

Jan-Cornelius Molnar

Wahrscheinlichkeitstheorie

Aufbauend auf einer Vorlesung von PD Dr. J. Dippon

Stuttgart, Wintersemester 2009 / 2010

Version: 6. November 2011 13:28

Für Hinweise auf Druckfehler und Kommentare jeder Art bin ich dankbar. Viel Spaß!¹

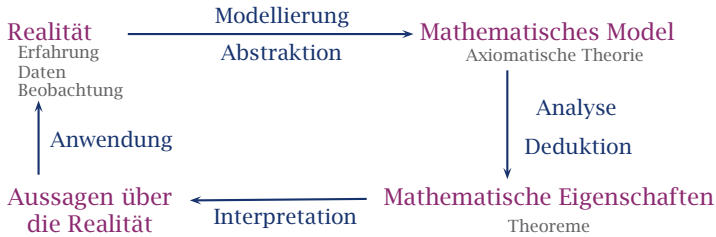
¹Jan-Cornelius Molnar jan.molnar@studentpartners.de

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 0 | Gegenstand der Vorlesung | 5 |
| 1 | Wahrscheinlichkeitsräume | 7 |
| 1-A | Algebren, Inhalte und Maße | 7 |
| 1-B | Wahrscheinlichkeitsräume und -maße | 13 |
| ■ | Binomialverteilung | 14 |
| ■ | Poissonverteilung | 17 |
| ■ | Weitere Eigenschaften von Wahrscheinlichkeitsmaßen | 18 |
| ■ | Fortsetzung von Wahrscheinlichkeitsmaßen | 22 |
| 1-C | Verteilungsfunktionen | 24 |
| 1-D | Bedingte Wahrscheinlichkeiten | 29 |
| 2 | Kombinatorische Wahrscheinlichkeitsrechnung | 35 |
| 3 | Zufallsvariablen und Verteilungen | 39 |
| 3-A | Meßbare Abbildungen und Zufallsvariablen | 39 |
| 3-B | Bildmaße und Verteilungen | 47 |
| ■ | Produktmessräume | 50 |
| 4 | Erwartungswerte und Dichten | 53 |
| 4-A | Erwartungswert, Varianz und Lebesgue Integral | 53 |
| ■ | Erwartungswert mittels Riemann-Stieltjes-Integral | 54 |
| ■ | Erwartungswert mittels Maß-Integral | 55 |
| ■ | Eigenschaften des Erwartungswerts | 59 |
| 4-B | Dichtefunktion und Zähldichte | 63 |
| ■ | Der Transformationsatz | 67 |
| 4-C | Momente von Zufallsvariablen | 71 |
| 5 | Unabhängigkeit | 75 |
| 5-A | Unabhängige Ereignisse und Zufallsvariablen | 76 |
| ■ | Unabhängige Zufallsvariablen | 78 |

| | | |
|-----------|--|------------|
| ■ | Faltungen | 82 |
| 5-B | Null-Eins-Gesetz | 84 |
| 6 | Erzeugende und charakteristische Funktionen, Faltungen | 89 |
| 6-A | Erzeugende Funktionen | 89 |
| ■ | Konvergenzsätze der Maßtheorie | 95 |
| 6-B | Charakteristische Funktionen | 97 |
| ■ | Eindeutigkeitssatz und Umkehrformeln | 103 |
| 6-C | Faltungen | 108 |
| 7 | Spezielle Verteilungen | 113 |
| 7-A | Diskrete Verteilungen | 113 |
| 7-B | Totalstetige Verteilungen | 117 |
| 8 | Gesetze der großen Zahlen | 123 |
| 8-A | Konvergenzbegriffe | 123 |
| 8-B | Schwache und starke Gesetze der großen Zahlen | 126 |
| 9 | Zentrale Grenzwertsätze | 135 |
| 9-A | Schwache Konvergenz von Wahrscheinlichkeitsmaßen in \mathbb{R} | 135 |
| 9-B | Zentrale Grenzwertsätze | 149 |
| ■ | Multivariate zentrale Grenzwertsätze | 160 |
| 10 | Bedingte Erwartungen | 163 |
| 10-A | Grundlagen | 164 |
| 11 | Martingale | 177 |
| A | Begriffe und Sätze der Maß- und Integrationstheorie | 191 |
| | Literaturverzeichnis | 193 |

o Gegenstand der Vorlesung



0.1 Schema Mathematische Modellierung.

Gegenstand der Vorlesung ist es nicht, Aussagen darüber zu treffen, was "Zufall" oder "Wahrscheinlichkeit" konkret bedeutet. Wir haben die Absicht zu klären, welche Axiome für eine Wahrscheinlichkeit gelten und welche Eigenschaften sie erfüllen muss - sofern sie existiert.

1 Wahrscheinlichkeitsräume

1-A Algebren, Inhalte und Maße

Ziel dieses Abschnittes ist es, Abbildungen μ zu definieren, die einer Teilmenge eines Raumes Ω ein “Maß” zuordnen,

$$\mu : \mathcal{P}(\Omega) \rightarrow \mathbb{R}.$$

Gehen wir von $\Omega = \mathbb{R}$ aus und betrachten zunächst nur Intervalle als Teilmengen, so ist es nur natürlich diesen auch ihre Länge als “Maß” zuzuordnen,

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

G. Vitali¹ hat jedoch bereits 1905 gezeigt, dass es nicht möglich ist, ein solches “Maß” dann für alle Teilmengen von \mathbb{R} , d.h. auf der Potenzmenge $\mathcal{P}(\mathbb{R})$ zu definieren, es ist also notwendig sich lediglich auf eine Teilmenge von $\mathcal{P}(\mathbb{R})$ zurückzuziehen, d.h. man kann auch nur bestimmten Teilmengen von \mathbb{R} mit einem solchen Maß “vermessen”.

¹Giuseppe Vitali (* 26. August 1875 in Ravenna; † 29. Februar 1932 in Bologna) war ein italienischer Mathematiker.

In der Wahrscheinlichkeitstheorie werden üblicherweise als Definitionsbereich von Maßen sogenannte σ -Algebren betrachtet.

1.1 **Definition** Sei Ω eine nichtleere Menge. Ein System \mathcal{A} von Teilmengen von Ω heißt *Algebra*, wenn

- 1.) $\emptyset \in \mathcal{A}$,
- 2.) $A \in \mathcal{A} \Rightarrow \Omega \setminus A \in \mathcal{A}$,
- 3.) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$
oder äquivalent
 $A_1, \dots, A_m \in \mathcal{A} \Rightarrow \bigcup_{i=1}^m A_i \in \mathcal{A}$.

das System \mathcal{A} heißt σ -Algebra, falls zusätzlich gilt,

- 3.*) $A_n \in \mathcal{A}$ für $n \in \mathbb{N} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. \times

Diese Definition einer Algebra ist nicht mit der aus der Vorlesung "Algebra" zu verwechseln. Dennoch existieren Parallelen. Vereinigung und Schnittbildung könnte man als Addition und Multiplikation betrachten, das Komplement als Inverses und die leere Menge als Identität.

1.1 *Bemerkung.* Ist \mathcal{A} σ -Algebra, so ist \mathcal{A} auch auch eine Algebra. \rightarrow

Wie wir noch sehen werden, besitzen Algebren zahlreiche angenehme Eigenschaften. Durch die Forderung dass Komplemente und Vereinigungen enthalten sind, ergibt sich beispielsweise automatisch, dass auch Schnitte und Differenzen enthalten sind.

1.1 **Lemma** a) Sei \mathcal{A} eine Algebra, so gilt

$$A_1, \dots, A_m \in \mathcal{A} \Rightarrow \bigcap_{n=1}^m A_n \in \mathcal{A}.$$

$$A, B \in \mathcal{A} \Rightarrow A \setminus B \in \mathcal{A}.$$

b) Sei \mathcal{A} eine σ -Algebra, so gilt außerdem

$$A_n \in \mathcal{A} \text{ für } n \in \mathbb{N} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}. \quad \times$$

Eine Algebra (bzw. σ -Algebra) enthält somit stets \emptyset und Ω und ist abgeschlossen gegenüber endlich (bzw. abzählbar) vielen üblichen Mengenoperationen.

