, n]$ with rational coefficients is countable and dense in $(C([0, n]), \|\cdot\|_\infty)$, see the discussion below Theorem 3.4.1. A diagonal argument then shows that all polynomials with rational coefficients are countable and dense in $(C([0, \infty)), d)$. Hence, the separability is proved.

Completeness: Suppose $(f_n)_{n \in \mathbb{N}}$ is Cauchy. By the definition of the metric, $(f_n)_{n \in \mathbb{N}}$ restricted to $[0, k]$ is also Cauchy in $(C([0, k]), \|\cdot\|_\infty)$. Since $(C([0, k]), \|\cdot\|_\infty)$ is complete, for all $k \in \mathbb{N}$ there is a uniform (continuous) limit g_k on $[0, k]$. Let us define a function $g : [0, \infty) \rightarrow \mathbb{R}$ as $g_k(x)$ for $x \in [0, k]$ and some $k > x$. Since $g_k = g_{k'}$ on $[0, k]$ for all $k' \geq k$ this construction is well-defined. Using dominated convergence once more yields

$$\lim_{n \rightarrow \infty} d(f_n, g) \stackrel{\text{DCT}}{=} \sum_{k=1}^{\infty} \lim_{n \rightarrow \infty} d_k(f_n, g) \frac{1}{2^k} = 0.$$

Hence, $(C([0, \infty)), \|\cdot\|_\infty)$ is complete. \square

Since $(C([0, \infty)), d)$ is a Polish metric space there is a natural σ -algebra on E , the Borel- σ -algebra generated by all open sets. Such a σ -algebra has much more structure than the product σ -algebra on the generic path space $\mathbb{R}^{[0, \infty)}$. As an example, all probability measures are inner regular. In contrast, $\mathbb{R}^{[0, \infty)}$ is topologically nothing useful. That's why the σ -algebra is not the topological σ -algebra induced by open sets (what are open sets without a topology?) but defined "by hands". Even though $\mathcal{B}(C([0, \infty)))$ is structurally much more rich there is a lot of similarity to the generic path σ -algebra. For example both σ -algebras have the same simple generators, the cylinder sets, but now cylinder sets of continuous functions.



Proposition 4.1.19. $\mathcal{B}(C([0, \infty)))$ is generated by the cylinder sets of continuous functions, i.e.

$$\begin{aligned} \mathcal{B}(C([0, \infty))) &= \sigma(\{\pi_t^{-1}(B) : t \geq 0, B \in \mathcal{B}(\mathbb{R})\}), \\ &= \sigma(\{\text{finitely generated cylinder sets of continuous functions}\}), \end{aligned}$$

where $\pi_t : C([0, \infty)) \rightarrow \mathbb{R}, f \mapsto f_t$, are the projections on time t .

In order to avoid confusion with the product σ -algebra on all generic functions let us emphasize the difference of projections on all continuous functions and generic functions. The cylinder sets of all functions are much larger since they contain all continuous functions passing through B at time t but also such discontinuous functions.

Proof. Intersecting cylinder sets generated by a single time-point leads to all finitely generated cylinder sets. Hence, the second equality holds and we only need to check the first. We will use the typical trick that in order to show $\sigma(\mathcal{E}) = \sigma(\mathcal{E}')$ it is enough to show $\mathcal{E} \subseteq \sigma(\mathcal{E}')$ and



cylinder sets of continuous and all functions

$\mathcal{E}' \subseteq \sigma(\mathcal{E})$. To do so note that the projections $\pi_t : C([0, \infty)) \rightarrow \mathbb{R}$ are continuous mappings. This follows immediately from the metric of $C([0, \infty))$. If $f_n \rightarrow f$, then the metric implies pointwise convergence $f_n(t) \rightarrow f(t)$ for all $t \in [0, \infty)$ which is nothing but $\pi_t(f_n) \rightarrow \pi_t(f)$.

" \supseteq ": Since the projections π_t are continuous and we are working with Borel- σ -algebras they are also measurable. Hence, all preimages of π_t (cylinder sets generated by one time-point) are in $\mathcal{B}(C([0, \infty)))$.

" \subseteq ": It is enough to show that some generator of $\mathcal{B}(C([0, \infty)))$ is contained in $\sigma(\{\pi_t : t \geq 0\})$. We chose the open sets. Since $(C([0, \infty)), d)$ is separable every open set can be written as a countable union of open balls (the open balls around polynomials with rational coefficients form a base of the topology). Hence, we need to show that open balls of functions are in $\sigma(\{\pi_t : t \geq 0\})$. Here is a trick. It is enough to prove that all mappings $f \mapsto d(f_0, f)$ are measurable as their preimages of $[0, \varepsilon] \in \mathcal{B}([0, \infty))$ are $B_\varepsilon(f_0)$. By the definition of d it is enough to show that all $f \mapsto d_k(f_0, f)$ are measurable for all $k \in \mathbb{N}$ as limits, sums, scalar multiplications, and taking minima are operations that preserve measurability by Proposition 2.3.6. But this follows by rewriting

$$d_k(f_0, f) = \sup_{t \leq k} |f_0(t) - f(t)| \wedge 1 = \sup_{t \leq k, t \in \mathbb{Q}} |f_0(t) - f(t)| \wedge 1 = \sup_{t \leq k, t \in \mathbb{Q}} |\pi_t(f_0) - \pi_t(f)| \wedge 1,$$

since the righthand side is a concatenation of measurable functions with measurable operations. \square

The previous proposition shows that $\mathcal{B}(C([0, \infty)))$ has the same type of generator as the product σ -algebra $\mathcal{E}^{\otimes I}$ on E^I . It is thus reasonable to copy ideas carried out before. If $J \subseteq [0, \infty)$ is a finite set with n time-points, then

$$\mathbb{P}_J(B_1 \times \dots \times B_n) := \mathbb{P}(\{f \in C([0, \infty)) : f(t_1) \in B_1, \dots, f(t_n) \in B_n\}), \quad B_i \in \mathcal{B}(\mathbb{R}),$$

is called a finite dimensional distribution of \mathbb{P} . Since probability measures are uniquely defined on a \cap -stable generator the finite dimensional distributions uniquely determine the law, just as Proposition 4.1.1 for probability measures on the generic path space.



Proposition 4.1.20. A probability measure \mathbb{P} on $\mathcal{B}(C([0, \infty)))$ is uniquely determined by the family $\{\mathbb{P}_J : |J| < \infty\}$ of all finite dimensional marginal distributions.

The generator of finitely generated cylinder sets reveals a striking fact. Restricting a cylinder set from the generator of $\mathcal{B}(\mathbb{R}^{[0, \infty)})$ to the continuous functions gives a cylinder set from the generator of $\mathcal{B}(C([0, \infty)))$. This is no coincidence, the same holds for the entire σ -algebras:



Proposition 4.1.21. $\mathcal{B}(\mathbb{R}^{[0, \infty)})|_{C([0, \infty))} = \mathcal{B}(C([0, \infty)))$

Recall that for a measurable space (Ω, \mathcal{A}) and some (possibly non-measurable) set $B \subseteq \Omega$ one defines the trace- σ -algebra on B as $\mathcal{A}|_B := \{A \cap B : A \in \mathcal{A}\}$. It is easy to check that this is again a σ -algebra and that $\mathcal{C} \cap B$ is a generator of $\mathcal{A}|_B$ if \mathcal{C} is a generator of \mathcal{A} .

Proof. We use again the strategy from Example 1.2.6 that in order to prove equality $\sigma(\mathcal{E}) = \sigma(\mathcal{E}')$ it suffices to show $\mathcal{E} \subseteq \sigma(\mathcal{E}')$ and $\mathcal{E}' \subseteq \sigma(\mathcal{E})$.

" \subseteq ": Note that the trace- σ -algebra is generated by all intersections of finitely generated cylinder sets of all functions with the continuous functions. Of course, all such sets are finitely generated cylinder sets of continuous functions (generated by the same time-points), which are in $\mathcal{B}(C([0, \infty)))$.

" \supseteq ": We use the cylinder sets of continuous functions generated by one time-point. Those are of course the same as the cylinder sets of all functions (through the same time-point) intersected with the continuous functions. But those sets are in the trace- σ -algebra. \square

Here is an important question that is motivated from the observation that $\mathcal{B}(C([0, \infty)))$ and $\mathcal{B}(\mathbb{R}^{[0, \infty)})$ have the same kind of semiring generators.



If the continuous paths σ -algebra is so similar to the generic product σ -algebra, can we also use the Carathéodory extension theorem to construct a measure from a consistent family of finite dimensional distributions? The frustrating answer is: no!

Going through the proof of the Kolmogorov extension theorem one finds that the proof fails in the very end when we constructed the contradiction. We stop the discussion here, there is no complete identification of all probability measures on $C([0, \infty))$. Nonetheless, we will see below how many stochastic processes with continuous sample paths can be constructed without the canonical representation.



Definition 4.1.22. A stochastic process $(X_t)_{t \geq 0}$ on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is said to have continuous/differentiable/etc. paths if there is an event C of probability 1 so that $t \mapsto X_t(\omega)$ is continuous/differentiable/etc. for all $\omega \in C$.

It is important to note that the definition does not use that $\{\omega : t \mapsto X_t(\omega)\}$ itself is measurable, only that it contains a measurable set C which has probability 1.

The law $\mathbb{P}_X(B) := \mathbb{P}(X \in B)$ of a stochastic process has already been defined on the product σ -algebra $\mathcal{B}(\mathbb{R}^{[0, \infty)})$, as a push-forward of the process seen as path-valued random variable (see Definition 4.1.13). If X has continuous sample paths it would certainly be nicer also to define the law of X on the continuous functions. This can be done, with one little refinement. Since $t \mapsto X_t(\omega)$ is only known to be continuous for all $\omega \in C$ for some event C of measure 1 we first restrict the entire probability space to C . Then paths are continuous for all $\omega \in \Omega' := \Omega \cap C$ and the law can be defined:



Definition 4.1.23. If $(X_t)_{t \geq 0}$ is a stochastic process with continuous sample paths then the law \mathbb{P}_X of X on $C([0, \infty))$ is defined as

$$\mathbb{P}_X(B) = \mathbb{P}_{|C}(X \in B), \quad B \in \mathcal{B}(C([0, \infty))).$$

Since C has full measure this little twist does not effect the finite dimensional distributions. The concept of the law of continuous processes will become much more important later when we prove the universal approximation property of a Brownian motion by random walks (Donsker's theorem). For now we only remark that any given probability measure P on $\mathcal{B}(C([0, \infty))$ allows a canonical construction:

- $\Omega := C([0, \infty))$,
- $\mathcal{A} := \mathcal{B}(C([0, \infty)))$,
- $\mathbb{P} := P$,
- X is the identity mapping from $C([0, \infty))$ to $C([0, \infty))$ and $X_t = \pi_t \circ X$.



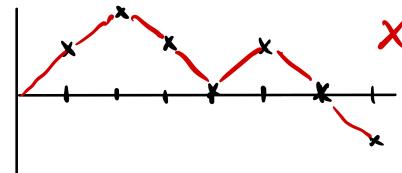
Check that $(X_t)_{t \geq 0}$ is a continuous stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$ with law $\mathbb{P}_X = P$.

The big trouble about continuous stochastic processes is that there is no generic construction theorem for probability measures on the continuous paths, hence, no generic construction theorem of continuous stochastic processes through the canonical construction. There are two major ways to overcome this problem:

- Some continuous stochastic processes can be written down explicitly.
- Some discontinuous stochastic processes can be modified into continuous processes.

For the first approach here is a simple example, a similar example appears later as one of two possible constructions of the Brownian motion.

Example 4.1.24. Take an iid sequence Y_1, Y_2, \dots on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ (this is based on the Kolmogorov extension theorem!) and define the random walk with jump-sizes Y as in Example ???. Now define X to be the linear interpolation as in the illustration. Then



cylinder sets of continuous and all functions

$t \mapsto X_t(\omega)$ is continuous even for all $\omega \in \Omega$. The corresponding law \mathbb{P}_X on $C([0, \infty))$ is called the **random walk law**. It will later be proved that random walk laws of scaled random walks converge weakly on $\mathcal{B}(C([0, \infty)))$ to the Wiener measures, the law of the Brownian motion.

We now turn to the powerful second approach, modifying discontinuous processes into continuous processes.

The constructing of continuous stochastic processes

There are different ways of defining the regularity of a real-valued functions. Continuity and differentiability are simple but not very precise notions, a more refined notion is Hölder continuity.