» a) Definition 1.1 besagt, $A_1, \dots, A_m \in \mathcal{A} \Rightarrow A_1^c, \dots, A_m^c \in \mathcal{A}$. Mit den Regeln von De-Morgan folgt,

$$\left(\bigcap_{n=1}^m A_n \right)^c = \bigcup_{n=1}^m A_n^c \in \mathcal{A},$$

d.h. $\bigcap_{n=1}^m A_n \in \mathcal{A}$.

b) Seien $A, B \in \mathcal{A}$. Man sieht leicht, dass $A \setminus B = A \cap B^c$ und daher ist auch $A \setminus B \in \mathcal{A}$ als Schnitt von A und B^c . «

1.2 **Definition** Sei Ω eine nichtleere Menge und \mathcal{A} eine σ -Algebra in Ω . Das Paar (Ω, \mathcal{A}) heißt *Messraum (measurable space)*; die Elemente $\in \mathcal{A}$ heißen *\mathcal{A} -messbare Mengen*. Entsprechendes gilt für σ -Algebra. \times

Man startet bei der Definition von Abbildungen auf Messräumen oft nur auf Teilmengen - wie z.B. den Intervallen in \mathbb{R} - und dehnt anschließend den Definitionsbereich aus. Für diese Vorgehensweise sind folgende Ergebnisse zentral.

1.2 **Lemma** a) Seien \mathcal{A}_α Algebren in Ω , $\alpha \in \mathcal{I}$ und \mathcal{I} beliebige Indexmenge, so ist $\bigcap_{\alpha \in \mathcal{I}} \mathcal{A}_\alpha$ eine Algebra in Ω .

b) Sei $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ ein Mengensystem in Ω , so es existiert eine kleinste \mathcal{C} enthaltende Algebra. \times

» a) Wir führen den Beweis für σ -Algebren; der für Algebren wird analog geführt. \mathcal{A}_α sei für jedes $\alpha \in \mathcal{I}$ eine σ -Algebra in Ω . Also ist auch $\emptyset \in \bigcap_{\alpha \in \mathcal{I}} \mathcal{A}_\alpha$, da $\emptyset \in \mathcal{A}_\alpha$ für alle $\alpha \in \mathcal{I}$. Weiterhin gilt

$$\begin{aligned} A \in \bigcap_{\alpha \in \mathcal{I}} \mathcal{A}_\alpha &\Leftrightarrow \forall \alpha \in \mathcal{I} : A \in \mathcal{A}_\alpha \Rightarrow \forall \alpha \in \mathcal{I} : A^c \in \mathcal{A}_\alpha \\ &\Leftrightarrow A^c \in \bigcap_{\alpha \in \mathcal{I}} \mathcal{A}_\alpha, \end{aligned}$$

sowie

$$\begin{aligned} A_1, A_2, \dots \in \bigcap_{\alpha \in \mathcal{I}} \mathcal{A}_\alpha &\Leftrightarrow \forall \alpha \in \mathcal{I} : A_1, A_2, \dots \in \mathcal{A}_\alpha \\ &\Rightarrow \forall \alpha \in \mathcal{I} : \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}_\alpha \Leftrightarrow \bigcup_{n=1}^{\infty} A_n \in \bigcap_{\alpha \in \mathcal{I}} \mathcal{A}_\alpha. \end{aligned}$$

b) Betrachte alle Obermengen von \mathcal{C} , die Algebren sind, und bilde den Schnitt,

$$\bigcap_{\substack{\mathcal{A}_\alpha \text{ Algebra} \\ \mathcal{C} \subseteq \mathcal{A}_\alpha}} \mathcal{A}_\alpha$$

Aus a) folgt, dass der Schnitt eine Algebra ist. Der Schnitt ist dann auch *per definitionem* die kleinste σ -Algebra, die \mathcal{C} enthält. «

1.3 **Definition** Sei \mathcal{C} ein Mengensystem in Ω . Die kleinste der \mathcal{C} enthaltenden Algebra heißt die *von \mathcal{C} erzeugte Algebra*; \mathcal{C} heißt ein *Erzeugersystem* dieser Algebra. Entsprechendes gilt für σ -Algebren.

Wir bezeichnen die von \mathcal{C} erzeugte Algebra mit $\mathcal{F}(\mathcal{C})$. \times

Das Konzept des Erzeugersystems ermöglicht es uns, Eigenschaften nur auf dem Erzeugersystem nachzuweisen, was oft einfacher ist, und diese Eigenschaften auf die erzeugte Algebra zu übertragen.

Im \mathbb{R}^n bilden die offenen Teilmengen ein wichtiges Mengensystem, die sie die Topologie erzeugen und somit Metrik, Norm, Konvergenz, Differenziation, etc. charakterisieren. Folgende Definition ist daher für alles Weitere zentral.

1.4 **Definition** Sei \mathcal{O}_n das System der offenen Mengen des Euklidischen Raumes \mathbb{R}^n . Setze $\mathcal{B}_n := \mathcal{F}(\mathcal{O}_n)$. \mathcal{B}_n wird als σ -Algebra der *Borelschen Mengen in \mathbb{R}^n* bezeichnet. $\mathcal{B} := \mathcal{B}_1$. \times

\mathcal{B}_n enthält alle abgeschlossenen, alle kompakten und alle höchstens abzählbaren Mengen in \mathbb{R}^n . Es gilt dennoch tatsächlich $\mathcal{B}_n \subsetneq \mathcal{P}(\mathbb{R}^n)$, der Beweis benötigt jedoch die Annahme des Auswahlaxioms.

Die Borelschen Mengen bilden die grundlegenden Mengen, mit denen wir uns beschäftigen. Wenn wir auf dem \mathbb{R}^n arbeiten, enthält also \mathcal{B}_n alle "für uns interessanten" Mengen.

Um das Konzept des Erzeugersystems ausnutzen können, verwenden wir das System J_n der halboffenen Intervalle (bzw. Rechtecke)

$$(a, b] := \{(x_1, \dots, x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i, i = 1, \dots, n\}$$

in \mathbb{R}^n , wobei $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n) \in \mathbb{R}^n$ mit $a_i < b_i$, oder auch das System der offenen Intervalle, der abgeschlossenen Intervalle, der Intervalle $(-\infty, a]$ oder der Intervalle $(-\infty, a)$.

1.1 **Satz** Das System J_n ist ein Erzeugersystem von \mathcal{B}_n . \times

» *per definitionem* gilt $(a, b] := \{(x_1, \dots, x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i, i = 1, \dots, n\}$. Nun lässt sich jedes offene Intervall als abzählbare Vereinigung halboffener Intervalle darstellen,

$$\begin{aligned} (a, b) &= \bigcup_{k \in \mathbb{N}} \left(a, b - \frac{1}{k} \right] \\ &= \bigcup_{k \in \mathbb{N}} \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i - \frac{1}{k} \right\} \in \mathcal{B}_n. \end{aligned}$$

Ferner lässt sich jede offene Menge in \mathbb{R}^n als *abzählbare* Vereinigung von offenen Intervallen mit rationalen Randpunkten darstellen. «

Wir haben nun die notwendigen Vorbereitungen getroffen um ein Maß zu definieren.

1.5 **Definition** Sei \mathcal{C} ein Mengensystem in Ω mit $\emptyset \in \mathcal{C}$. Eine Mengenfunktion

$$\mu : \mathcal{C} \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$$

heißt (mit den Konventionen $a + \infty = \infty$ ($a \in \mathbb{R}$), $\infty + \infty = \infty$ usw.)

1.) ein *Inhalt (content)* auf \mathcal{C} , wenn

(a) μ ist nulltreu, d.h. $\mu(\emptyset) = 0$,

(b) μ ist positiv, d.h. $\forall A \in \mathcal{C} : \mu(A) \geq 0$,

(c) μ ist additiv, d.h. für paarweise disjunkte $A_n \in \mathcal{C}$ mit $n = 1, 2, \dots, m$ und $\sum_{n=1}^m A_n \in \mathcal{C}$ gilt²

$$\mu \left(\sum_{n=1}^m A_n \right) = \sum_{n=1}^m \mu(A_n).$$

2.) ein *Maß (measure)* auf \mathcal{C} , wenn

(a) μ ist nulltreu, d.h. $\mu(\emptyset) = 0$,

(b) μ ist positiv, d.h. $\forall A \in \mathcal{C} : \mu(A) \geq 0$,

² $A + B$ bzw. $\sum_{n=1}^m A_n$ steht für die disjunkte Vereinigung von A und B bzw. der A_n . Wir fordern bei der Verwendung von $+$ und \sum implizit, dass A und B bzw. die A_n disjunkt sind.

(c) μ ist σ -additiv, d.h. für paarweise disjunkte $A_n \in \mathcal{C}$ mit $n \in \mathbb{N}$ und $\sum_{n=1}^{\infty} A_n \in \mathcal{C}$ gilt

$$\mu \left(\sum_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n).$$

Ist \mathcal{A} eine σ -Algebra in Ω , μ ein Maß auf \mathcal{A} , so heißt $(\Omega, \mathcal{A}, \mu)$ ein **Maßraum (measure space)**. \times

Offensichtlich ist jedes Maß ein Inhalt.