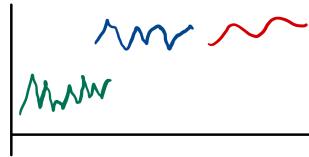
Definition 4.1.25. A function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is called Hölder-continuous of index $\gamma \in (0, 1]$

$$|f(x) - f(y)| \leq K|x - y|^\gamma, \quad \forall x, y \in \mathbb{R}^d.$$

A function f is called locally Hölder continuous of index $\gamma > 0$ if for all x there is a neighborhood U of x and a constant K_x such that

$$|f(x) - f(y)| \leq K_x|x - y|^\gamma, \quad \forall y \in U.$$

The intuition should be that the roughness decreases for increasing γ and $\gamma = 1$ gives pretty smooth functions. Why is this? Suppose x is close to y . Then the difference of the function between x and y , also called an increment, is bounded by the distance raised to the power γ . If x is very close to y then increasing the power γ decreases the size of the increment $|f(x) - f(y)|$, f has infinitesimally small fluctuations (the spikes are not very sharp). To sharpen your intuition please think about the following exercise.



from rough to smooth: small γ , medium γ , $\gamma = 1$



- If f is (locally) Hölder continuous of index γ , then f is also (locally) Hölder continuous of index γ' for all $\gamma' < \gamma$.
- If f is continuously differentiable, then f is locally Hölder continuous of index 1 (which is the same as Lipschitz continuous).
- If f would be Hölder continuous with index $\gamma > 1$, then f must be constant. That's why we directly restrict to $\gamma \in (0, 1]$.
- Polynomials $f(x) = |x|^\beta$, $\beta < 1$, are not differentiable at 0 but Hölder continuous of index β . Away from 0 they are continuously differentiable, hence, locally Lipschitz continuous.

Motivated by the last example we might think of Hölder continuous functions of index β to be spiky functions with (possibly many) rough tips that can be as rough as $|x|^\beta$ around 0.

We will prove below that under certain conditions a stochastic process can be "modified" in a way that sample paths are continuous. Even better, there is a relatively simple condition under which sample paths are Hölder continuous with an identifiable index γ . Caused by the additional complexity of time there are different notions of equality of stochastic processes that go beyond the uniqueness of the laws when seen as a path-valued random variable.



Definition 4.1.26. Suppose X and Y are stochastic processes with index set I on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then X is a **modification** of Y if $\mathbb{P}(X_t = Y_t) = 1$ for all $t \in I$.

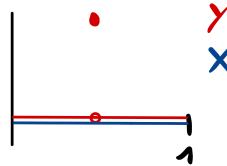
Being modifications of each other is a very different concept of equality of processes than having the same finite dimensional distributions (i.e. same law). Here, X and Y must be defined on the same probability space and, intersecting events of probability 1, X and Y automatically have the same finite dimensional distributions. In a way, being modifications of each other is much more equal than only having same law. Still, the paths of X and Y could be very different. As an example if all X_t are symmetric random variables, then mirroring the path gives processes that look very different but X and $-X$ are modifications of each other. The natural stronger notion would move "for all" in the probability. We want to call two processes indistinguishable if $\mathbb{P}(X_t = Y_t \text{ for all } t \in I) = 1$. But there is a technical problem: Writing $\{X_t = Y_t \text{ for all } t \in I\} = \cap_{t \in I} \{X_t = Y_t\}$ there is again an uncountability problem. The righthand side is measurable if I is countable, but not necessarily if I is uncountable. Here is the good definition:



Definition 4.1.27. Suppose X and Y are stochastic processes with index set I on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then X and Y are called **indistinguishable** if there is an event C with probability 1 such that $X_t(\omega) = Y_t(\omega)$, $t \in I$, for all $\omega \in C$.

Here is the simplest example to separate both notions of equality for processes. Suppose $\mathcal{U} \sim$

$\mathcal{U}([0, 1])$ is a uniformly distributed random variable on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and define the stochastic processes $X_t := 0$ and $Y_t := \mathbf{1}_{\{U\}}(t)$ on $(\Omega, \mathcal{A}, \mathbb{P})$ with index set $I = [0, 1]$. Then X is a modification of Y as $\mathbb{P}(X_t = Y_t) = \mathbb{P}(U \neq t) = 1$ but their paths are almost surely different.



modifications but not indistinguishable

Please check the connections of the equalities for processes and think about simple counter examples:



Let

- (i) X and Y are indistinguishable.
- (ii) X is a modification of Y .
- (iii) X and Y have the same finite dimensional distributions (and same law).

Show that (i) \Rightarrow (ii) \Rightarrow (iii) but both converse implications are generally wrong.

In order to construct continuous processes from discontinuous processes we will refer to the famous Kolmogorov-Chentsov theorem. The theorem actually much better, under a Hölder-type condition (which is easy to check in many situations) a modification with Hölder-continuous paths is constructed.



Theorem 4.1.28. (Kolmogorov-Chentsov theorem)

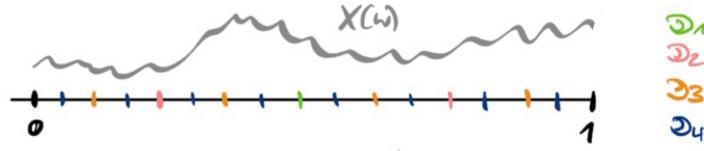
Suppose $(X_t)_{t \geq 0}$ is a stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$ such that there are $\alpha, \beta, T > 0$ with

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq c \cdot |t - s|^{1+\beta}, \quad \forall t, s \leq T.$$

Then there is a modification \tilde{X} of X such that almost surely \tilde{X} has locally Hölder continuous paths of order γ for all $\gamma < \frac{\beta}{\alpha}$.

Proof. We only prove the theorem for processes indexed by $[0, 1]$. Since $[0, \infty)$ can be split into countably many intervals a sequence of modifications on $[n, n+1]$ can be combined to a modification on $[0, \infty)$ and the almost sure Hölder continuity holds on the intersection of the countably many almost sure events.

Fix $\gamma < \frac{\beta}{\alpha}$ and set $K := \frac{2}{1-2^{-\gamma}}$. We will construct a set $C \in \mathcal{A}$ of measure 1 such that $\omega \mapsto X_t(\omega)$ is (locally) Hölder continuous with constant K and index γ on a set of time-points that is dense D in $[0, \infty)$, that is, the local Hölder estimate holds pathwise for the mapping $t \mapsto X_t(\omega)$ for $t \in D$ and $\omega \in C$. Then we redefine $X_t(\omega)$ for $t \notin D$ and this will be the modification. The set D of good time-points will be the set of all dyadic in $[0, 1]$, defined as follows: $\mathcal{D}_n = \{0, \frac{1}{2^n}, \frac{2}{2^n}, \dots, 1\}$ and $\mathcal{D} := \bigcup_{n=1}^{\infty} \mathcal{D}_n$. Note that $\mathcal{D}_1 \mathcal{D}_2 \subseteq \dots \subseteq \mathcal{D}$. The argument goes as follows: We prove the almost sure Hölder continuity on \mathcal{D} by chaining from one dyadic to another by a shortest path along neighboring dyadic points. Such increments are first estimated by Markov's inequality to obtain an almost sure uniform upper bound through the Borel-Cantelli lemma. The Hölder continuity is then extended to \mathcal{D}^c and this is the modification.



approximation on the dyadics

- (i) In a first step we estimate increments over neighboring dyadic numbers in \mathcal{D}^n by combining the Markov inequality and Borel-Cantelli. First note that

$$\mathbb{P}(|X_t - X_s| > \varepsilon) \stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}[|X_t - X_s|^\alpha]}{\varepsilon^\alpha} \stackrel{\text{ass.}}{\leq} c \frac{|t - s|^{1+\beta}}{\varepsilon^\alpha} \quad (4.6)$$

so that, in particular,

$$\mathbb{P}\left(|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}| \geq 2^{-\gamma n}\right) \leq c 2^{\alpha \gamma n} \cdot 2^{-n(1+\beta)} = c 2^{-n(1+\beta-\alpha\gamma)}.$$

Since a maximum is larger than a value if each element of the maximum is larger than the value, we obtain

$$\begin{aligned} \mathbb{P}\left(\underbrace{\max_{k \leq 2^n} |X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}|}_{=: A_n} \geq 2^{-\gamma n}\right) &\leq \mathbb{P}\left(\bigcup_{k \leq 2^n} \{|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}| \geq 2^{-\gamma n}\}\right) \\ &\stackrel{\text{sub. add.}}{\leq} \sum_{k=1}^{2^n} \mathbb{P}\left(|X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}| \geq 2^{-\gamma n}\right) \\ &\leq c \cdot \sum_{k=1}^{2^n} 2^{-n(1+\beta-\alpha\gamma)} \\ &= c \cdot 2^{-n(\beta-\alpha\gamma)}. \end{aligned}$$

The righthand side is summable by the choice of γ . Hence, Borel-Cantelli ?? implies that $\limsup A_n$ is a zero set. Defining $C := \{A_n \text{ i.o.}\}^C \in \mathcal{A}$ then yields $\mathbb{P}(A) = 1$, so that, for all $\omega \in A$, there are numbers $n_0(\omega)$ so that

$$\max_{k \leq 2^n} |X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}}| < \frac{1}{2^{\gamma n}}, \quad \forall n \geq n_0(\omega). \quad (4.7)$$

The important point is that we can estimate all increments over neighboring dyadics.

- (ii) The next step is the so-called chaining ("hangeln") trick. Using the above estimate for increments over neighboring dyadics we can chain cleverly from one dyadic to another by going only over neighboring dyadics of same type (see the picture).



chaining trick (orange is subset of blue)

Fix $\omega \in \mathcal{A}$ so that we can use (4.10) for all $m \geq N \geq n_0(\omega)$. Now suppose $m > N$ and $t, s \in \mathcal{D}_m$ are such that $|t - s| < \frac{1}{2^N}$. We use induction (starting at $m = N = 1$) to prove that

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{k=N+1}^m \frac{1}{2^{\gamma k}}$$

holds for all $m > N$:

- Induction start at $m = N+1$: For $m = N+1$, t and s must be neighbors ($|t-s| = \frac{1}{2^m} < \frac{1}{2^N}$) in \mathcal{D}_m , the claim follows from (4.10) as $m \geq n_0(\omega)$.
- Induction step: We suppose the claim holds for some $m-1 > N$ and take $s, t \in \mathcal{D}_m$. We chose $t', s' \in \mathcal{D}_{m-1}$ as in the picture (the closest to t, s)

$$\begin{aligned} |X_t(\omega) - X_s(\omega)| &\stackrel{m \geq N \geq n_0(\omega)}{\leq} |X_t(\omega) - X_{t'}(\omega)| + |X_{t'}(\omega) - X_{s'}^i(\omega)| + |X_{s'} - X_s(\omega)| \\ &\stackrel{\text{Ind. hyp.}}{\leq} 2 \frac{1}{2^{\gamma(m-1)}} + 2 \sum_{k=N+1}^{m-1} \frac{1}{2^{\gamma k}} + 2 \frac{1}{2^{\gamma(m-1)}} \\ &= 2 \sum_{k=N+1}^m \frac{1}{2^{\gamma k}} \end{aligned}$$

(iii) We can now show that $t \mapsto X_t(\omega)$ is Hölder continuous on \mathcal{D} for all $\omega \in C$. To do so define $h(\omega) = 2^{-n_0(\omega)}$. If $t, s \in \mathcal{D}$ are such that $|t-s| < h(\omega)$, then there is some $N \geq n_0(\omega)$ with $\frac{1}{2^{N+1}} \leq |t-s| \leq \frac{1}{2^N}$ so that

$$|X_t(\omega) - X_s(\omega)| \stackrel{(ii)}{\leq} 2 \sum_{k=N+1}^{\infty} \frac{1}{2^{\gamma k}} = \frac{2}{2^{\gamma(N+1)}} \sum_{k=0}^{\infty} \frac{1}{2^{\gamma k}} \leq \underbrace{K}_{:= \frac{2}{1-\frac{1}{2^\gamma}}} |t-s|^\gamma. \quad (4.8)$$

But this means that for all t there is a neighborhood (interval of radius $h(\omega)$) on which the Hölder estimate holds on \mathcal{D} with index γ and constant K . Hence, $t \mapsto X_t(\omega)$ is locally Hölder continuous with index γ for all $\omega \in C$.

(iv) We now define the modification \tilde{X} of X . Fix $t \notin \mathcal{D}$ and take any sequence (s_n) of dyadic numbers that decreases to t . By (4.8), for all $\omega \in C$ the sequence $(X_{s_n}(\omega))$ is a real Cauchy sequence, hence, converges. Then we define $\tilde{X}_t(\omega) := \lim_{n \rightarrow \infty} X_{s_n}(\omega)$ for $\omega \in C$ and $\tilde{X}_t(\omega) = 0$ otherwise. Note that (4.8) also shows that the limit is independent of the choice of the sequence so that, \mathbb{P} -almost surely,

$$\tilde{X}_t = \begin{cases} \lim_{s \rightarrow t, s \in \mathcal{D}} X_s & : t \notin \mathcal{D}, \\ X_t & : t \in \mathcal{D}, \end{cases} \quad t \in [0, 1].$$

Then

- \tilde{X} is a stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$ (limits of measurable random variables are measurable),
- $X_t = \tilde{X}_t$ for $t \in \mathcal{D}$,
- For all $\omega \in C$, $t \mapsto \tilde{X}(\omega)$ inherits the local Hölder continuity of index γ on $[0, 1]$ (\cdot is continuous)

It only remains to show that \tilde{X} is a modification of X . For $t \in \mathcal{D}$ we trivially have $\mathbb{P}(X_t = \tilde{X}_t) = 1$. For $t \notin \mathcal{D}$ we use that (4.6) implies that $X_s \xrightarrow{P} X_t$ as $s \rightarrow t$ for all $t \in [0, 1]$. Additionally, by definition of \tilde{X} , X_s converges almost surely to \tilde{X}_t so that in particular $X_s \xrightarrow{P} \tilde{X}_t$ as $s \rightarrow t$. But then $\mathbb{P}(X_t = \tilde{X}_t) = 1$. \square

There is a simple example which shows best how things work. If we return to the zero process X on $[0, 1]$ and the random one-point process $Y = \mathbf{1}_{\{U\}}$ from above, then Y is discontinuous and X is an almost surely continuous modification of Y .