“Natürliche” Definitionsbereiche für Inhalt und Maß sind Algebren bzw. σ -Algebren. In diesem Fall können wir auch auf die Voraussetzung $\sum_{n=1}^m A_n \in \mathcal{C}$ bzw. $\sum_{n=1}^{\infty} A_n \in \mathcal{C}$ verzichten. Häufig werden wir nur diese Definitionsbereiche verwenden.

In der Wahrscheinlichkeitstheorie arbeiten wir häufig auf abzählbaren Messräumen wie z.B. \mathbb{N} dann lassen sich alle Teilmengen messen, indem wir ihnen als Maß die Anzahl ihrer Elemente zuordnen.

1.6 **Definition** Sei (Ω, \mathcal{A}) ein Messraum und $Z = \{z_1, z_2, \dots\}$ eine höchstens abzählbare Teilmenge von Ω , so heißt

$$\mu : \mathcal{A} \rightarrow \overline{\mathbb{R}}, \text{ mit } \mu(A) := \text{Anzahl der Elemente in } A \cap Z, \quad A \in \mathcal{A},$$

ein **abzählendes Maß (counting measure)**. \times

1.2 **Bemerkung.** Sei \mathcal{A} Algebra in Ω , μ endlicher Inhalt auf \mathcal{A} , ferner $A, B \in \mathcal{A}$ und $A \subset B$, so gilt

$$\mu(B \setminus A) = \mu(B) - \mu(A).$$

Diese Eigenschaft nennt sich **Subtraktivität**. \rightarrow

» Da $B = A + B \setminus A$, gilt $\mu(B) = \mu(A) + \mu(B \setminus A)$. \leftarrow

1-B Wahrscheinlichkeitsräume und -maße

1.7 **Definition** Sei \mathcal{A} eine σ -Algebra in Ω . Eine Abbildung $P : \mathcal{A} \rightarrow \mathbb{R}$ heißt ein *Wahrscheinlichkeitsmaß (W-Maß)* auf \mathcal{A} , wenn

(a) P nulltreu, d.h. $P(\emptyset) = 0$,

(b) P positiv, d.h. $P(A) \geq 0$,

(c) P σ -additiv, d.h. für alle paarweise disjunkte $A_n \in \mathcal{A}$ ($n = 1, 2, \dots$) gilt

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

(d) P normiert, d.h. $P(\Omega) = 1$.

(Ω, \mathcal{A}, P) heißt dann ein *Wahrscheinlichkeitsraum (W-Raum) (probability space)*.

Ist (Ω, \mathcal{A}, P) ein W-Raum, so bezeichnet man die Grundmenge Ω auch als *Merkmalsraum, Stichprobenraum (sample space)*, die Mengen $A \in \mathcal{A}$ als *Ereignisse*, $P(A)$ als *Wahrscheinlichkeit von A* und die Elemente $\omega \in \Omega$ bzw. die 1-Punkt-Mengen $\{\omega\} \subset \Omega$ (nicht notwendig $\in \mathcal{A}$) als *Elementarereignisse (auch Realisierungen)*.

Ein W-Raum ist ein normierter Maßraum. \times

Ein Wahrscheinlichkeitsmaß, ist also ein normiertes Maß auf (Ω, \mathcal{A}) mit Wertebereich \mathbb{R} anstatt $\overline{\mathbb{R}}$.

Bei der Definition eines Wahrscheinlichkeitsmaßes, ergeben sich Nulltreue, Positivität und Normiertheit ganz intuitiv. In der Vergangenheit wurde lange darüber diskutiert, ob es notwendig ist, die σ -Additivität für Ereignisse zu fordern. Es hat sich jedoch herausgestellt, dass dies durchaus sinnvoll und notwendig ist.

Damit haben wir die Axiome der Wahrscheinlichkeitstheorie formuliert und alles Folgende wird auf diesen Axiomen aufbauen. Nun stellt sich natürlich die Frage, ob diese Definition einer "Wahrscheinlichkeit" mit unserer Alltagserfahrung übereinstimmt. Im Laufe der Vorlesung werden wir feststellen, dass dieses "Modell" in vielen Fällen eine sehr gute Approximation der Wirklichkeit darstellt.

Zur Motivation des Wahrscheinlichkeitsraumes. Wir suchen nach einem alternativen Weg, eine Wahrscheinlichkeit einzuführen. Sei dazu $h_n(A)$ die absolute Häufigkeit des Eintretens eines Ereignisses A (z.B. Würfeln einer "6" bei n Würfeln) und

$H_n(A) = \frac{h_n(A)}{n}$ die relative Häufigkeit des Auftretens von Ereignis A .

Offensichtlich gilt für jedes A und jedes n :

$$0 \leq H_n(A) \leq 1, \quad H_n(\Omega) = 1, \quad H_n(\emptyset) = 0.$$

Für zwei disjunkte Ereignisse A_1, A_2 gilt weiterhin

$$H_n(A_1 + A_2) = H_n(A_1) + H_n(A_2).$$

H_n verfügt also über alle Eigenschaften eines Wahrscheinlichkeitsmaßes.

Führt man das (Würfel)-Experiment hinreichend oft durch, kann man das “**Empirische Gesetz der großen Zahlen**” vermuten:

$$\lim_{n \rightarrow \infty} H_n(A) \text{ existiert.}$$

Im Fall der Existenz kann man dem Ereignis A den obigen Grenzwert als Wahrscheinlichkeit zuordnen (R.v. Mises 1919³),

$$P(A) = \lim_{n \rightarrow \infty} H_n(A).$$

Dieser Grenzwert muss aber nicht für alle möglichen Folgen von Versuchsergebnissen existieren - es lässt sich nicht ausschließen, dass man in einer Versuchsfolge ausschließlich 6er würfelt.

Die endgültige, axiomatische Formulierung der Wahrscheinlichkeitstheorie, wie wir sie eben eingeführt haben, hat 1933 mit dem Werk “Grundbegriffe der Wahrscheinlichkeitsrechnung” von Kolmogorov⁴ seinen Abschluss gefunden. Der axiomatische Ansatz hat sich bisher als der erfolgreichste erwiesen. Wir zahlen aber einen Preis, denn dieser Ansatz erklärt nicht, was Wahrscheinlichkeit eigentlich bedeutet.

■ Binomialverteilung

Oft sind nur endlich viele Elementarereignisse möglich und alle diese treten mit der gleichen Wahrscheinlichkeit ein. Wir sprechen in diesem Fall von einem Laplace-Experiment.

³ Richard Edler von Mises (* 19. April 1883 in Lemberg, Österreich-Ungarn, heute Lwiw, Ukraine; † 14. Juli 1953 in Boston, Massachusetts) war ein österreichischer Mathematiker.

⁴ Andrei Nikolajewitsch Kolmogorow (* 25. April 1903 in Tambow; † 20. Oktober 1987 in Moskau) war einer der bedeutendsten Mathematiker des 20. Jahrhunderts. Kolmogorow leistete wesentliche Beiträge auf den Gebieten der Wahrscheinlichkeitstheorie und der Topologie, er gilt als der Gründer der Algorithmischen Komplexitätstheorie. Seine bekannteste mathematische Leistung war die Axiomatisierung der Wahrscheinlichkeitstheorie.

Definition Ein W -Raum (Ω, \mathcal{A}, P) mit Ω endlich, $\mathcal{A} = \mathcal{P}(\Omega)$ und

$$P(\{\omega\}) = \frac{1}{|\Omega|}, \quad \omega \in \Omega, \quad |\Omega| = \text{Anzahl der Elemente in } \Omega,$$

heißt *Laplacescher W -Raum*:

$$\text{Für } A \in \mathcal{A} \text{ gilt } P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der "günstigen" Fälle}}{\text{Anzahl der "möglichen" Fälle}}. \quad \times$$

BSP 1 Seien $n \in \mathbb{N}$, $p \in [0, 1]$ fest. Betrachte n gleichartige Experimente ohne gegenseitige Beeinflussung mit den möglichen Ausgängen

0 ("Misserfolg", "Kopf") und 1 ("Erfolg", "Zahl")

mit jeweiliger Erfolgswahrscheinlichkeit p . Dies bezeichnet man auch als n -Tupel von Bernoulli-Versuchen.⁵ ■

Wie groß ist die Wahrscheinlichkeit in n Bernoulli-Versuchen genau k , wobei $k \in \{0, 1, \dots, n\}$. Erfolge zu erhalten?

Dieses Experiment entspricht der Platzierung von k Kugeln auf n Plätzen ohne Mehrfachbelegung eines Platzes.

| Experiment | Wahrscheinlichkeit |
|---|---|
| $\underbrace{1 \dots 1}_k \underbrace{0 \dots 0}_{n-k}$ | $p \cdots p \cdot (1-p) \cdots (1-p) = p^k (1-p)^{n-k}$ |
| $0 \ 1 \ 1 \dots 1 \ 0 \dots 0$ | $(1-p)p \cdots p(1-p) \cdots (1-p) = p^k (1-p)^{n-k}$. |

Die Reihenfolge der Erfolge und Misserfolge ist unerheblich für die Wahrscheinlichkeit. Das Experiment besitzt genau

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

mögliche Ausgänge. Aufsummation liefert die Gesamtwahrscheinlichkeit,

$$W = \binom{n}{k} p^k (1-p)^{n-k} =: b(n, p; k),$$

wobei $\binom{n}{k} = 0$, falls $k \notin \{0, 1, \dots, n\}$.

⁵Jakob I. Bernoulli (* 6. Januar 1655 in Basel; † 16. August 1705 ebenda) war ein Schweizer Mathematiker und Physiker.

Definition Durch $P(\{k\}) := b(n, p; k)$ für $k \in \mathbb{N}_0$ ist ein W -Maß auf \mathcal{B} definiert, wobei für $A \subseteq \mathbb{N}_0$,

$$P(A) = \sum_{k \in \mathbb{N}_0 \cap A} b(n, p; k) = \sum_{k \in \mathbb{N}_0 \cap A} \binom{n}{k} p^k (1-p)^{n-k}.$$

P ist ein W -Maß und heißt **Binomialverteilung** $b(n, p)$ mit Parameteren $n \in \mathbb{N}$ und $p \in [0, 1]$. \times

Bsp 2 500g Teig und 36 Rosinen seien rein zufällig verteilt. Daraus werden 10 Brötchen mit je 50g Teig geformt. Greife ein Brötchen heraus.

Wie groß ist die Wahrscheinlichkeit, dass dieses Brötchen 2 oder mehr Rosinen enthält?

Rosine Nr. 1-36 ist unabhängig von den anderen mit Wahrscheinlichkeit $p = \frac{1}{10}$ im betrachteten Brötchen. Es handelt sich also um 36 Bernoulli Versuche mit jeweiliger Erfolgswahrscheinlichkeit $p = \frac{1}{10}$.

$$\begin{aligned} P(A) &= \sum_{k=2}^{36} b(n, p; k) = \sum_{k=2}^{36} \binom{36}{k} \frac{1}{10^k} \left(1 - \frac{1}{10}\right)^{36-k} \\ &= 1 - b\left(36, \frac{1}{10}; 0\right) - b\left(36, \frac{1}{10}; 1\right) \approx 0.89. \quad \blacksquare \end{aligned}$$

Wir wollen das Beispiel der Binomialverteilung nun dahingehend verallgemeinern, dass wir die Menge Ω auf die ganze reelle Achse erweitern.

Definition Sei $\Omega = \mathbb{R}$, $\mathcal{A} = \mathcal{B}$ und $(p_k)_{k \in \mathbb{N}_0}$ eine reelle Zahlenfolge mit

$$0 \leq p_k \leq 1 \quad \text{für } k \in \mathbb{N}_0, \quad \sum_{k \in \mathbb{N}_0} p_k = 1.$$

(p_k) bezeichnet man als **Zähldichte**. Durch $P(\{k\}) = p_k$, $\forall k \in \mathbb{N}_0$ also

$$P(A) = \sum_{k \in A \cap \mathbb{N}_0} p_k, \quad A \in \mathcal{B}.$$

wird ein W -Maß auf \mathcal{B} definiert. P ist auf \mathbb{N}_0 **konzentriert**, d.h. $P(\mathbb{N}_0) = 1$ und besitzt die Zähldichte $(p_k)_{k \in \mathbb{N}_0}$. \times

» Nulltreue, Positivität und Normiertheit sind klar. Wir weisen nun die σ -Additivität nach: Seien A_n für $n \in \mathbb{N}$ paarweise disjunkt. Ohne Einschränkung können wir davon ausgehen, dass $A_n \subseteq \mathbb{N}$, es gilt also,

$$P\left(\sum_{n=1}^{\infty} A_n\right) = \sum_{k \in \sum_n A_n} p_k.$$

Da die Reihe absolut konvergent ist, dürfen wir den großen Umordnungssatz aus der Analysis anwenden, und die Reihenglieder umgruppieren,

$$\dots = \sum_{n=1}^{\infty} \underbrace{\sum_{k \in A_n} p_k}_{=: P(A_n)} = \sum_{n=1}^{\infty} P(A_n). \quad \ll$$

Für Beispiel 1 gilt $(b(n, p; k))_{k \in \mathbb{N}_0} = (p_k)_{k \in \mathbb{N}_0}$.

■ Poissonverteilung

Seien nun $p_n \in (0, 1)$ mit $n \cdot p_n \rightarrow \lambda \in (0, 1)$ für $n \rightarrow \infty$, so ist

$$\begin{aligned} b(n, p_n; k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} p_n^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} p_n^k \cdot (1 - p_n)^{-k} \left[(1 - p_n)^{1/p_n} \right]^{np_n} \end{aligned}$$

Für den Grenzübergang $n \rightarrow \infty$ gilt,

$$\begin{aligned} \frac{n(n-1) \cdots (n-k+1)}{k!} p_n^k &\rightarrow \frac{\lambda^k}{k!}, \\ (1 - p_n)^{-k} &\rightarrow 1, \\ (1 - p_n)^{1/p_n} &\rightarrow e^{-1}. \end{aligned}$$

Wir erhalten also insgesamt,

$$b(n, p_n; k) \xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda} =: \pi(\lambda; k).$$

$\pi(\lambda; k)$ definiert eine Zähldichte, denn

$$\sum_{k \in \mathbb{N}_0} \pi(\lambda; k) = e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{=: e^{\lambda}} = 1.$$

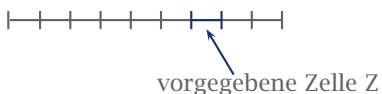
1.8 **Definition** Das zu $\pi(\lambda; k)$ gehörige W-Maß $\pi(\lambda)$ auf \mathcal{B} ist gegeben durch,

$$\pi(\lambda)(\{k\}) = \pi(\lambda; k)$$

und heißt *Poisson-Verteilung*⁶ mit Parameter λ . \times

Die Poisson-Verteilung hat sehr viele angenehme Eigenschaften, insbesondere ist sie numerisch gut handhabbar. Daher approximiert man oft eine Binomialverteilung durch eine Poissonverteilung. Die Approximation ist gut für p “klein” und n “groß”.

BSP 3 Eine Anwendung zu Beispiel 2. Gegeben seien m gleichgroße mehrfach besetzbare Zellen.



1.1 Zur Anwendung von 2.

n Teilchen seien rein zufällig auf die Zellen verteilt. Die Wahrscheinlichkeit, dass Teilchen Nr. 1 in Zelle Z landet, ist $\frac{1}{m}$. Also ist die Wahrscheinlichkeit, dass k Teilchen in Z landen $b(n, \frac{1}{m}; k)$. Die “Belegungsintensität” ist $\frac{n}{m}$.

Für $n \rightarrow \infty$ und $m \rightarrow \infty$ gelte $\frac{n}{m} \rightarrow \lambda \in (0, \infty)$. In der Grenze ist die Wahrscheinlichkeit, dass genau $k \in \mathbb{N}_0$ Teilchen in der Zelle Z landen gegeben durch $\pi(\lambda; k)$. Damit können wir das Beispiel 2 wieder aufgreifen.

Betrachte eine große Anzahl von Rosinen in einer großen Teigmenge. Teile den Teig in m gleichgroße Brötchen auf, mit $\lambda := \frac{n}{m}$. Nehmen wir nun ein Brötchen heraus, so ist die Wahrscheinlichkeit, dass es genau k Rosinen enthält

$$\approx e^{-\lambda} \frac{\lambda^k}{k!}.$$

Im Zahlenbeispiel $n = 36$, $m = 10$, $\lambda = 3.6$ gilt

$$\sum_{k=2}^{36} e^{-\lambda} \frac{\lambda^k}{k!} \approx 1 - e^{-\lambda} - \lambda e^{-\lambda} \approx 0.874, \quad \text{statt } 0.89. \quad \blacksquare$$

■ Weitere Eigenschaften von Wahrscheinlichkeitsmaßen

Die reellen Zahlen \mathbb{R} sind geordnet und für “monoton fallende” bzw. “monoton steigende” Folgen wissen wir, dass ein eigentlicher oder uneigentlicher Grenzwert

⁶Siméon-Denis Poisson (* 21. Juni 1781 in Pithiviers (Département Loiret); † 25. April 1840 in Paris) war ein französischer Physiker und Mathematiker.

existiert. Ein Mengensystem lässt sich durch “ \subseteq ” oder “ \supseteq ” halbordnen, somit lässt sich der Monotoniebegriff in gewisser Weise übertragen.