Go through the construction and check why \tilde{X} from the proof is indistinguishable from X . Can you identify the set C ?

Let us now summarize the section for the Kolmogorov-Chentsov approach of constructing continuous stochastic processes with given finite-dimensional marginals:



- (i) Write down a consistent family $\{\mathbb{P}_J : |J| < \infty\}$ on I .
- (ii) Use Theorem 4.1.16 to obtain a stochastic process X on some $(\Omega, \mathcal{A}, \mathbb{P})$ with finite-dimensional marginals \mathbb{P}_J .
- (iii) Check the Hölder-type condition of Kolmogorov-Chentsov to obtain a continuous (even Hölder-continuous) modification \tilde{X} . Since modifications have the same finite dimensional marginals, \tilde{X} has the finite-dimensional marginals \mathbb{P}_J .

If of interest, use the law $\mathbb{P}_{\tilde{X}}$ of \tilde{X} induced on $\mathcal{B}(C([0, \infty)))$ with finite-dimensional marginals \mathbb{P}_J and work with the canonical construction.

There is a large class of processes for which the approach can be applied easily. Centered Gaussian processes are natural candidates as we already know how to construct from the covariance function K a family of consistent finite-dimensional distributions and the additional information $X_t - X_s \sim \mathcal{N}(0, K(t-s))$ is very well suited to check the Hölder-type condition from the explicit moment formulas of $\mathcal{N}(0, \sigma^2)$. In the next section the approach will be carried out for the Brownian motion.

4.2 Foundation of Brownian motion

We now come to the big final of this set of lectures notes. The Brownian motion. While the origin was to describe the two-dimensional movement of pollen in water by the biologist Robert Brown in 1827⁶ the first rigorous formulations date back to the beginning of the last century. The wide use of Brownian motion is perhaps best reflected in citing more applications in Einstein's treatment of the movement of molecules⁷ and Bachelier's famous model of a stock in his „Théorie de la spéculation“⁸. Also within mathematics the Brownian motion and its generalizations have turned out to be universal objects that appear very naturally when discrete structures are made to converge to continuous structures. The aim of this section is to present the foundations and to prove the universal approximation property of a Brownian motion in the simplest setting of stochastic processes.



Definition 4.2.1. A real-valued stochastic process $(B_t)_{t \geq 0}$ on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is called a **(standard) Brownian motion** if

- (i) $B_0 = 0$ \mathbb{P} -almost surely
- (ii) B has **stationary and independent increments**, i.e.
 - the distribution of $B_{t+h} - B_t$ depends on h but not on t ,
 - the increments $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_{t_0}$ are independent random variables for all $0 \leq t_0 \leq \dots \leq t_n$.

⁶Brown, Robert: "A brief account of microscopical observations made in the months of June, July and August, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies", Philosophical Magazine, 4 (21), 1828, 161–173.

⁷Einstein, Albert: "Investigations on the Theory of Brownian Movement", 1956, New York: Dover.

⁸Bachelier, Louis: "Théorie de la spéculation", Annales Scientifiques de l'École Normale Supérieure, 3 (17), 1900, 21–86.



- (iii) the paths $t \mapsto B_t$ are \mathbb{P} -almost sure continuous
- (iv) $B_t \sim \mathcal{N}(0, t)$ for all $t > 0$

The corresponding probability measure on $\mathcal{B}(C([0, \infty)))$ is called **Wiener measure**. If B_1, \dots, B_d are independent Brownian motions on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, then $B_t := (B_1, \dots, B_d))_{t \geq 0}$ is called a **d -dimensional Brownian motion**.

As discussed in the previous section if we prefer we can always use the canonical construction on the continuous functions as soon as the existence of a Brownian motion is known. This, in fact, is non-trivial! There are plenty of clever constructions, we will present a construction based on the Kolmogorov extension theorem and a more direct construction (which will not be discussed in class).

There are several equivalent definitions of a Brownian motion, viewing the Brownian motion as special case of two of the most important classes of stochastic processes.

- The first two properties are the so-called **Lévy properties**. A Lévy process is a stochastic process with sample paths that are right continuous with left limits (RCLL) that start in 0 and have stationary and independent increments. A Brownian motion is the only non-deterministic Lévy process with continuous sample paths.
- A Brownian motion is a centered Gaussian process process with covariance function $K(s, t) = \min(s, t)$.

Since we will use the Gaussian point of view let us prove the claimed equivalence:



Proposition 4.2.2. A real-valued stochastic process X with index-set $I = [0, \infty)$ is a Brownian motion if and only if X is a centered Gaussian process with covariance function $K(s, t) = \min(t, s)$ and continuous sample paths.

Proof. " \Rightarrow ": First note that the increments $X_t - X_s$ are $\mathcal{N}(0, t - s)$ by combining properties (i), (ii), and (iii) since $B_t - B_s \sim B_{t-s} - B_0 \sim B_{t-s} \sim \mathcal{N}(0, t - s)$. Writing

$$\begin{pmatrix} B_s \\ B_t \end{pmatrix} = \begin{pmatrix} B_s \\ B_s + (B_t - B_s) \end{pmatrix} = A \cdot \begin{pmatrix} B_s \\ B_t - B_s \end{pmatrix}$$

with $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ shows that (B_s, B_t) is a Gaussian vector as $(B_t, B_t - B_s)$ is a vector of independent Gaussian random variables. The same trick works for all finite dimensional distributions X_1, \dots, X_n which can be written as matrix multiplication of some (more complicated) matrix with a vector of increments which are independent and Gaussian. Hence, we proved that B is a Gaussian process. Furthermore, as $B_t \sim \mathcal{N}(0, t)$ the Gaussian process is also centered. To compute the covariance function $K(s, t)$ we proceed similarly:

$$\begin{aligned} \varphi_{(X_s, X_t)}(t_1, t_2) &= \mathbb{E}[\exp(i(t_1 X_s + t_2 X_t))] \\ &= \mathbb{E}[\exp(i(t_1 + t_2) X_s) \exp(it_2(X_t - X_s))] \\ &\stackrel{\text{ind. incr.}}{=} \mathbb{E}[\exp(i(t_1 + t_2) X_s)] \mathbb{E}[\exp(it_2(X_t - X_s))] \\ &= \exp\left(-\frac{1}{2}(t_1 + t_2)^2 s^2\right) \exp\left(-\frac{1}{2}t_2(t - s)^2\right) \\ &= \exp\left(-\frac{1}{2}\langle \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}, \begin{pmatrix} s & s \\ s & t \end{pmatrix} \cdot \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \rangle\right), \end{aligned}$$

which is the characteristic function of a $\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s & s \\ s & t \end{pmatrix}\right)$ random vector. Hence, $K(s, t) = \min(s, t)$.

" \Leftarrow ": We check the three defining properties of a Brownian motion:

(i) $\text{Var}[X_0] = \text{Cov}(X_0, X_0) = \min(0, 0) = 0$, hence, $X_0 = 0$ almost surely.

(ii) The assumption implies that the finite dimensional marginals $(X_{t_0}, \dots, X_{t_n})$ are centered Gaussian vectors. Writing the vector of increments $(X_{t_n} - X_{t_{n-1}}, \dots, X_{t_1} - X_{t_0})$ as a matrix multiplication of $(X_{t_0}, \dots, X_{t_n})$ shows that the vector of increments is centered Gaussian, hence, is uniquely determined by all covariances. To show that all covariances are zero we use linearity of expectations. For $t_k \geq t_{k-1} \geq t_j \geq t_{j-1}$ this gives

$$\begin{aligned} & \text{Cov}(X_{t_j} - X_{t_{j-1}}, X_{t_k} - X_{t_{k-1}}) \\ &= \text{Cov}(X_{t_j}, X_{t_k}) - \text{Cov}(X_{t_{j-1}}, X_{t_k}) + \text{Cov}(X_{t_{j-1}}, X_{t_{k-1}}) - \text{Cov}(X_{t_j}, X_{t_{k-1}}) \\ &= t_j - t_{j-1} + t_{j-1} - t_j = 0. \end{aligned}$$

Hence, the covariance matrix is a diagonal matrix and thus the increments are independent. Since (X_t, X_{t+h}) is a Gaussian vector also their linear combination $X_{t+h} - X_t$ is Gaussian. The variance can be computed from the assumed variance structure:

$$\text{Var}[X_{t+h} - X_t] = \text{Var}[X_{t+h}] - 2\text{Cov}(X_{t+h}, X_t) - \text{Var}[X_t] = (t+h) - 2t + t = h.$$

(iii) The continuous sample paths have been assumed. \square

Before continuing to discuss any properties let us prove the existence of a Brownian motion.



Theorem 4.2.3. There is a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a stochastic process $(X_t)_{t \geq 0}$ on $(\Omega, \mathcal{A}, \mathbb{P})$ that satisfies the properties of an Brownian motion. In other words: Brownian motion exists.

There are many ways of constructing the Brownian motion, we give two proofs.

Abstract proof. The proof uses the abstract Kolmogorov theory of the previous section in combination with the previous proposition.



There is a centered Gaussian process X indexed by $I = [0, \infty)$ with covariance function $K(s, t) = \min(s, t)$.

This follows from Kolmogorov's extension theorem for Gaussian processes with the symmetric positive semidefinite function $K(t, s) = \min(t, s)$.



There is a modification B of X with continuous sample paths.

We use that $X_t - X_s \sim \mathcal{N}(0, t-s)$ and $\mathbb{E}[X^4] = 3\sigma^4$ if $X \sim \mathcal{N}(0, \sigma^2)$. Hence, the Kolmogorov-Chentsov property holds with $\alpha = 4$, $\beta = 1$, and $K = 3$ even with equality:

$$\mathbb{E}[|X_t - X_s|^4] = 3|t-s|^{1+1}.$$

But then X has a modification B with continuous (even Hölder-continuous with index $\gamma < \frac{1}{4}$) sample paths.



B is a Brownian motion.

Being a modification B has the same finite dimensional marginals as X . Hence, B is a centered Gaussian process with covariance function $K(t, s) = \min(t, s)$. Since B has continuous samples paths the theorem follows from Proposition 4.2.2. \square

Lecture 23

Here is a second proof that is more, not using the Kolmogorov extension theorem. The proof is not part of the lecture.

Lévy's construction of the Brownian motion. The Brownian motion will only be constructed on $[0, 1]$. Pasting together ⁹independent copies of $(B_t)_{t \in [0, 1]}$ one can check readily that the defining properties hold. As in the proof of Kolmogorov-Chentsov we work with the dyadic numbers $\mathcal{D}_n = \{\frac{k}{2^n} : k = 0, \dots, 2^n\}$ in $[0, 1]$. The idea is to define a sequence of "discrete Brownian motions" on all \mathcal{D}_n , i.e. stochastic processes that satisfy the right distributional properties (properties (i) and (ii)) on \mathcal{D}_n and are linearly interpolated in between. The entire sequence is constructed on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and converges almost surely uniformly on $[0, 1]$. The uniform convergence ensures that the limit has continuous paths (property (iii)), the distributional properties hold on the union \mathcal{D} of the \mathcal{D}_n , thus, by continuity of sample paths also on $[0, 1]$ as the dyadics are dense.



Construction of the probability space and the discrete Brownian motion on \mathcal{D}_n .

Since \mathcal{D} is countable there is a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a sequence of iid $\mathcal{N}(0, 1)$ random variables $(Z_d)_{d \in \mathcal{D}}$ on $(\Omega, \mathcal{A}, \mathbb{P})$. This actually uses the Kolmogorov extension theorem to ensure the existence of the iid sequence (compare Theorem ??). The construction of the discrete Brownian motions on \mathcal{D}_n is inductively:

- $B_0 = 0, B_1 = Z_1$
- Suppose B_d are already defined for all $d \in \mathcal{D}_{n-1}$. Then define $B_d, d \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}$ as

$$B_d = \frac{B_{d-\frac{1}{2^n}} + B_{d+\frac{1}{2^n}}}{2} + \frac{Z_d}{2^{\frac{(n+1)}{2}}}.$$

The construction is best understood in a picture:¹⁰



Check the distributional properties (i) and (ii) of a Brownian motion on the dyadics.

The construction ensures that $(B_d)_{d \in \mathcal{D}_n}$ and $(Z_d)_{d \in \mathcal{D} \setminus \mathcal{D}_n}$ are independent random vectors. This can be used to show that

- $B_t - B_s \sim \mathcal{N}(0, t - s)$ for $s < t$ in \mathcal{D}_n ,
- $B_s - B_r$ independent of $B_t - B_s$ for $r < s < t$ in \mathcal{D}_n .

The first property follows immediately from the construction, the second is a bit more involved. For the second recall two facts on Gaussian vectors:

- X, Y independent $\mathcal{N}(0, \sigma^2)$ implies that $X + Y, X - Y$ are independent $\mathcal{N}(0, 2\sigma^2)$.
- If (X_1, \dots, X_n) is a Gaussian vector, then the X_1, \dots, X_n are independent iff X_1, \dots, X_n is pairwise independent.

The second statement follows directly by checking the covariance matrix Σ , the first by writing

$$\begin{pmatrix} X + Y \\ X - Y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} X \\ Y \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^T = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

⁹Bild

¹⁰Bild

The distribution of the increments is now computed by induction. A bit similarly to the proof of Kolmogorov-Chentsov we first consider neighboring dyadics. Rewriting the definition of B_d for $d \in \mathcal{D}_{n+1}$ gives

$$B_d - B_{d-2^{-n}} = \frac{B_{d+2^{-n}} - B_{d-2^{-n}}}{2} + \frac{Z_d}{2^{\frac{(n+1)}{2}}}$$

and

$$B_{d+2^{-n}} - B_d = \frac{B_{d+2^{-n}} - B_{d-2^{-n}}}{2} - \frac{Z_d}{2^{\frac{(n+1)}{2}}}$$

The summands on the right hand side are independent and both $\mathcal{N}(0, 2^{-(n+1)})$ by the induction hypothesis. Hence, from the above, the two increments are independent and $\mathcal{N}(0, 2^{-n})$. The same holds for increments of $d \in \mathcal{D}_{n-1}$ which are not adjacent. (Bild) Left increment only constructed from increments in $(n-1)$ st step on the left, right increment from increments in $(n-1)$ st step on the right (those are independent by induction hypothesis) plus independent of Z_d . \Rightarrow increments are independent. Hence, the (Gaussian) increments are pairwise independent and, thus, independent. But then all increments are also independent.

Now we need to transfer from \mathcal{D}_n to $[0, 1]$. We now consider the functions obtained from the B_d in \mathcal{D}_n by linear interpolation. Note: Once a point $t \in \mathcal{D}_n$ is set it is not changed again. Claim: The sequence converges uniformly on $[0, 1]$ to a (continuous) function. Why? Let us formalize the functions:

$$F_0(t) := \begin{cases} Z_1 & : t = 1 \\ 0 & : t = 0 \\ \text{linear in between} & \text{else} \end{cases}$$

$$F_n(t) := \begin{cases} \frac{Z_t}{2^{\frac{(n+1)}{2}}} & : t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1} \\ 0 & : t \in \mathcal{D}_{n-1} \\ \text{linear in between} & \text{else} \end{cases}$$

Then for all $t \in \mathcal{D}_n$:

$$B_d(t) = \sum_{k=0}^n F_k(t) = \sum_{k=0}^{\infty} F_k(t)$$

We use the series representation to prove the uniform convergence. Recall from Stochastik 1 (exercise) that $\mathbb{P}(|Z_d| \geq c \cdot \sqrt{n}) \leq \exp(-\frac{c^2 n}{2})$ holds for $Z_d \sim \mathcal{N}(0, 1)$. So,

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbb{P}(\exists d \in \mathcal{D}_k : |Z_d| \geq c\sqrt{k}) &\stackrel{\text{subadd.}}{\leq} \sum_{k=0}^{\infty} \sum_{d \in \mathcal{D}_k} \mathbb{P}(|Z_d| \geq c\sqrt{k}) \\ &\leq \sum_{k=0}^{\infty} \sum_{d \in \mathcal{D}_k} e^{-\frac{c^2 k}{2}} \\ &= \sum_{k=0}^n (2^k + 1) e^{-\frac{c^2 k}{2}} \stackrel{*}{<} \infty \end{aligned}$$

* if $c > \sqrt{2 \log(2)}$ If we fix such a c , then Borel-Cantelli implies $\mathbb{P}(\exists d \in \mathcal{D}_k : |Z_d| \geq c\sqrt{k} \text{ i.o.}) = 0$. Hence, a.s. there is a random $N = N(\omega)$ such that $|Z_d| < c\sqrt{n}$ for all $d \in \mathcal{D}_n$ and $n \geq N$. In terms of the (random) functions F_n this means that $\|F_n\|_{\infty} \leq c \cdot \sqrt{n} \cdot 2^{-\frac{(n+1)}{2}}$ $\forall n \geq N$. But then

$$\left\| \sum_{k=0}^{\infty} F_k - \sum_{k=0}^n F_k \right\|_{\infty} \stackrel{\triangle}{\leq} \sum_{k=n+1}^{\infty} \|F_k\|_{\infty} \leq \sum_{k=n+1}^{\infty} c\sqrt{n} 2^{-\frac{(k+1)}{2}} \rightarrow 0, n \rightarrow \infty$$

Hence, $B := \sum_{k=0}^{\infty} F_k$ is almost surely the uniform limit of continuous functions and as such also continuous.

Claim: B satisfies the properties of a Brownian motion on $[0, 1]$.

(i) ✓

(ii) Use that centered Gaussian r.v. are independent if all pairwise covariances are zero.

Fix $t_1 < \dots < t_n$ and sequences $(t_{i,k})_{i,k \in \mathbb{N}} \subseteq \mathcal{D}$ with $t_{1,k} < \dots < t_{n,k}$ and $t_{i,k} \downarrow t_i$. As in (ii), by continuity, vectors $(B_{t_n} - B_{t_{n-1}}, \dots, B_{t_1} - B_0)$ are a.s. limits of $(B_{t_{n,k}} - B_{t_{n-1,k}}, \dots, B_{t_{1,k}} - B_{0,k})$. Hence, they are Gaussian vectors for which we only need covariances. Since the Brownian properties hold on \mathcal{D} we get

$$\begin{aligned}\text{Cov}(B_{t_{i,k}} - B_{t_{i-1,k}}, B_{t_{j,k}} - B_{t_{j-1,k}}) &= \mathbf{1}_{i=j}(t_{i,k} - t_{i-1,k}) \\ &\stackrel{k \rightarrow \infty}{=} \mathbf{1}_{i=j}(t_i - t_{i-1}) \\ &= \text{Cov}(B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}})\end{aligned}$$

Hence, the increments are independent and $B_{t+h} - B_t \sim \mathcal{N}(0, h)$ which is independent of h .

(iii) ✓

(iv) Statement holds for $t_k \in \mathcal{D}$ which is dense in $[0, 1]$. We use a further property of Gaussian vectors: If (X_n) is a sequence of Gaussian vectors for which the expectation vectors and covariance matrices converge and (X_n) converge a.s. to a random vector X , then X is Gaussian with the limiting expectation vector and covariance matrix.

Take a sequence $(t_n)_{n \in \mathbb{N}} \subseteq \mathcal{D}$ with $t_n \rightarrow t$, $n \rightarrow \infty$. Since B is continuous, $B_{t_n} \rightarrow B_t$, $n \rightarrow \infty$. Using that $B_{t_n} \sim \mathcal{N}(0, t_n)$ we obtain $B_t \sim \mathcal{N}(0, t)$.

□

Now that the existence is settled we can turn towards properties of the Brownian motion.


Proposition 4.2.4. (Brownian scaling property)

If B is a Brownian motion and $a > 0$, then the process $X_t := \frac{1}{a}B_{a^2 \cdot t}$, $t \geq 0$, is also a Brownian motion.

Proof. All we need to do is check the defining properties:

(i) ✓

(ii) Since

$$(X_{t_n} - X_{t_{n-1}}, \dots, X_{t_1} - X_{t_0}) = \frac{1}{a}(B_{a^2 t_n} - B_{a^2 t_{n-1}}, \dots, B_{a^2 t_1} - B_{a^2 t_0})$$

the independence of increments is inherited from X , just as the independence of h in $X_{t+h} - X_t = \frac{1}{a}(B_{a^2 t+a^2 h} - B_{a^2 t})$ is inherited.

(iii) ✓

(iv) $X_t = \frac{1}{a}B_{a^2 t} \sim \mathcal{N}(0, t)$ follows from the scaling property of $\mathcal{N}(\mu, \sigma^2)$

□

In order to get an idea why the scaling property is useful let us consider first hitting times. The scaling property tell us that enlarging (resp. shrinking) space by a constant has the same effect than speeding up/slowing down time with the square-root of the constant. Morally, the time to a point which is b -times further away should thus be b^2 -times longer. Here is an example computation for

$$T(a, b) = \inf\{t: B_t = a \text{ or } B_t = b\}.$$

Using the scaling property gives

$$T(a, b) \stackrel{(d)}{=} \inf\left\{t: bB_{\frac{1}{b^2}t} = a \text{ or } bB_{\frac{1}{b^2}t} = b\right\} = b^2 \inf\left\{t: B_t = \frac{a}{b} \text{ or } B_t = 1\right\} = b^2 T\left(\frac{a}{b}, 1\right).$$

In particular, choosing $a = b$ we have the identity $\mathbb{E}[T(-b, b)] = b^2 \mathbb{E}[T(-1, 1)]$. This is a very typical application of the scaling property, in order to compute the expectation of all hitting times it is only needed to compute $\mathbb{E}[T(-1, 1)]$.

Extremely useful tools to work with the Brownian motion are the Markov and martingale property for which we use the natural filtration $\mathcal{F}_s := \sigma(B_t : t \leq s)$. Here are the two properties:



Proposition 4.2.5. Let $(B_t)_{t \geq 0}$ be a Brownian motion.

- (i) B is a simple Markov process, that is,

$$B_t^{(s)} := B_{t+s} - B_s, \quad t \geq 0,$$

is a Brownian motion independent of \mathcal{F}_s for all $s \geq 0$.

- (ii) B is a continuous-time martingale, that is,

$$\mathbb{E}[|B_t|] < \infty \quad \text{and} \quad \mathbb{E}[B_t | \mathcal{F}_s] = B_s \quad \text{a.s. for all } t \geq s.$$

Before proceeding let us recall the definition of independence to be clear what the statement means and how it should be proved. The statement of (i) is to be read as independence of $\mathcal{F}_s = \sigma(B_t : t \leq s)$ from $\sigma(B_t^{(s)} : t \geq 0)$ because independence of random variables is always defined through the generated σ -algebras.

Proof. (i) Independence of σ -algebras only needs to be checked on arbitrary generators, Proposition ???. Hence, it suffices to check that B_r and $B_t^{(s)}$ are independent random variables for all choices of $r \leq s$ and $t \geq 0$. Since the Brownian motion $(B_t)_{t \geq 0}$ is a Gaussian process the vector (B_r, B_s, B_{s+t}) of finite dimensional marginals is a Gaussian vector. Multiplying the vector with a cleverly chosen matrix (which?) gives $(B_r, B_{s+t} - B_s)$ which is then also a Gaussian vector. But then the random variables are independent if and only if they are uncorrelated (generally wrong, jointly Gaussian is needed here). Since both are centered and the covariance function of the Brownian motion is $K(s, t) = \min(s, t)$ this is simple:

$$\begin{aligned} \text{Cov}(B_r, B_{s+t} - B_s) &= \mathbb{E}[B_r(B_{s+t} - B_s)] \\ &= \mathbb{E}[B_r B_{s+t}] - \mathbb{E}[B_r B_s] \\ &= \text{Cov}(B_r, B_{s+t}) + \text{Cov}(B_r, B_s) \\ &= r - r = 0 \end{aligned}$$

- (ii) Trick: Use $B_t = (B_t - B_s) + B_s$ and the statements from (i)

$$\mathbb{E}[B_t | \mathcal{F}_s] = \mathbb{E}[B_t - B_s | \mathcal{F}_s] + \mathbb{E}[B_s | \mathcal{F}_s] = \mathbb{E}[B_t^{(s)} | \mathcal{F}_s] + B_s \stackrel{(i)}{=} \underbrace{\mathbb{E}[B_t^{(s)}]}_{=0} + B_s = B_s.$$

□

Here is a nice application of the martingale property:



Proposition 4.2.6. (Brownian law of large numbers)

If $(B_t)_{t \geq 0}$ is a Brownian motion, then $\lim_{t \rightarrow \infty} \frac{1}{t} B_t = 0$ a.s.