Definition Die Konvergenz *von unten* bzw. *von oben* ist für die Mengen A und A_n ($n = 1, 2, \dots$) folgendermaßen definiert:

$$A_n \uparrow A : A_1 \subset A_2 \subset A_3 \subset \dots, \bigcup_{n=1}^{\infty} A_n = A,$$

$$A_n \downarrow A : A_1 \supset A_2 \supset A_3 \supset \dots, \bigcap_{n=1}^{\infty} A_n = A. \quad \times$$

Da wir nun über eine Art Konvergenzbegriff für Mengen verfügen, lässt sich nach der Stetigkeit von Abbildungen von Mengensystemen in die reellen Zahlen fragen. Natürlich handelt es sich hierbei nicht um Konvergenz und Stetigkeit im Sinne der Analysis, denn wir haben auf \mathcal{A} keine Norm zur Verfügung.

1.2 **Satz** Sei \mathcal{A} eine Algebra in Ω , μ ein endlicher (d.h. $\mu(\Omega) < \infty$) Inhalt auf \mathcal{A} . Dann sind die folgenden Aussagen äquivalent:

- a) μ ist σ -additiv
- b) μ ist stetig von unten,
d.h. $[A_n \in \mathcal{A}, A_n \uparrow A \in \mathcal{A} \Rightarrow \mu(A_n) \rightarrow \mu(A)]$
- c) μ ist stetig von oben,
d.h. $[A_n \in \mathcal{A}, A_n \downarrow A \in \mathcal{A} \Rightarrow \mu(A_n) \rightarrow \mu(A)]$
- d) μ ist \emptyset -stetig,
d.h. $[A_n \in \mathcal{A}, A_n \downarrow \emptyset \Rightarrow \mu(A_n) \rightarrow 0]$.

Auch ohne die Voraussetzung der Endlichkeit von μ gilt a) \Leftrightarrow b). \times

» a) \Rightarrow b): Sei $A_n = A_1 + A_2 \setminus A_1 + \dots + A_n \setminus A_{n-1}$ und $A = A_1 + \sum_{k=2}^{\infty} A_k \setminus A_{k-1}$. Wir verwenden die Additivität von μ und erhalten somit,

$$\begin{aligned} \mu(A_n) &= \mu(A_1) + \mu(A_2 \setminus A_1) + \dots + \mu(A_n \setminus A_{n-1}) \\ &= \mu(A_1) + \sum_{k=2}^n \mu(A_k \setminus A_{k-1}) \xrightarrow{n \rightarrow \infty} \mu(A_1) + \sum_{k=2}^{\infty} \mu(A_k \setminus A_{k-1}) = \mu(A). \end{aligned}$$

Die übrigen Implikationen folgen analog. \ll

Da Wahrscheinlichkeitsräume *per definitionem* endlich sind, ist Satz 1.2 dort stets anwendbar. Insbesondere ist jedes W-Maß stetig.

1.3 **Lemma** Sei \mathcal{A} eine Algebra in Ω , μ ein Inhalt auf \mathcal{A} .

a) μ ist monoton,

$$\text{d.h. } [A, B \in \mathcal{A}, A \subset B \Rightarrow \mu(A) \leq \mu(B)].$$

b) μ ist subadditiv,

$$\text{d.h. } [A_1, \dots, A_m \in \mathcal{A} \Rightarrow \mu(\bigcup_{n=1}^m A_n) \leq \sum_{n=1}^m \mu(A_n)].$$

c) Ist μ ein Maß, dann ist μ sogar σ -subadditiv,

$$\text{d.h. } [A_n \in \mathcal{A} \text{ für } n \in \mathbb{N}, \bigcup_{n=1}^{\infty} A_n \in \mathcal{A} \Rightarrow \mu(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)]. \quad \times$$

» a): Seien $A, B \in \mathcal{A}$ und $A \subseteq B$. Nun gilt $B = A + B \setminus A$, also

$$\begin{aligned} \mu(B) &= \mu(A) + \underbrace{\mu(B \setminus A)}_{\geq 0}, \\ &\Rightarrow \mu(A) \leq \mu(B). \end{aligned}$$

b): Wir untersuchen den Spezialfall $n = 2$. Seien $A_1, A_2 \in \mathcal{A}$, so ist

$$\begin{aligned} \mu(A_1 \cup A_2) &= \mu(A_1 + A_2 \setminus A_1) = \mu(A_1) + \underbrace{\mu(A_2 \setminus A_1)}_{\leq \mu(A_2)} \\ &\leq \mu(A_1) + \mu(A_2). \end{aligned}$$

Der allgemeine Fall folgt durch Induktion.

c): Für $N \in \mathbb{N}$ gilt,

$$\mu(A_1 \cup \dots \cup A_N) \leq \sum_{n=1}^N \mu(A_n) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

Beim Grenzübergang für $N \rightarrow \infty$ erhalten wir so,

$$\mu(A_1 \cup \dots \cup A_N) \rightarrow \mu\left(\bigcup_{n=1}^{\infty} A_n\right),$$

und die Behauptung folgt. «

W-Maße sind also σ -additiv, σ -subadditiv, monoton und stetig. Diese Eigenschaften sind zwar unscheinbar, es lassen sich damit aber sehr starke Aussagen beweisen, wie wir gleich sehen werden.

Erinnern wir uns aber zunächst an den *limes superior* einer reellen Zahlenfolge (a_n) ,

$$\limsup_{n \rightarrow \infty} a_n := \inf_{n \in \mathbb{N}} \sup_{k \geq n} a_k.$$

Wir können nun diese Definitionen auf Mengen übertragen.

Definition Sei (Ω, \mathcal{A}, P) ein W-Raum, $A_n \in \mathcal{A}$ für $n \in \mathbb{N}$

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega : \text{es ex. unendlich viele } n \text{ mit } \omega \in A_n\} \\ &= \text{“Ereignis, dass unendlich viele } A_n \text{ eintreten”}. \quad \times \end{aligned}$$

1.3 **1. Lemma von Borel und Cantelli** Sei (Ω, \mathcal{A}, P) ein W-Raum und $A_n \in \mathcal{A}$ für $n \in \mathbb{N}$, dann gilt

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \Rightarrow \quad P(\limsup A_n) = 0,$$

d.h. P-f.s. gilt: A_n tritt nur endlich oft auf. \times

Bsp 4 Lernen durch Erfahrung, d.h. je öfter man eine Tätigkeit durchführt, je kleiner ist die Wahrscheinlichkeit, dass ein Fehler auftritt.

Betrachte eine Folge von Aufgaben. Ein Misserfolg in der n -ten Aufgabe tritt mit der Wahrscheinlichkeit

$$P(A_n) = \frac{1}{n^2}$$

auf. Es gilt dann

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

1.3 \Rightarrow Mit Wahrscheinlichkeit 1 wird von einem zufälligen Index n an *jede* Aufgabe immer richtig gelöst. ■

» *Beweis des 1. Lemma von Borel und Cantelli.* Sei $B := \limsup_n A_n$, d.h.

$$B = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k.$$

Offensichtlich gilt $\bigcup_{k \geq n} A_k \downarrow B$. Aufgrund der Monotonie von P gilt somit,

$$P(B) \leq P\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k) \rightarrow 0,$$

da die Reihe nach Voraussetzung konvergent ist. Also ist $P(B) = 0$. «

■ Fortsetzung von Wahrscheinlichkeitsmaßen

Bisher haben wir ausschließlich mit Maßen gearbeitet, die auf \mathbb{N}_0 konzentriert sind, d.h. $P(\mathbb{N}_0) = 1$. Unser Ziel ist es jedoch ein Maß auf \mathcal{B}_n zu definieren, das jedem Intervall (Rechteck) seinen elementargeometrischen Inhalt zuordnet.

Betrachten wir wieder das Erzeugersystem J_n der halboffenen Intervalle, so bildet dieses einen Halbring.

Definition $h \subseteq \mathcal{P}(\Omega)$ heißt *Halbring* über Ω , falls

- 1.) $\emptyset \in h$,
- 2.) $A, B \in h \Rightarrow A \cap B \in h$,
- 3.) $A, B \in h, A \subseteq B \Rightarrow \exists k \in \mathbb{N} : C_1, \dots, C_k \in h : B \setminus A = \sum_{i=1}^k C_i$. ✕

Auf dem Halbring J_n können wir das gesuchte Maß definieren durch,

$$\mu : J_n \rightarrow \mathbb{R}, \quad (a_i, b_i] \mapsto \mu((a_i, b_i]) = \prod_{i=1}^n (b_i - a_i).$$

Mit Hilfe des folgenden Satzes können wir μ *eindeutig* auf \mathcal{B}_n fortsetzen.