Proof. It follows from the definition that the increments $(B_n - B_{n-1})_{n \in \mathbb{N}}$ are iid $\mathcal{N}(0, 1)$ because independence of infinitely many events or random variables is only defined through all finite subsets of the index set (Definition ??). Hence, the strong law of large numbers yields

$$\lim_{n \rightarrow \infty} \frac{1}{n} B_n = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (B_{t_k} - B_{t_{k-1}}) \stackrel{\text{LLN}}{=} \mathbb{E}[B_{t_1} - 0] = 0$$

almost surely. Thus the statement of the theorem is simple on \mathbb{N} and it suffices to control B between the integers. For that purpose define

$$Y_n := \sup_{t \in [n, n+1]} |B_t - B_n| = \sup_{t \in [n, n+1] \cap \mathbb{Q}} |B_t - B_n|.$$

The second equality holds almost surely and follows from the continuity of sample paths. Then Y_0, Y_1, \dots is an iid sequence because $B_t - B_n$ and $B_s - B_m$ are independent increments. If we can prove that $\mathbb{E}[|Y_0|] < \infty$ then, again by the strong law of large numbers,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^n Y_k = \mathbb{E}[Y_0] < \infty$$

almost surely which in particular implies $\frac{1}{n} Y_n \rightarrow 0$, $n \rightarrow \infty$, almost surely. Combined with the law of large number on the integers it follows that ??? obere Gaußklammer?

$$0 = \liminf_{t \rightarrow \infty} \frac{B_{\lfloor t \rfloor} + Y_{\lfloor t \rfloor}}{t} \leq \liminf_{t \rightarrow \infty} \frac{1}{t} B_t \leq \limsup_{t \rightarrow \infty} \frac{1}{t} B_t \leq \limsup_{t \rightarrow \infty} \frac{B_{\lfloor t \rfloor} + Y_{\lfloor t \rfloor}}{t} = 0$$

almost surely and the proof of the theorem is complete. However, we need to verify that $\mathbb{E}[|Y_0|] < \infty$ which, surprisingly, follows from martingale theory. Define

$$M_n^{(m)} := B_{n \cdot 2^{-m}}, \quad n = 0, \dots, 2^m,$$

which is a $(\mathcal{F}_n^m)_{n=0, \dots, 2^m}$ martingale. In fact, $M^{(m)}$ is the Brownian motion at discrete times so that the martingale property follows by restricting process and filtration to the same increasing subset of time. The Doob inequality with $p = 2$ then implies

$$\mathbb{E} \left[\left(\sup_{n \leq 2^m} M_n^{(m)} \right)^2 \right] \stackrel{\text{Doob for } p=2}{\leq} 4 \cdot \mathbb{E}[(M_{2^m}^{(m)})^2] = 4 \cdot \mathbb{E}[B_1^2] = 4,$$

for all $m \in \mathbb{N}$. Hence, the sequence $(\sup_{n \leq 2^m} M_n^{(m)})_{m \in \mathbb{N}}$ is bounded in L^2 , in particular uniformly integrable. The proof can be finished the following nice trick. Continuity of Brownian motion implies

$$Y_0 = \sup_{t \in [0, 1]} |B_t| = \lim_{m \rightarrow \infty} \sup_{n \leq 2^m} M_n^{(m)} \text{ a.s.}$$

We proved that u.i. and a.s. convergece implies L^1 -convergence. Hence, $Y_0 \in \mathcal{L}^1$. \square

¹¹The law of large number is a very rough statement about the fluctuations of B at infinity. It can be interpreted such that a Brownian motion eventually stays in every linear cone. In fact, a much better statement holds, a Brownian motion behaves more than the square-root of its time. The law of the iterated logarithm states that path stay in $[-\sqrt{2t \log \log t}, \sqrt{2t \log \log t}]$ almost surely and this is the smallest possible as $\lim_{t \rightarrow \infty} \frac{|B_t|}{\sqrt{2t \log \log t}} = 1$ almost surely.



Proposition 4.2.7. (time-inversion)

If $(B_t)_{t \geq 0}$ is a Brownian motion, then also

$$X_t := \begin{cases} t \cdot B_{\frac{1}{t}} & , t > 0 \\ 0 & , t = 0 \end{cases}$$

is a Brownian motion.

Proof. • X is a Gaussian process

- $\text{Cov}(X_t, X_s) = t \cdot s \cdot \text{Cov}(B_{\frac{1}{t}}, B_{\frac{1}{s}}) = t \cdot s \cdot \min(\frac{1}{t}, \frac{1}{s}) = \min(t, s)$

¹¹Bild!

- X is continuous for $t > 0$ as concatenation of continuous functions

$$\lim_{t \rightarrow \infty} X_t = \lim_{t \rightarrow \infty} t \cdot B_{\frac{1}{t}} = \lim_{t \rightarrow \infty} \frac{1}{t} B_t = 0 = X_0$$

Hence, X is almost sure continuous on $[0, \infty)$.
 $\Rightarrow X$ is a BM □

Why is time-inversion useful? If you can prove theorems for $t \rightarrow 0$ you get theorems for $t \rightarrow \infty$ for free (and vice versa). The next theorem is extremely useful to study very carefully the path behavior of the Brownian sample paths.



Theorem 4.2.8. (Blumenthal 0-1-law)

Define $\mathcal{F}_{0+} := \bigcap_{s>0} \mathcal{F}_s$, then \mathcal{F}_{0+} is trivial, i.e. $\mathbb{P}(A) \in \{0, 1\}$ for all $A \in \mathcal{F}_{0+}$.

Proof. Let $A \in \mathcal{F}_{0+}$, $0 < t_1 < \dots < t_k$ and $g: \mathbb{R}^k \rightarrow \mathbb{R}$ bounded continuous. Using continuity of g and almost sure continuity of Brownian paths yields

$$\mathbb{E}[\mathbf{1}_A g(B_{t_1}, \dots, B_{t_k})] \stackrel{\text{DCT}}{=} \lim_{\varepsilon \rightarrow 0} \mathbb{E}[\mathbf{1}_A \cdot g(B_{t_1} - B_\varepsilon, \dots, B_{t_k} - B_\varepsilon)]$$

If $\varepsilon < t_1$ then by the simple Markov property the random variables $B_{t_1} - B_\varepsilon, \dots, B_{t_k} - B_\varepsilon$ are independent of \mathcal{F}_ε and thus independent of its sub- σ -algebra \mathcal{F}_{0+} (the independence property needs to hold only for less events). Hence, for all $A \in \mathcal{F}_{0+}$,

$$\mathbb{E}[\mathbf{1}_A g(B_{t_1}, \dots, B_{t_k})] = \lim_{\varepsilon \rightarrow 0} \mathbb{P}(A) \mathbb{E}[g(B_{t_1} - B_\varepsilon, \dots, B_{t_k} - B_\varepsilon)] = \mathbb{P}(A) \mathbb{E}[g(B_{t_1}, \dots, B_{t_k})]. \quad (4.9)$$

From (4.9) one can directly check that \mathcal{F}_{0+} is independent of $\sigma(B_{t_1}, \dots, B_{t_k})$. To see why recall that $\sigma(B_{t_1}, \dots, B_{t_k})$ consists of preimages of the random vectors $(B_{t_1}, \dots, B_{t_k})$ so that (4.9) with $\mathbf{1}_C(B_{t_1}, \dots, B_{t_k})$ for all $C \in \mathcal{B}(\mathbb{R}^k)$ yields the independence. We approximate as usually $\mathbf{1}_C$ by the continuous bounded functions $g = f_C^\varepsilon$ from the proof of Proposition ?? and use dominated convergence to get rid of ε . Then (4.9) holds for all indicators of events from $\sigma(B_{t_1}, \dots, B_{t_k})$ so that the definition of independence is checked.

Since this holds for all $0 < t_1 < \dots < t_k$ this shows \mathcal{F}_{0+} is independent of $\sigma(B_t : t > 0)$. Next, $\sigma(B_t : t > 0)$ is equal to $\sigma(B_t : t \geq 0)$ because $B_t = \lim_{s \rightarrow 0} B_{t+s}$, so that B_0 is $\sigma(B_t : t > 0)$ -measurable. Since $\mathcal{F}_{0+} \subseteq \sigma(B_t : t \geq 0)$ which shows \mathcal{F}_{0+} is independent of itself, that is, $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A) \cdot \mathbb{P}(A)$ which implies $\mathbb{P}(A) \in \{0, 1\}$. □

As an application we deduce some path properties:



Corollary 4.2.9. We have $\forall \varepsilon > 0$, almost surely,

$$\sup_{t \in [0, \varepsilon]} B_t > 0 \text{ and } \inf_{t \in [0, \varepsilon]} B_t < 0$$

With the simple Markov property the same holds everywhere.

Message: Brownian path fluctuate enormously!

Proof. Let $\varepsilon_n \downarrow 0$ and define

$$A := \bigcap_{n \in \mathbb{N}} \{\omega \in \Omega : \sup_{0 \leq s \leq \varepsilon_n} B_s > 0\} \in \mathcal{F}_{0+}$$

so that $\mathbb{P}(A) \in \{0, 1\}$. By monotonicity of measures we have

$$\mathbb{P}(A) = \lim_{n \rightarrow \infty} \mathbb{P}(\sup_{s \leq \varepsilon_n} B_s > 0) \geq \lim_{n \rightarrow \infty} \mathbb{P}(B_{\varepsilon_n} > 0) \stackrel{\sim \mathcal{N}(0, \varepsilon_n)}{=} \frac{1}{2}.$$

Hence, $\mathbb{P}(A) = 1$. For the infimum just replace B by $-B$ which is also a BM. □



Theorem 4.2.10. Almost surely the Brownian paths are not monotone on any bounded interval

Beweis. The claim is a direct consequence of Corollary 4.2.9 and the simple Markov property. By the simple Markov property we find a.s for all $q \in \mathbb{Q}_+$ that

$$0 < \sup_{t \leq \varepsilon} B_t^{(q)} \stackrel{\text{Def}}{=} \sup_{t \leq \varepsilon} (B_{q+t} - B_q) = \sup_{t \leq \varepsilon} B_{q+t} - B_q,$$

hence $B_q < \sup_{t \leq \varepsilon} B_{q+t}$ and similarly $\inf_{t \leq \varepsilon} B_{q+t} < B_q$. \square



Theorem 4.2.11. The sample paths of a Brownian motion are almost surely not Hölder continuous of index $\frac{1}{2}$ but they are Hölder continuous of index γ for all $\gamma < \frac{1}{2}$.

, i.e. there is $A \in \mathcal{A}$, $\mathbb{P}(A) = 1$ and $t \rightarrow B_t(\omega)$ not $\frac{1}{2}$ -Hölder at 0 $\forall \omega \in A$.
Simple markov implies BM is a.s. not $\frac{1}{2}$ -Hölder continuous at every fixed $s > 0$.

Proof. We show that sample paths are not Hölder continuous at 0.

$$\begin{aligned} \mathbb{P}\left(\inf\{t > 0 : B_t \geq k \cdot \sqrt{t}\} = 0\right) &\quad \forall k < 0 \\ \mathbb{P}\left(\inf\{t > 0 : B_t \leq -k \cdot \sqrt{t}\} = 0\right) &\quad \forall k < 0 \end{aligned}$$

Bild fehlt!

Why? With $A_s := \{\inf\{t > 0 : B_t \geq k \cdot \sqrt{t}\} \leq s\} \in \mathcal{F}_s$ we have $A = \bigcap_{s>0} A_s \in \mathcal{F}_{0+}$. Hence, $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$. Brownian scaling gives

$$\begin{aligned} \mathbb{P}(A) &\stackrel{\text{monot. of meas.}}{=} \lim_{s \rightarrow 0} \mathbb{P}(A_s) \\ &= \lim_{s \rightarrow 0} \mathbb{P}\left(\inf\{t > 0 : \frac{1}{\sqrt{t}} B_t \geq k\} \leq s\right) \\ &\stackrel{\text{monoton}}{\geq} \lim_{s \rightarrow 0} \mathbb{P}\left(\frac{1}{\sqrt{s}} B_s \geq k\right) \\ &\stackrel{\text{scaling}}{=} \lim_{s \rightarrow 0} \mathbb{P}(B_1 \geq k) \\ &= \mathbb{P}(B_1 \geq k) > 0 \end{aligned}$$

Blumenthal 0-1 gives $\mathbb{P}(A) = 1$. If $t \rightarrow B_t(\omega)$ $\frac{1}{2}$ -Hölder at 0 then $|B_t(\omega) - B_0(\omega)| \leq K|t - 0|^{-0.5}$, t small $\Rightarrow \frac{1}{\sqrt{t}}|B_t| \leq k$. Hence, $\omega \in A^C$ with $\mathbb{P}(A^C) = 0$ as above. \square

In fact, much more is known:

- (i) For all $\gamma \geq \frac{1}{2}$, almost surely the Brownian sample paths are nowhere γ -Hölder continuous.
- (ii) For all $\gamma < \frac{1}{2}$, almost surely the Brownian sample paths are everywhere γ -Hölder continuous.

We could prove (ii) in this lecture ((i) is harder) but only prove non-differentiability:



Theorem 4.2.12. The path of a Brownian motion is almost surely nowhere differentiable, i.e. there is $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ and $t \rightarrow B_t(\omega)$ is nowhere differentiable for all $\omega \in A$.

Beweis. We show that, almost surely, for every $t_0 \geq 0$ the rightderivate $\lim_{h \downarrow 0} \frac{B_{t_0+h} - B_{t_0}}{h}$ does not exist. Then B is not differenetiabale in t_0 . Define

$$\begin{aligned} D^+ f(t) &= \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \\ D^- f(t) &= \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \end{aligned}$$

and $A := \{\omega \in \Omega \mid \exists t_0 \in [0, 1]: -\infty < D^- B_{t_0}(\omega) \leq D^+ B_{t_0}(\omega) < +\infty\}$. We will show $\mathbb{P}(A) = 0$, i.e. no limits of the differential quotient exist. Define

$$A_M = \left\{ \omega \in \Omega \mid \exists t_0 \in [0, 1]: \sup_{h \in [0, 1]} \left| \frac{B_{t_0+h}(\omega) - B_{t_0}(\omega)}{h} \right| \leq M \right\}$$

so that $A \subseteq \bigcup_{M=1}^{\infty} A_M$. We show that $\mathbb{P}(A) = 0$ for all M . Now fix some $n \geq 3$ and define the events

$$\begin{aligned} F_{n,k}^{(1)} &:= \left\{ \omega \in \Omega: \left| B_{\frac{k+1}{2^n}}(\omega) - B_{\frac{k}{2^n}}(\omega) \right| \leq \frac{3M}{2^n} \right\} \\ F_{n,k}^{(2)} &:= \left\{ \omega \in \Omega: \left| B_{\frac{k+2}{2^n}}(\omega) - B_{\frac{k+1}{2^n}}(\omega) \right| \leq \frac{5M}{2^n} \right\} \\ F_{n,k}^{(3)} &:= \left\{ \omega \in \Omega: \left| B_{\frac{k+3}{2^n}}(\omega) - B_{\frac{k+2}{2^n}}(\omega) \right| \leq \frac{7M}{2^n} \right\} \end{aligned}$$

Claim: $A_M \subseteq \bigcup_{k=1}^{2^n} F_{n,k}$ with $F_{n,k} = F_{n,k}^{(1)} \cap F_{n,k}^{(2)} \cap F_{n,k}^{(3)}$. Why? If $\omega \in A_M$ with $t_0 \in [\frac{k-1}{2^n}, \frac{k}{2^n}]$ for some $k \leq 2^n$, then Bild (!!!) and $|B_{t_0+h} - B_{t_0}| \leq h \cdot M$ gives:

$$\begin{aligned} \left| B_{\frac{k+1}{2^n}}(\omega) - B_{\frac{k}{2^n}}(\omega) \right| &\stackrel{\Delta}{=} \left| B_{\frac{k+1}{2^n}}(\omega) - B_{t_0}(\omega) \right| + \left| B_{t_0}(\omega) - B_{\frac{k}{2^n}}(\omega) \right| \leq \frac{3M}{2^n} \Rightarrow \omega \in F_{n,k}^{(1)} \\ \left| B_{\frac{k+2}{2^n}}(\omega) - B_{\frac{k+1}{2^n}}(\omega) \right| &\stackrel{\Delta}{=} \left| B_{\frac{k+2}{2^n}}(\omega) - B_{t_0}(\omega) \right| + \left| B_{t_0}(\omega) - B_{\frac{k+1}{2^n}}(\omega) \right| \leq \frac{5M}{2^n} \Rightarrow \omega \in F_{n,k}^{(2)} \\ \left| B_{\frac{k+3}{2^n}}(\omega) - B_{\frac{k+2}{2^n}}(\omega) \right| &\stackrel{\Delta}{=} \left| B_{\frac{k+3}{2^n}}(\omega) - B_{t_0}(\omega) \right| + \left| B_{t_0}(\omega) - B_{\frac{k+2}{2^n}}(\omega) \right| \leq \frac{7M}{2^n} \Rightarrow \omega \in F_{n,k}^{(3)} \end{aligned}$$

Hence, $\mathbb{P}(A) \leq \sum_{k=1}^{2^n} \mathbb{P}(F_{n,k})$, $\forall n \geq 3$. Next, $F_{n,k}^{(1)}, F_{n,k}^{(2)}, F_{n,k}^{(3)}$ are independent since the increments are independent. Thus,

$$\mathbb{P}(A_M) \leq \sum_{k=1}^{2^n} \mathbb{P}(F_{n,k}^{(1)}) \cdot \mathbb{P}(F_{n,k}^{(2)}) \cdot \mathbb{P}(F_{n,k}^{(3)}), \quad \forall n \geq 3$$

and we can estimate the 3 probabilities separately:

$$\begin{aligned} \mathbb{P}(F_{n,k}^{(3)}) &= \mathbb{P}\left(\left| B_{\frac{k+3}{2^n}} - B_{\frac{k+2}{2^n}} \right| \leq \frac{7M}{2^n}\right) \\ &= \mathbb{P}\left(|N| \leq \frac{7M}{2^n}\right) \text{ with } N \sim \mathcal{N}\left(0, \frac{1}{2^n}\right) \\ &= \mathbb{P}\left(|Y| \leq \frac{7M}{2^{\frac{n}{2}}}\right) \text{ with } Y \sim \mathcal{N}\left(0, \frac{1}{\sqrt{2^n}}\right) \\ &= \int_{-\frac{7M}{2^{\frac{n}{2}}} \underbrace{\sqrt{2\pi}}_{\leq 1}}^{\frac{7M}{2^{\frac{n}{2}}} \underbrace{\sqrt{2\pi}}_{\leq 1}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &\leq \underbrace{\frac{1}{\sqrt{2\pi}}}_{\leq 1} \cdot \frac{7M}{2^{\frac{n}{2}}} \leq \frac{7M}{2^{\frac{n}{2}}} \end{aligned}$$

In the same way we can estimate the other probabilities: $\mathbb{P}(F_{n,k}^{(2)}) \leq \frac{7M}{2^{\frac{n}{2}}}$ and $\mathbb{P}(F_{n,k}^{(1)}) \leq \frac{7M}{2^{\frac{n}{2}}}$. Thus,

$$\begin{aligned}\mathbb{P}(A_M) &\leq \sum_{k=1}^{2^n} \mathbb{P}F_{n,k} = \sum_{k=1}^{2^n} \mathbb{P}(F_{n,k}^{(1)}) \cdot \mathbb{P}(F_{n,k}^{(2)}) \cdot \mathbb{P}(F_{n,k}^{(3)}) \\ &\leq \sum_{k=1}^{2^n} \frac{(7M)^3}{2^{3\frac{n}{2}}} \\ &\stackrel{\text{ind. of } k}{\leq} 2^n \frac{(7M)^3}{2^{3\frac{n}{2}}} = \frac{(7M)^3}{2^{\frac{n}{2}}}\end{aligned}$$

Since this holds for all $n \in \mathbb{N}$ we find that $\mathbb{P}(A_M) = 0$. Thus, $\mathbb{P}(A) \leq \sum_{M=1}^{\infty} \mathbb{P}(A_M) = 0$. Hence, for all $\omega \in A^C$, $\omega \mapsto B_t(\omega)$ is nowhere differentiable in $[0, 1]$. Using the Markov property the same statement holds for all $[n, n+1]$ and intersecting all these events of probability 1 gives the claim. \square

4.3 Weak convergence in $C([0, \infty))$

We are now approaching Donkser, the convergence of scaled random walks towards the Brownian motion. The right notion will be weak convergence of the random walk laws towards the Wiener measure on $C([0, \infty))$. Hence, the first step is to discuss weak convergence on the Polish space $C([0, \infty))$.

Recall:

If P, P_1, P_2, \dots are probability measure on $\mathcal{B}(C([0, \infty)))$ we know

$$P_n \xrightarrow{(w)} P, n \rightarrow \infty \Leftrightarrow \text{tightness} + \text{convergence of enough integrals}$$

Tightness of measures on $C([0, \infty))$

Recall the definition of tightness for a family $(P_n)_{n \in \mathbb{N}}$ on $M_1(E)$ for some Polish space E :

$$\forall \varepsilon > 0 \exists K \subseteq E \text{ compact: } P_n(K^C) < \varepsilon \quad \forall n \in \mathbb{N}$$

Now the only difficulty of this section: your intuition of compact sets fails!!! So far the intuition came from \mathbb{R}^d : measures are tight if mass is not lost towards infinity, there is a large box where most of the mass of all P_i is concentrated. Probably you are mislead to believe in $C([0, \infty))$ we need to bound functions (or you have no intuition at all). This intuition is wrong anyways. A compact set is a set which you cannot „leave with a sequence“, formally, every sequence has a conv. subsequence with limit in the compact set. A is called relatively compact if \bar{A} is compact or (we are in a metric space) all sequences have a converging subsequence with limit in \bar{A} .

Beispiel. (i) $A := \{f\}$ is compact

(ii) $A := \{\text{all constant functions}\} \subseteq$

(iii) (Bild) $A := \{f_n : n \in \mathbb{N}\}$ not compact in $C([0, \infty)) \Leftarrow$ unif. conv. on compacts. But A is relatively compact.

$$(iv) f_n(x) = \begin{cases} x^{\frac{1}{n}}, & x \leq 1 \\ 2-x, & x \in (1, 2) \\ 0, & x \geq 1 \end{cases} \quad A := \{f_n : n \in \mathbb{N}\} \text{ is not compact. (Bild!)}$$

(v) Bild!! not compact

What goes wrong in (iii), (iv)? We may loose „regularity“. In this example the „strength if continuity“ - here in the sence of Hoelder continuity (f_n is $\frac{1}{n}$ -Hoelder) is not „uniformly good“.

The next theorem from functional analysis addresses exactly this point: If a set of functions is uniformly equicontinuous (this is the good notion of „same strength of continuity“) and are in some sense bounded, then subsequences with continuous limit exist. Arzéla-Ascoli characterizes relatively compact sets in $C([0, \infty))$. Keep in mind: You could also formulate A is closed and (i)+(ii) from Arzéla-Ascoli hold.?????????


Theorem 4.3.1. (Arzéla-Ascoli)

$A \subseteq C([0, \infty))$ is relatively compact if and only if

- (i) $\{f(0) : f \in A\} \subseteq \mathbb{R}$ is bounded
- (ii) $\lim_{\delta \rightarrow 0} \sup_{f \in A} V_f^N(\delta) = 0 \quad \forall N \in \mathbb{N}$, where

$$V_f^N(\delta) = \sup_{|t-s| < \delta, t, s \leq N} |f(t) - f(s)|$$

$\delta \mapsto V_f^N(\delta)$ is called modulus of continuity of f , maximal deviation of increments of length $\leq \delta$.

Bemerkung. (i) The same holds in $C([0, 1])$, we just need to remove N .

(ii) A is uniformly equicontinuous on $[0, N]$:

$$\begin{aligned} & \forall \varepsilon > 0 \exists \delta > 0 \forall f \in A: |t - s| < \delta \Rightarrow |f(t) - f(s)| < \varepsilon \\ & \Leftrightarrow \forall \varepsilon > 0 \exists \delta > 0 \sup_{f \in A} \sup_{|t-s| < \delta} |f(t) - f(s)| < \varepsilon \\ & \Leftrightarrow \lim_{\delta \rightarrow 0} \sup_{f \in A} V_f^N(\delta) = 0 \end{aligned}$$

Beispiel. $A_M = \{f : [0, \infty) \rightarrow \mathbb{R} \mid f|_{[0, N]} \text{ } \gamma\text{-Hoelder with constant } K \forall N\}$ is relatively compact. Why?

$$V_f^N(\delta) = \sup_{|t-s| < \delta, t, s \leq N} |f(t) - f(s)| \leq K \cdot \sup_{|t-s| < \delta, t, s \leq N} |t - s|^\gamma = K \cdot \delta^\gamma$$

The RHS is independent of f and vanishes for $\delta \rightarrow 0$. Since $|f(t) - f(s)| = \lim_{n \rightarrow \infty} |f_n(t) - f_n(s)| \leq k|t - s|^\gamma$ and $|f(0)| = \lim_{n \rightarrow \infty} |f_n(0)| \leq K$, A_M is even compact. Hence, we know plenty of compact sets: Bild!!! In essence these are the compact sets used for tightness of laws of stochastic processes.

We now use the description of (relatively) compact sets to characterize tightness of measures on $C([0, \infty))$.


Theorem 4.3.2. A family $(P_i)_{i \in I}$ of probability measures on $C([0, \infty))$ is weakly relatively compact (tight) if and only if

- (i) $\forall \varepsilon > 0 \exists K > 0: P_i(\{f : |f(0)| > K\}) \leq \varepsilon \quad \forall i \in I$
- (ii) $\forall \varepsilon, \eta > 0, N \in \mathbb{N} \exists \delta > 0: \sup_{i \in I} P_i(\{f : V_f^N(\delta) > \eta\}) \leq \varepsilon$

This is somewhat similar to saying a family of probability measures is tight iff $\forall \varepsilon > 0 \exists M > 0: \sup_{i \in I} P_i([-M, M]^C) \leq \varepsilon$, small mass away from the center. Here far away from the center means rough in the sense of modulus of continuity.

Beweis. „ \Rightarrow “ We skip this direction, not needed for this course as we only need conditions to ensure tightness.

„ \Leftarrow “ Suppose (i) and (ii) hold. For $\varepsilon > 0, k, N \in \mathbb{N}$ choose K_ε and $\delta_{N, K_\varepsilon}$ with

$$\sup_{i \in I} P_i(\{f : |f(0)| > K_\varepsilon\}) \leq \frac{\varepsilon}{2}$$

and

$$\sup_{i \in I} P_i \left(\left\{ f : V_f^N(\delta_{N,K,\varepsilon}) > \frac{1}{k} \right\} \right) \leq \frac{\varepsilon}{2^{K+N+1}}.$$

Define

$$C_{N,\varepsilon} := \left\{ f : V_f^N(\delta_{N,K,\varepsilon}) \leq \frac{1}{k} \forall k \in \mathbb{N} \right\}, \quad D = \{f : |f(0)| \leq K_\varepsilon\}$$

By A-A, $C_\varepsilon := D \cap \bigcap_{N=1}^\infty C_{N,\varepsilon}$ is relatively compact. Why? If $f \in C_\varepsilon$, then $V_f^N(\delta_{N,K,\varepsilon}) \leq \frac{1}{K}$ $\forall N, K$ so that

$$\lim_{\delta \rightarrow 0} \sup_{f \in C_\varepsilon} V_f^N(\delta_{N,K,\varepsilon}) = 0 \text{ as } \delta_{N,K,\varepsilon} \rightarrow 0 \Leftrightarrow K \rightarrow 0$$

We now prove that \overline{C}_ε fulfills the definition that makes $(P_i)_{i \in I}$ tight in $C([0, \infty))$:

$$\begin{aligned} P_i(C_\varepsilon^C) &\leq P_i \left(D^C \cup \bigcup_{N=1}^\infty C_{N,\varepsilon}^C \right) \\ &= P_i \left(\{f : |f(0)| > K_\varepsilon\} \cup \bigcup_{N=1}^\infty \left\{ f : V_f^N(\delta_{N,K,\varepsilon}) > \frac{1}{K} \text{ for some } K \right\} \right) \\ &\leq P_i(\{f : |f(0)| > K_\varepsilon\}) + \sum_{N=1}^\infty \sum_{K=1}^\infty P_i \left(\left\{ f : V_f^N(\delta_{N,K,\varepsilon}) > \frac{1}{K} \right\} \right) \\ &\leq \frac{\varepsilon}{2} + \sum_{N=1}^\infty \sum_{K=1}^\infty \frac{\varepsilon}{2^{K+N+1}} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

If C_ε is relatively compact, then \overline{C}_ε is compact and we have $P_i(\overline{C}_\varepsilon) \stackrel{\text{monot.}}{\geq} P_i(C_\varepsilon) \geq 1 - \varepsilon \forall i \in I$. Hence, $(P_i)_{i \in I}$ is tight. \square

Weak convergence of stochastic processes in $C([0, \infty))$:

Definition. Suppose X, X^1, X^2, \dots are stochastic processes with a.s. continuous paths and laws $\mathbb{P}_X, \mathbb{P}_{X^1}, \dots$ on $C([0, \infty))$.