1.4 Fortsetzungs- und Eindeutigkeitssatz Sei μ ein Maß auf einem Halbring h über Ω , so besitzt μ eine Fortsetzung μ^* zu einem Maß auf $\mathcal{F}(h)$.

Gilt außerdem $\Omega = \bigcup_{i=1}^{\infty} A_i$ mit $\mu(A_i) < \infty$, so ist μ^* *eindeutig*. ✕

» Siehe Elstrodt, Kap. III, §§ 4.5. «

Insbesondere lassen sich auf Halbringen bzw. Algebren über W -Räumen definierte Maße immer eindeutig fortsetzen, da hier bereits $\mu(\Omega) < \infty$.

1.3 *Bemerkung.* μ^* in Satz 1.4 ist für $A \in \mathcal{F}(\mathcal{A})$ gegeben durch

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(B_n) : B_n \in \mathcal{A} \ (n = 1, 2, \dots) \text{ mit } A \subset \bigcup_n B_n \right\}. \quad \circ$$

1.4 *Bemerkung.* Es ist im Allgemeinen nicht möglich, μ von \mathcal{A} auf $\mathcal{P}(\Omega)$ fortzusetzen. H. Lebesgue⁷ hat sogar bewiesen, dass es unter Annahme des Auswahlaxioms nicht möglich ist, ein Maß auf $\mathcal{P}([0, 1])$ fortzusetzen. Ist Ω jedoch abzählbar, so kann P stets auf die Potenzmenge fortgesetzt werden. \circ

1.5 **Korollar** *Es gibt genau ein Maß λ auf \mathcal{B}_n , das jedem beschränkten (nach rechts halboffenen) Intervall in \mathbb{R}^n seinen elementargeometrischen Inhalt zuordnet.* \times

1.9 **Definition** *Das Maß λ in Korollar 1.5 heißt Lebesgue-Borel-Maß (L-B-Maß) in \mathbb{R}^n .* \times

BSP 5 *Rein zufälliges Herausgreifen eines Punktes aus dem Intervall $[0, 1]$.*

Sei $\Omega = [0, 1]$, $\mathcal{A} := [0, 1] \cap \mathcal{B} := \{B \cap [0, 1] : B \in \mathcal{B}\}$. Das W-Maß P auf \mathcal{A} soll jedem Intervall in $[0, 1]$ seine elementare Länge zuordnen. Also ist

$$P = \lambda|_{[0,1] \cap \mathcal{B}}$$

eine Restriktion von λ auf $[0, 1] \cap \mathcal{B}$. \blacksquare

Das λ -Maß hat die Eigenschaft, dass nicht nur die leere Menge sondern auch jede höchstens abzählbare Menge das Maß Null hat. Eine Menge $N \subseteq \Omega$ mit $\mu(N) = 0$ nennt man **Nullmenge**.

Oft kann man Eigenschaften nicht auf ganz Ω nachweisen aber auf Ω bis auf eine Nullmenge. Solche Eigenschaften treten vor allem in der Maßtheorie auf und sind daher auch in der Wahrscheinlichkeitstheorie von großer Bedeutung.

Definition *Sei Maßraum $(\Omega, \mathcal{A}, \mu)$, W-Raum (Ω, \mathcal{A}, P) . Eine Aussage gilt*

- 1.) *μ -fast überall (μ -f.ü.) in Ω bzw. P -fast sicher (P -f.s.) in Ω , falls sie überall bis auf einer Menge $\in \mathcal{A}$ vom μ -Maß bzw. P -Maß Null [oder eine Teilmenge hiervon] gilt.*
- 2.) *L -fast überall (L -f.ü.) in $\mathbb{R}^n \dots$, wenn sie überall bis auf einer Menge $\in \mathcal{B}_n$ vom L -B-Maß Null [oder eine Teilmenge hiervon] gilt.* \times

⁷ Henri Léon Lebesgue (* 28. Juni 1875 in Beauvais; † 26. Juli 1941 in Paris) war ein französischer Mathematiker.

1-C Verteilungsfunktionen

W-Maße sind Abbildungen der Art

$$P : \sigma \subseteq \mathcal{P}(\Omega) \rightarrow \mathbb{R}.$$

Ihr Definitionsbereich ist somit ein Mengensystem. Wir konnten einige elementare Eigenschaften wie Stetigkeit und Monotonie aus der reellen Analysis auf diese Abbildungen übertragen, dennoch lassen sie sich nicht immer so leicht handhaben wie Funktionen.

Betrachten wir nun ein W-Maß auf \mathcal{B} , so ist durch folgende Vorschrift eine Funktion gegeben

$$F : \mathbb{R} \rightarrow \mathbb{R}, \quad F(x) := P((-\infty, x]), \quad x \in \mathbb{R}.$$

F wird als die zu P gehörende **Verteilungsfunktion (Verteilungsfunktion)** bezeichnet. Verteilungsfunktionen auf \mathbb{R} lassen sich in der Regel viel leichter handhaben als Maße auf \mathcal{B} .

In diesem Abschnitt werden wir zeigen, dass eine Bijektion zwischen Maßen und Verteilungsfunktionen auf einem W-Raum existiert, wir also jedem Maß eine Verteilungsfunktion zuordnen können und umgekehrt. Durch die Verteilungsfunktionen werden die Maße “greifbar”.

BSP 6 a.) Würfeln mit Wahrscheinlichkeiten $\underbrace{p_1, \dots, p_5}_{\geq 0}$ für die Augenzahlen $1, \dots, 6$

wobei $p_1 + \dots + p_6 = 1$.

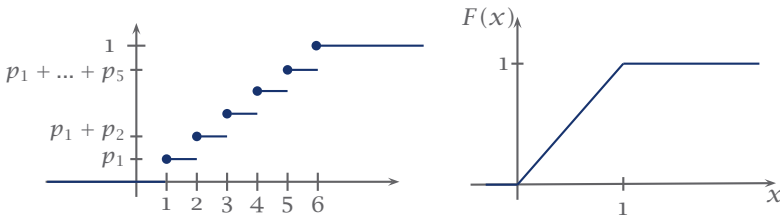
Als zugehörigen Wahrscheinlichkeitsraum wählen wir $\Omega = \mathbb{R}$ mit $\mathcal{A} = \mathcal{B}$. Das W-Maß auf \mathcal{B} erfüllt $P(\{1\}) = p_1, \dots, P(\{6\}) = p_6$.

Zur Verteilungsfunktion betrachte die Skizze.

b.) Betrachte erneut Beispiel 5: Sei $(\Omega, \mathcal{A}, P) = ([0, 1], [0, 1] \cap \mathcal{B}, \lambda|_{[0,1] \cap \mathcal{B}})$ abgewandelt zu $(\Omega, \mathcal{A}, P) = (\mathbb{R}, \mathcal{B}, P)$. Das Lebesgue-Maß ist nur dann ein Wahrscheinlichkeitsmaß, wenn man den zugrundeliegenden Raum auf ein Intervall der Länge Eins einschränkt.

Wir setzen daher $P(B) := \lambda([0, 1] \cap B)$. Die zugehörige Verteilungsfunktion ist dann gegeben durch,

$$F(x) = \begin{cases} 0, & x \leq 0, \\ x, & x \in (0, 1), \\ 1, & x \geq 1. \end{cases} \blacksquare$$



1.2 Verteilungsfunktionen zum Würfelexperiment und zu Beispiel 5

1.10 **Definition** Eine Funktion $F : \mathbb{R} \rightarrow \mathbb{R}$, die monoton wachsend (im Sinne von nicht fallend) und rechtsseitig stetig ist mit

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1,$$

heißt (eindimensionale) **Verteilungsfunktion (Verteilungsfunktion)**. \times

Der folgende Satz stellt nun eine eindeutige Beziehung zwischen einem Maß auf \mathcal{B} und einer eindimensionalen Verteilungsfunktion her.

1.6 **Satz** Zu jedem W -Maß P auf \mathcal{B} gibt es genau eine Verteilungsfunktion $F : \mathbb{R} \rightarrow \mathbb{R}$, so dass gilt

$$F(b) - F(a) = P((a, b]), \quad a \leq b, \quad a, b \in \mathbb{R} \quad (*)$$

gilt, und umgekehrt. Es existiert also eine Bijektion $P \longleftrightarrow F$. \times

» Wir wählen für die Verteilungsfunktion den Ansatz,

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

1.) Die (*)-Bedingung ist erfüllt, denn es gilt für $a, b \in \mathbb{R}$ mit $a < b$,

$$P((a, b]) = P((-\infty, b] \setminus (-\infty, a]) = P((-\infty, b]) - P((-\infty, a]).$$

2.) F ist monoton steigend, denn P ist monoton.