- (i) We say $(X^n)_{n \in \mathbb{N}}$ converges weakly to X if $\mathbb{P}_{X^n} \xrightarrow{(w)} \mathbb{P}_X$, $n \rightarrow \infty$. Usually one writes rather $X^n \Rightarrow X$, $n \rightarrow \infty$, but also $X^n \xrightarrow{(w)} X$, $n \rightarrow \infty$, or $X^n \xrightarrow{(d)} X$, $n \rightarrow \infty$.
- (ii) We say the finite dimensional distributions (fdd) of $(X^n)_{n \in \mathbb{N}}$ converge to those of X if for all $0 \leq t_1 < \dots < t_k$, $k \in \mathbb{N}$

$$(X_{t_1}^n, \dots, X_{t_k}^n) \xrightarrow{(w)} (X_{t_1}, \dots, X_{t_k}), \quad n \rightarrow \infty$$

In that case we write $X^n \xrightarrow{\text{fdd}} X$, $n \rightarrow \infty$.

Bemerkung.

$$X^n \Rightarrow X, \quad n \rightarrow \infty \Rightarrow X^n \xrightarrow{\text{fdd}} X, \quad n \rightarrow \infty$$

Why? The projections $\pi_{t_1, \dots, t_k}(f) := (f(t_1), \dots, f(t_k))$ are continuous mappings from $C([0, \infty))$ to \mathbb{R}^k . Then

$$(X_{t_1}^n, \dots, X_{t_k}^n) = \pi_{t_1, \dots, t_k}(X^n) \xrightarrow{(w)} \pi_{t_1, \dots, t_k}(X) = (X_{t_1}, \dots, X_{t_k})$$



Theorem 4.3.3. Suppose X, X^1, \dots are stochastic processes with continuous paths. Then

$$X^n \Rightarrow X, n \rightarrow \infty \Leftrightarrow \begin{array}{l} \text{(i) } (\mathbb{P}_{X^n})_{n \in \mathbb{N}} \text{ is tight} \\ \text{(ii) } X^n \xrightarrow{\text{fdd}} X, n \rightarrow \infty \end{array}$$

Beweis. „ \Rightarrow “ (i) is Prohorov for $E = C([0, \infty))$, (ii) continuous mapping as in 4.3.

„ \Leftarrow “ By (i) and Prohorov $(X^n)_{n \in \mathbb{N}}$ is relatively sequential compact. Hence, there is a subsequence $(X^{n_k})_{k \in \mathbb{N}}$ that converges weakly to some process Y . By „ \Rightarrow “ the fdd of $(X^{n_k})_{k \in \mathbb{N}}$ converge to the ones of Y . Since the law of a stochastic process is uniquely determined by fdd we have $\mathbb{P}_X = \mathbb{P}_Y$. Hence, $X^{n_k} \Rightarrow X, k \rightarrow \infty$. Now we argue exactly as in the proof of 3.3.4. \square

Before we turn to Donsker we need a more practical tool to check the tightness for a sequence of continuous stochastic processes:



Theorem 4.3.4. (Kolmogorov tightness criterion)

Suppose $(X^n)_{n \in \mathbb{N}}$ is a sequence of continuous stochastic processes on $(\Omega, \mathcal{A}, \mathbb{P})$ such that

- (i) $(X_0^n)_{n \in \mathbb{N}}$ is tight
- (ii) There are $\alpha, \beta, T > 0$ such that, $\forall n \in \mathbb{N}$,

$$\mathbb{E}[|X_t^n - X_s^n|^\alpha] \leq c \cdot |t - s|^{1+\beta}, \quad \forall t, s \leq T$$

Then $(X^n)_{n \in \mathbb{N}}$ is tight in $C([0, \infty))$.

Beweis. Fix $\gamma \in (0, \frac{\beta}{\alpha})$ and set $K = \frac{2}{1-2-\gamma}$. We will construct a set $C \in \mathcal{A}$ of measure 1 such that $\omega \mapsto X_t^n(\omega)$ is (locally) $K - \gamma$ -Hoelder continuous, that is, for all $t \geq 0$, $|X_t^n(\omega) - X_s^n(\omega)| \leq K \cdot |t - s|^\gamma$ in a neighborhood around t . Note: constructing such a C with $\mathbb{P}(C) = 1$ is more than needed (compare 4.3.2 (ii)): As $V_N(\delta) \leq K \cdot \delta^\gamma$ for $K - \gamma$ -Hoelder functions, we could even choose $\varepsilon = 0$. The set C will be constructed using only Borel-Cantelli. As several times before we only consider times in $[n, n+1]$ and intersecting countably many measurable sets of measure 1 gives again a measurable set of measure 1.

1. Estimating increments of X^i on neighboring dyadics:

$$\mathbb{P}(|X_t^i - X_s^i| > \varepsilon) \stackrel{\text{Markov}}{\leq} \frac{\mathbb{E}[|X_t^i - X_s^i|^\alpha]}{\varepsilon^\alpha} \stackrel{\text{ass.}}{\leq} c \cdot \frac{|t - s|^{1+\beta}}{\varepsilon^\alpha}$$

so that, in particular,

$$\mathbb{P}\left(|X_{\frac{k}{2^n}}^i - X_{\frac{k-1}{2^n}}^i| \geq 2^{-\gamma \cdot n}\right) \leq c \cdot 2^{\alpha \gamma n} \cdot 2^{-n(1+\beta)} = c \cdot 2^{-n(1+\beta-\alpha\gamma)}$$

which gives

$$\begin{aligned} \mathbb{P}\left(\underbrace{\max_{k \leq 2^n} |X_{\frac{k}{2^n}}^i - X_{\frac{k-1}{2^n}}^i|}_{A_n} \geq 2^{-\gamma \cdot n}\right) &\leq \mathbb{P}\left(\bigcup_{k \leq 2^n} \left\{ |X_{\frac{k}{2^n}}^i - X_{\frac{k-1}{2^n}}^i| \geq 2^{-\gamma \cdot n} \right\}\right) \\ &\stackrel{\text{sub. add.}}{\leq} \sum_{k=1}^{2^n} \mathbb{P}\left(|X_{\frac{k}{2^n}}^i - X_{\frac{k-1}{2^n}}^i| \geq 2^{-\gamma \cdot n}\right) \\ &\leq c \cdot \sum_{k=1}^{2^n} 2^{-n(1+\beta-\alpha \cdot \gamma)} \\ &= c \cdot 2^{-n(\beta-\alpha \cdot \gamma)} \end{aligned}$$

By Borel-Cantelli there is $A = \{A_n \text{ i.o.}\}^C \in \mathcal{A}$ with $\mathbb{P}(A) = 1$ so that for all $\omega \in A$ there are $n_0(\omega)$ so that

$$\max_{k \leq 2^n} |X_{\frac{k}{2^n}}^i - X_{\frac{k-1}{2^n}}^i| < \frac{1}{2^{\gamma \cdot n}}, \quad \forall n \geq n_0(\omega). \quad (4.10)$$

2. Estimating increments of X on all dyadics: Fix $\omega \in \mathcal{A}$ so that we can use 4.10 for all $m \geq N \geq n_0(\omega)$. We now estimate increments $|X_t(\omega) - X_s(\omega)|$ on the dyadic numbers. (Bild!!)

$\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \dots$ dyadics $\mathcal{D}_n = \{0, \frac{1}{2^n}, \frac{2}{2^n}, \dots, 1\}$ and $\mathcal{D} := \bigcup_{n=1}^{\infty} \mathcal{D}_n$.

Claim: Suppose $m > N$ and $t, s \in \mathcal{D}_m$ are such that $|t - s| < \frac{1}{2^N}$. Then

$$|X_t^i(\omega) - X_s^i(\omega)| \leq 2 \cdot \sum_{k=N+1}^m \frac{1}{2^{\gamma \cdot k}}.$$

Why? Clever telescopic sum argument, by induction: (Bild!!!)

$m = N + 1$: For $m = N + 1$, t and s must be neighbors ($|t - s| = \frac{1}{2^m} < \frac{1}{2^N}$) in \mathcal{D}_m , the claim follows from 4.10 as $m \geq n_0(\omega)$.

$m > N + 1$: We suppose the claim holds for $m - 1$ and take $s, t \in \mathcal{D}_m$. We chose $t', s' \in \mathcal{D}_{m-1}$ as in the picture.

$$\begin{aligned} &|X_t^i(\omega) - X_s^i(\omega)| \\ &\stackrel{m \geq N \geq n_0(\omega)}{\leq} |X_t^i(\omega) - X_{t'}^i(\omega)| + |X_{t'}^i(\omega) - X_{s'}^i(\omega)| + |X_{s'}^i(\omega) - X_s^i(\omega)| \\ &\leq 2 \cdot \frac{1}{2^{\gamma(m-1)}} + 2 \cdot \sum_{k=N+1}^{m-1} \frac{1}{2^{\gamma \cdot k}} + 2 \cdot \frac{1}{2^{\gamma(m-1)}} \\ &= 2 \cdot \sum_{k=N+1}^m \frac{1}{2^{\gamma \cdot k}} \end{aligned}$$

3. Hölder on \mathcal{D} :

Now define $h(\omega) = 2^{-n_0(\omega)}$. If $t, s \in \mathcal{D}$ are such that $|t - s| < h(\omega)$, then there is some $N \geq n_0(\omega)$ with $\frac{1}{2^{N+1}} \leq |t - s| \leq \frac{1}{2^N}$ so that

$$|X_t^i(\omega) - X_s^i(\omega)| \stackrel{2.}{\leq} 2 \cdot \sum_{k=N+1}^{\infty} \frac{1}{2^{\gamma \cdot k}} = \frac{2}{2^{\gamma(N+1)}} \sum_{k=0}^{\infty} \frac{1}{2^{\gamma \cdot k}} \leq \underbrace{K}_{=\frac{2}{1-2^{-\gamma}}} \cdot |t - s|^{\gamma}$$

$$\Rightarrow |X_t^i(\omega) - X_s^i(\omega)| \leq K \cdot |t - s|^{\gamma} \text{ in } (t - h(\omega), t + h(\omega)).$$

4. Hoelder on $[0, 1]$:

Since X is a.s. continuous there is some $B \in \mathcal{A}$ with $\mathbb{P}(B) = 1$ so that $t \mapsto X_t(\omega)$ is continuous. Now define $C := A \cap B \in \mathcal{A}$. Then $\mathbb{P}(C) = 1$ and $|X_t^i(\omega) - X_s^i(\omega)| \leq K \cdot |t - s|^\gamma$ locally at all t . Hence, all X^i have almost sure K - γ -Hoelder continuous paths.

5. Tightness:

$$(i) \quad \mathbb{P}_{X^i}(\{f: |f(0)| > K\}) = \mathbb{P}(|X_i| > K) \checkmark$$

(ii) For small δ we have:

$$\begin{aligned} \mathbb{P}_{X^i}(\{f: V_j^N(\delta) > \eta\}) &= \mathbb{P}\left(\sup_{|t-s|<\delta} |X_t^i - X_s^i| > \eta\right) \\ &\leq \mathbb{P}(K\delta^\gamma > \eta) = 0 \end{aligned}$$

\Rightarrow Theorem 4.3.2 implies tightness. □

Bemerkung. The argument is usually differently, to construct continuous modifications of processes for which the continuity is not known. Most importantly, if processes are constructed using the canonical construction on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}(\mathbb{R})^{[0,\infty)})$ using Kolmogorov \rightarrow Kolmogorov/Chentsov, spatial statistics lecture

Bemerkung. If you look at the proof you see we proved that a continuous process X satisfying $\mathbb{E}[|X_t - X_s|^\alpha] \leq c \cdot |t - s|^{1+\beta}$ has almost surely $\frac{\beta}{\alpha}$ -Hoelder continuous sample paths. Note: $|t - s|^a \leq c_T \cdot |t - s|^b \forall t, s \leq T$ for $a \leq b$. That's why it is harder to prove the assumption for larger β but you also get better (i.e. larger γ) Hoelder continuity.

Beispiel. • $\mathbb{E}[|B_t - B_s|^2] = |t - s|$ gives nothing

- $\mathbb{E}[|B_t - B_s|^4] \leq 3|t - s|^2$ gives $\frac{1}{4}$ -Hoelder
- $\mathbb{E}[|B_t - B_s|^{2m}] \leq c_m \cdot |t - s|^m$ gives $\frac{m-1}{2m}$ -Hoelder $\forall m \in \mathbb{N}$
which implies that Brownian paths are $(\frac{1}{2} - \varepsilon)$ -Hoelder.

4.4 Donsker's invariance principle

We collected all tools to prove the so-called functional central limit theorem, the random function (stochastic process) version of the CLT. Recall the definition of the random walk measure and the Wiener measure:

1. For ξ_1, ξ_2, \dots are iid on $(\Omega, \mathcal{A}, \mathbb{P})$, $S_n := \sum_{k=1}^n \xi_k$ is called random walk: Bild!!!! Interpolated linearly we obtain a stochastic process on $(\Omega, \mathcal{A}, \mathbb{P})$ with continuous paths. Bild linear. Note: $t \mapsto X_t(\omega)$ is even continuous for all $\omega \in \Omega$ by construction. The induced probability law on $C([0, \infty))$ is called random walk measure.
2. The wiener measure is the law of a Brownian motion on $C([0, \infty))$.

To turn random walks into a Brownian motion we need to increase the jump frequency. Let's chose to have n jumps in all time periods $[0, 1], [1, 2], \dots$. Hence, at time 1 we get $X_1^n = \sum_{k=1}^n \xi_k$. If $(X^n)_{n \in \mathbb{N}}$ should converge to a Brownian motion weakly in $C([0, \infty))$ we already know that finite dimension marginals need to converge weakly. At least we need $X_1^n \rightarrow B_1 \sim \mathcal{N}(0, 1)$. Of course this does not work, we know the CLT! What we need to do is subtract the expectation and divide by $\sigma\sqrt{n}$. On top, we need finite second moments.

Definition. Let ξ_1, ξ_2, \dots iid then we call

$$\tilde{X}_t^n := \frac{\sum_{k=1}^{\lfloor nt \rfloor} (\xi_k - \mathbb{E}[\xi_1])}{\sigma \sqrt{n}}$$

the scaled random walk and

$$X_t^n := \tilde{X}_t^n + (nt - \lfloor nt \rfloor) \frac{1}{\sigma \sqrt{n}} \xi_{\lfloor nt \rfloor + 1}, \quad t \geq 0,$$

the linearly interpolated scaled random walk. Here we use $\lfloor x \rfloor = \sup\{n \in \mathbb{N} : n \leq x\}$. zwei mal Bilder!!!!!!



Theorem 4.4.1. (Donsker functional CLT)

Let X^n the scaled and linearly interpolated random walk for a jump distribution with finite variance $\sigma^2 = \mathbb{V}(\xi_1)$. Then $X^n \xrightarrow{\text{fdd}} B$, $n \rightarrow \infty$, where B is a Brownian motion.

Beweis. Idea: $X^n \xrightarrow{\text{fdd}} B$ and tightness in $C([0, \infty))$. Wlog $\mathbb{E}[\xi_1] = 0$ and $\sigma^2 = 1$.

1. $X^n \xrightarrow{\text{fdd}} B$:

We first show $\tilde{X}^n \xrightarrow{\text{fdd}} B$. For $s < t$ we get g

$$\tilde{X}_t^n - \tilde{X}_s^n = \frac{\sum_{k=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} \xi_k}{\sqrt{n}} \xrightarrow{(d)} \frac{\sum_{k=1}^{\lfloor n(t-s) \rfloor} \xi_k}{\sqrt{n}} \xrightarrow{\text{CLT}} \mathcal{N}(0, t-s)$$

Since, for $t_0 < \dots < t_k$, the sums $\tilde{X}_{t_1}^n - \tilde{X}_{t_0}^n, \dots, \tilde{X}_{t_k}^n - \tilde{X}_{t_{k-1}}^n$ are independent, we get, for a BM B ,

$$(\tilde{X}_{t_1}^n - \tilde{X}_{t_0}^n, \dots, \tilde{X}_{t_k}^n - \tilde{X}_{t_{k-1}}^n) \xrightarrow{(d)} (B_{t_1} - B_{t_0}, \dots, B_{t_k} - B_{t_{k-1}})$$

and then continuous mapping theorem, for $f(x) = A \cdot x$,

$$(\tilde{X}_{t_0}^n, \dots, \tilde{X}_{t_k}^n) \xrightarrow{(d)} (B_{t_0}, \dots, B_{t_k})$$

This shows $\tilde{X}^n \xrightarrow{\text{fdd}} B$. Recall Slutsky for random vectors from Stochastik 2:

$$X_n \xrightarrow{(d)} X, \quad n \rightarrow \infty, \quad Y_n \xrightarrow{P} c, \quad n \rightarrow \infty \Rightarrow X_n + Y_n \xrightarrow{(d)} X + c, \quad n \rightarrow \infty$$

Hence, we would like to estimate

$$\begin{aligned} \mathbb{P}(|\tilde{X}_t^n - X_t^n| > \varepsilon) &\leq \frac{\mathbb{E}[|\tilde{X}_t^n - X_t^n|^2]}{\varepsilon^2} \\ &\leq (nt - \lfloor nt \rfloor)^2 \frac{1}{\varepsilon^2 \cdot n} \mathbb{E}[|\xi_{\lfloor nt \rfloor + 1}|^2] \\ &\leq \frac{\mathbb{E}[\xi_1^2]}{\varepsilon^2 \cdot n} \rightarrow 0, \quad n \rightarrow \infty \end{aligned}$$

Writing $X^n = \underbrace{\tilde{X}_t^n}_{\xrightarrow{\text{fdd}} B} + \underbrace{(X^n - \tilde{X}_t^n)}_{\xrightarrow{P} 0}$ we find $X^n \xrightarrow{\text{fdd}} B$ as $n \rightarrow \infty$.

2. Tightness: We assume $\mathbb{E}[\xi_1^4] < \infty$

We use the Kolmogorov tightness criterion, ugly only because of the interpolation

- (i) The constant sequence $(X_0^n)_{n \in \mathbb{N}}$ is tight.
- (ii) We need to estimate $\mathbb{E}[|X_t^n - X_s^n|^\alpha] \leq c|t-s|^{1+\beta}$ for some $\alpha > 0$. We use $\alpha = 4$.

Differences are most easily understood on $\{\frac{k}{n} : n \in \mathbb{N}\}$:

$$X_{\frac{k}{n}}^n - X_{\frac{l}{n}}^n = \sum_{i=1}^k \frac{\xi_i}{\sqrt{n}} - \sum_{i=1}^l \frac{\xi_i}{\sqrt{n}} = \sum_{j=l+1}^k \frac{\xi_j}{\sqrt{n}}$$

and we know how to estimate $\mathbb{E}[|X_t - X_s|^4]$ as in the proof of LNN with 4th moments. For general t, s everything just looks ugly due to the linear interpolation:

$$X_t^n - X_s^n = (nt - \lfloor nt \rfloor) \frac{\xi_{\lfloor nt \rfloor + 1}}{\sqrt{n}} + \sum_{j=\lfloor ns \rfloor + 2}^{\lfloor nt \rfloor} \frac{\xi_j}{\sqrt{n}} - (ns - \lfloor ns \rfloor) \frac{\xi_{\lfloor ns \rfloor + 1}}{\sqrt{n}} \quad (4.11)$$

3 mal Bilder:

- (1a) $|X_t^n - X_s^n|^4 = |t-s|^4 \cdot |\text{slope}|^4 = |t-s|^4 \cdot n^4 \cdot \frac{\xi_{k+1}^4}{n^2} \leq |t-s|^2 \xi_{k+1}^4$
(1b)

$$\begin{aligned} |X_t^n - X_s^n| &\leq |t-s| \cdot |\text{slope}| + |t-s| \cdot |\text{slope}| \\ &= |t-s| \cdot n \cdot \frac{|\xi_{k+1}|}{\sqrt{n}} + |t-s| \cdot n \cdot \frac{|\xi_{k+2}|}{\sqrt{n}} \\ &= |t-s| \sqrt{n} (|\xi_{k+1}| + |\xi_{k+2}|) \end{aligned}$$

$$\text{so that } |X_t^n - X_s^n|^4 \underbrace{\leq}_{|t-s| \leq \frac{1}{n}} |t-s|^2 (|\xi_{k+1}| + |\xi_{k+2}|)^4.$$

$$\stackrel{(a),(a')}{\Rightarrow} \mathbb{E}[|X_t^n - X_s^n|^4] \leq C_1 \cdot |t-s|^2 \text{ if } |t-s| \leq \frac{1}{n}.$$

- (2) We now estimate $\mathbb{E}[|X_t^n - X_s^n|^4]$ if $|t-s| > \frac{1}{n}$. As in the proof of the LLN in Stochastik 1 recall that

$$\begin{aligned} \mathbb{E} \left[\left(\sum_{j=1}^N Y_j \right)^4 \right] &= \mathbb{E} \left[\sum_{i,j,l,n=1}^N Y_i \cdot Y_j \cdot Y_l \cdot Y_n \right] \\ &\stackrel{Y_i \text{iid and } \mathbb{E}[Y_i]=0}{=} N \cdot \mathbb{E}[Y_1^4] + \frac{N(N-1)}{2} \mathbb{E}[Y_1^2] \end{aligned} \quad (4.12)$$

Also note that, for $a \in [-1, 1]$, for X, Y centered and independent

$$\begin{aligned} \mathbb{E}[(aX + Y)^4] &= a^4 \mathbb{E}[X^4] + 6a^2 \mathbb{E}[X^2] \mathbb{E}[Y^2] + \mathbb{E}[Y^4] \\ &\leq \mathbb{E}[X^4] + 6\mathbb{E}[X^2] \mathbb{E}[Y^2] + \mathbb{E}[Y^4] = \mathbb{E}[(X + Y)^4] \end{aligned}$$

We use both facts for 4.11, the second facts twice to remove $nt - \lfloor nt \rfloor$ and $ns - \lfloor ns \rfloor$ to obtain:

$$\begin{aligned} \mathbb{E}[|X_t - X_s|^4] &\leq \frac{1}{n^2} \mathbb{E} \left[\left(\sum_{j=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor + 1} \xi_j \right)^4 \right] \\ &\stackrel{4.12}{\leq} \frac{(\lfloor nt \rfloor - \lfloor ns \rfloor) \cdot \mathbb{E}[\xi_1^4] + (\lfloor nt \rfloor - \lfloor ns \rfloor)^2 \cdot 1}{n^2} \\ &\leq \frac{2|t-s|n\mathbb{E}[\xi_1^4] + 4|t-s|^2 \cdot n^2 \cdot \sigma^2}{n^2} \\ &\leq (2\mathbb{E}[\xi_1] + 4) \cdot |t-s|^2 \\ &:= C_2 \cdot |t-s|^2 \end{aligned}$$

Now we can combine (1) and (2) to obtain

$$\mathbb{E} [|X_t - X_s|^4] \leq C \cdot |t - s|^2$$

for all $t, s \geq 0$ with $C := \max\{C_1, C_2\}$.

□

If $\sigma^2 = \infty$, then the CLT and Donsker are wrong! Replacing the scaling \sqrt{n} by $n^{\frac{1}{\alpha}}$ for the right choice of α leads to stable distributions and α -stable Lévy processes (see simulation). Those processes have discontinuous paths! Magic: The scaling limit depends on α , the max „number“ of finite moments. α small means the limit is ugly. As soon as $\alpha \geq 2$, the limit is always the same, the underlying law is only reflected in σ . This magic is called the universality.

Bemerkung. Donsker is the first step into the world of modern probability, where „universal objects“ are studied that are scaling limits of large classes of discrete random objects:

- random trees: (Bild)
- growing surfaces → ICPZ equation (Bild)

Bemerkung. We proved $X^n \Rightarrow B$, $n \rightarrow \infty$ in $C([0, \infty))$. You will sometimes see for other sequences of processes that people only prove $X^n \xrightarrow{\text{fdd}} X$. This has multiple reasons: (a) sometimes only this weaker statement holds, (b) sometimes people failed to prove the stronger statement $X^n \Rightarrow X$. Why do we prefer the weak convergence? Because then automatically we get $F(X^n) \xrightarrow{(d)} F(X)$ for all $F: C([0, \infty)) \rightarrow \mathbb{R}$ continuous. Example:

$$F(f) := \sup_{t \leq 1} |f(t)|$$

is continuous. Hence,

$$\begin{aligned} \frac{1}{\sigma \sqrt{n}} \sup_{k \leq 1} \left| \sum_{l=1}^k \xi_l \right| &= F(X^n) \\ &\xrightarrow{(d)} F(B) \\ &= \sup_{t \leq 1} |B_t| \\ &\sim |X| \end{aligned}$$

with $X \sim \mathcal{N}(0, 1)$ which is called the „reflection principle“. Supremum of any random walk with finite second moments roughly behaves like $\sigma \sqrt{n}$!