3.) F ist rechtsstetig, denn für $x_n \downarrow x$ folgt $(-\infty, x_n] \downarrow (-\infty, x]$. Somit gilt nach Satz 1.2,

$$F(x_n) = P((-\infty, x_n]) \downarrow P((-\infty, x]) = F(x).$$

4.) $F(x) \rightarrow 0$ für $x \rightarrow -\infty$, denn

$$x_n \downarrow -\infty \Rightarrow (-\infty, x_n] \downarrow \emptyset.$$

Mit Satz 1.2 folgt daraus,

$$F(x_n) = P((-\infty, x_n]) \downarrow P(\emptyset) = 0.$$

5.) $F(x) \rightarrow 1$ für $x \rightarrow \infty$, denn

$$x_n \rightarrow \infty \Rightarrow (-\infty, x_n] \uparrow \mathbb{R}.$$

Ebenfalls mit Satz 1.2 folgt

$$F(x_n) = P((-\infty, x_n]) \rightarrow P(\mathbb{R}) = 1.$$

6.) *Eindeutigkeit.* Seien F, F^* zwei Verteilungsfunktion gemäß der (*)-Bedingung in Satz 1.6. Dann unterscheiden sich F und F^* lediglich um eine Konstante, d.h.

$$F^* = F + c.$$

Nun gilt aber $F^*(x), F(x) \rightarrow 1$ für $x \rightarrow +\infty$, also $c = 0$, d.h. $F = F^*$. «

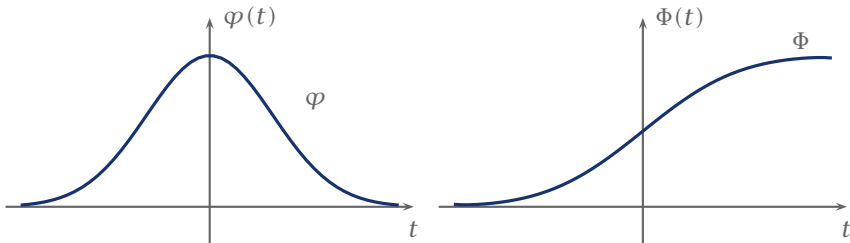
Wir betrachten nun zwei sehr wichtige Verteilungsfunktionen als Beispiele.

BSP 7 Die Funktion $\Phi : \mathbb{R} \rightarrow [0, 1]$ mit

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \underbrace{e^{-\frac{t^2}{2}}}_{\varphi(t)} dt, \quad x \in \mathbb{R},$$

ist eine Verteilungsfunktion, denn Φ ist streng monoton, da $\varphi > 0$, die Rechtsstetigkeit folgt aus der Stetigkeit des Integrals. $F(x) \rightarrow 0$ für $x \rightarrow -\infty$ ist offensichtlich, einzig $F(x) \rightarrow 1$ für $x \rightarrow \infty$ muss man explizit nachrechnen (Funktionentheorie).

Der Integrand $\varphi(t)$ heißt übrigens Gaußsche Glockenkurve und das zu Φ gehörige W-Maß auf \mathcal{B} , **standardisierte Normverteilung** $N(0, 1)$.



1.3 Gaußsche Glockenkurve und standardisierte Normalverteilung

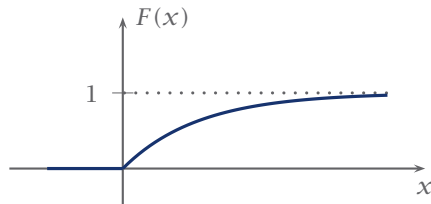
Eine Verallgemeinerung davon ist die **Normalverteilung (Gauß-Verteilung)** $N(a, \sigma^2)$, mit $a \in \mathbb{R}$ und $\sigma > 0$. $N(a, \sigma^2)$ ist ein W-Maß auf \mathcal{B} mit Verteilungsfunktion

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

Diese Funktion ergibt sich mittels Substitution aus der Verteilungsfunktion von $N(0, 1)$, alle Eigenschaften übertragen sich daher entsprechen. ■

BSP 8 Wähle $\lambda > 0$ fest, so ist eine Verteilungsfunktion gegeben durch,

$$F: \mathbb{R} \rightarrow \mathbb{R}, \quad F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$



1.4 Verteilungsfunktion der Exponentialverteilung.

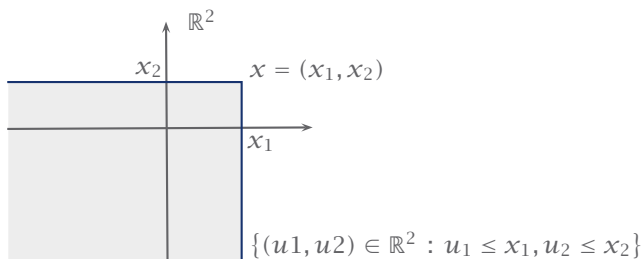
Das zugehörige W-Maß heißt **Exponentialverteilung** mit Parameter λ (kurz $\exp(\lambda)$).



1.5 *Bemerkung.* Definition 1.10 und Satz 1.6 lassen sich auf \mathcal{B}_n und \mathbb{R}^n als Definitionsbereiche von P bzw. F anstatt \mathcal{B} bzw. \mathbb{R} verallgemeinern. Dabei ist die Verteilungsfunktion F zu einem W-Maß P auf \mathcal{B}_n gegeben durch

$$F(x_1, \dots, x_n) = P(\{(u_1, \dots, u_n) \in \mathbb{R}^n \mid u_1 \leq x_1, \dots, u_n \leq x_n\})$$

für $(x_1, \dots, x_n) \in \mathbb{R}^n$. \rightarrow



1.5 Zur Verallgemeinerung der Verteilungsfunktion.

1-D Bedingte Wahrscheinlichkeiten

Wenn wir beim Durchführen eines Zufallsexperiments wissen, dass bereits Ereignis A eingetreten ist, so ändert sich die Wahrscheinlichkeit dafür, dass auch noch B eintritt. Für diese Wahrscheinlichkeit schreiben wir

$$P(B | A).$$

$P(B | A)$ ist die bedingte Wahrscheinlichkeit, dass unter der Voraussetzung, dass A eingetreten ist, nun B eintritt.

Bedingte Wahrscheinlichkeiten spielen in vielen aktuellen Fragestellungen der Wahrscheinlichkeitstheorie eine zentrale Rolle. Beispielsweise in der Finanzmathematik, denn es ist offensichtlich, dass der nächste Aktienkurs von allen vorherigen abhängt.

BSP 9 Unter Studenten einer Vorlesung wird eine Umfrage gemacht.

| | m | w |
|------------|----|----|
| Sport | 12 | 18 |
| Kein Sport | 16 | 20 |

1.6 Ergebnis der Umfrage.

Wir nummerieren die Personen so, dass Nr. 1-12 die Sport-treibenden Frauen, 13-30 die Sport-treibenden Männer, 31-46 die nicht Sport-treibenden Frauen und 47-66 die nicht Sport-treibenden Männer sind.

Nun modellieren wir einen Ω -Raum für die rein zufällige Auswahl einer Person,

$$\Omega = \{1, \dots, 66\}, \quad \mathcal{A} = \mathcal{P}(\Omega), \quad P(\{\omega\}) = \frac{1}{66}.$$

Ereignis $A = \{1, \dots, 12, 31, \dots, 46\}$, die ausgewählte Person ist weiblich,

Ereignis $B = \{1, \dots, 30\}$, die ausgewählte Person treibt Sport.

Es sei bekannt, dass die ausgewählte Person Sport treibt. *Wie groß ist die Wahrscheinlichkeit, dass die ausgewählte Person eine Frau ist?* Abzählen der Elemente in Ω ergibt,

$$\frac{12}{30} = \frac{2}{5}.$$

Andererseits erhalten wir durch

$$\frac{P(A \cap B)}{P(B)} = \frac{\frac{12}{66}}{\frac{30}{66}} = \frac{2}{5}$$

dasselbe Ergebnis. Wir verwenden dies nun zur Definition der bedingten Wahrscheinlichkeit. ■

1.11 **Definition** Es sei (Ω, \mathcal{A}, P) ein W-Raum, $A \in \mathcal{A}$, $B \in \mathcal{A}$ mit $P(B) > 0$. Dann heißt

$$P(A | B) := \frac{P(A \cap B)}{P(B)}$$

die *bedingte Wahrscheinlichkeit (conditional probability)* von A unter der Bedingung B . ✕

Es folgen einige elementare Eigenschaften der bedingten Wahrscheinlichkeit.

1.7 **Satz** Sei (Ω, \mathcal{A}, P) ein W-Raum.

a) Bei festem $B \in \mathcal{A}$ mit $P(B) > 0$ ist $P(\cdot | B)$ ein W-Maß auf \mathcal{A} , das auf B konzentriert ist [d.h. $P(B | B) = 1$].

b) Für $A, B \in \mathcal{A}$ mit $P(A) > 0$, $P(B) > 0$ gilt

$$P(A | B) = P(B | A) \cdot \frac{P(A)}{P(B)}.$$

c) Für $A_n \in \mathcal{A}$ ($n = 1, \dots, m$) mit $P(\bigcap_{n=1}^{m-1} A_n) > 0$ gilt

$$P\left(\bigcap_{n=1}^m A_n\right) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot \dots \cdot P\left(A_m | \bigcap_{n=1}^{m-1} A_n\right). \quad \times$$

» a) Der Beweis ist eine leichte Übungsaufgabe.

b) Betrachte die rechte Seite,

$$P(B | A) \cdot \frac{P(A)}{P(B)} = \frac{P(A \cap B)}{P(A)} \frac{P(A)}{P(B)} = \frac{P(A \cap B)}{P(B)} =: P(A | B).$$

c) Vollständige Induktion. Für $n = 2$

$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1).$$

Der Induktionsschluss ist eine leichte Übung. «

1.8 **Satz** Sei (Ω, \mathcal{A}, P) ein W -Raum, $\{B_n : n \in I\}$ eine höchstens abzählbare Familie paarweise disjunkter Ereignisse $B_n \in \mathcal{A}$ mit $P(B_n) > 0$ ($n \in I$) und $\sum_{n \in I} B_n = \Omega$. Dann gilt für $A \in \mathcal{A}$

a) die *Formel von der totalen Wahrscheinlichkeit*

$$P(A) = \sum_{n \in I} P(B_n)P(A | B_n).$$

b) falls $P(A) > 0$ die *Formel von Bayes*

$$P(B_k | A) = \frac{P(B_k) \cdot P(A | B_k)}{\sum_{n \in I} P(B_n)P(A | B_n)}, \quad k \in I. \quad \times$$

Mit der Formel von der totalen Wahrscheinlichkeit kann man somit mit bedingten Wahrscheinlichkeiten auf die Wahrscheinlichkeit für ein bestimmtes Ereignis rückschließen. Die Formel von Bayes hingegen erlaubt es die Bedingung der Bedingten Wahrscheinlichkeit umzukehren.

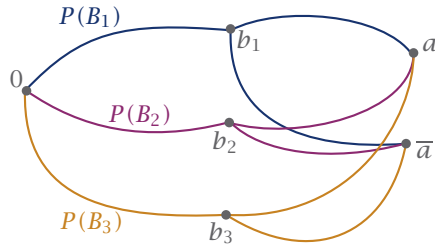
» a) $A = A \cap \Omega = \sum_{n \in I} (A \cap B_n)$. Weiterhin gilt aufgrund der σ -Additivität,

$$P(A) = \sum_{n \in I} P(A \cap B_n) = \sum_{n \in I} P(B_n)P(A | B_n).$$

b) Satz 1.7 besagt, $P(B_k) = P(A|B_k) \frac{P(B_k)}{P(A)}$. Die Behauptung folgt nun unter Verwendung des Teils a). «

BSP 10 Zur Veranschaulichung des Satzes 1.8. Betrachte einen Weg von 0 über b_1 oder b_2 oder b_3 oder ... nach a oder \bar{a} . [a könnte z.B. für das Auftreten von Krebs, \bar{a} für das Nichtauftreten von Krebs, b_1 für Rauchen, b_2 für das Ausgesetztsein von giftigen Dämpfen, b_3 ... stehen].

An jeder Verzweigung b_1, b_2, b_3, \dots wird ein Zufallsexperiment durchgeführt. Das Ereignis B_k sei definiert durch, dass der Weg über b_k führt.



1.7 Wegediagramm.

Die Formel von der totale Wahrscheinlichkeit ergibt,

$$P(A) = \sum_n P(A|B_n)P(B_n).$$

Deutung. Betrachte B_1, B_2, \dots als “Ursachen” und A als “Wirkung”.

$P(B_k)$ bezeichnet man als sog. “a priori Wahrsch.” von B_k .

Nun stehe A durch Erfahrung zur Verfügung.

$P(B_k|A)$ bezeichnet man als sog. “a posteriori Wahrsch.” von B_k .

$P(B_k|A)$ ist die Wahrscheinlichkeit, dass der Weg über b_k verlaufen ist, wobei bereits bekannt sei, dass die Ankunft in a (= Ereignis A) stattgefunden hat.

Die Formel von Bayes erlaubt also den Rückschluss von einer Wirkung auf ihre Ursache - zumindest in einem Wahrscheinlichkeitstheoretischen Sinne.

Anwendung. Ein Arzt beobachtet das Symptom a , welches ausschließlich die Ursachen b_1, b_2, \dots haben kann. Gesucht ist $P(B_k|A)$, die Wahrscheinlichkeit für die jeweilige Ursache unter Berücksichtigung des Eintretens von a . ■

BSP 11 *Betrieb mit einer Alarmanalge.* Bekannt seien folgende Daten:

- Bei Einbruch erfolgt Alarm mit Wahrscheinlichkeit 0.99,
- bei Nichteinbruch mit Wahrscheinlichkeit 0.005.
- Die Einbruchswahrscheinlichkeit ist 0.001.

Gesucht ist nun die Wahrscheinlichkeit, dass bei einem Alarm auch tatsächlich ein Einbruch stattfindet. Übertragen wir dies auf unser Modell, so erhalten wir folgendes Schema:

| Ereignis | Beschreibung |
|----------|---------------|
| E | Einbruch |
| e^c | kein Einbruch |
| A | Alarm |
| A^c | kein Alarm |

Die Wahrscheinlichkeiten der einzelnen Ereignisse sind,

$$P(A|E) = 0.99, \quad P(A|e^c) = 0.005, \quad P(E) = 0.001.$$

Gesucht ist $P(E|A)$. Mit dem Satz von Bayes erhalten wir,

$$\begin{aligned} P(E|A) &= \frac{P(A|E)P(E)}{P(A|E)P(E) + P(A|e^c)P(e^c)} = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.005 \cdot 0.999} \\ &= \frac{22}{133} \approx 0.165. \end{aligned}$$

D.h. die Wahrscheinlichkeit, dass bei Auslösung eines Alarms, auch tatsächlich ein Einbruch stattfindet, beträgt lediglich 16.5%.

Die Wahrscheinlichkeit für einen Alarm ist nach der Formel von der totalen Wahrscheinlichkeit

$$P(A) = P(A|E)P(E) + P(A|e^c)P(e^c) = 0.006. \quad \blacksquare$$

2 Kombinatorische Wahrscheinlichkeitsrechnung

In der kombinatorischen Wahrscheinlichkeitsrechnung legen wir Laplacesche W-Räume zugrunde. Die Abzählung der “günstigen” und der “möglichen” Fälle erfolgt systematisch mit Hilfe der Kombinatorik.

- 2.1 **Definition** Gegeben seien n — nicht notwendiger Weise verschiedene — Elemente a_1, \dots, a_n . Das n -Tupel $(a_{i_1}, \dots, a_{i_n})$ mit $i_j \in \{1, \dots, n\}$, $i_j \neq i_k$ für $j \neq k$ ($j, k = 1, \dots, n$) heißt eine *Permutation* der gegebenen Elemente. \times

In der Kombinatorik spielt die mögliche Anzahl an Permutationen eine sehr große Rolle.

- 2.1 **Satz** Die Anzahl der Permutationen von n verschiedenen Elementen ist $n!$.

Eine Verallgemeinerung für nicht notwendiger Weise verschiedene Elemente liefert der folgende

- 2.2 **Satz** Gegeben seien n Elemente; die verschiedenen Elemente darunter seien mit a_1, \dots, a_p bezeichnet (mit $p \leq n$). Tritt a_i n_i -fach auf ($i = 1, \dots, p$), wobei $\sum_{i=1}^p n_i = n$, dann können die n Elemente auf

$$\frac{n!}{n_1! \dots n_p!}$$

verschiedene Arten permutiert werden. \times

- 2.2 **Definition** Sei A eine n -elementige Menge, $k \in \mathbb{N}$.

- 1.) [2.] Jedes k -Tupel (a_1, \dots, a_k) mit nicht notwendig verschiedenen [lauter verschiedenen] $a_i \in A$ heißt eine *Kombination k -ter Ordnung aus A mit [ohne] Wiederholung und mit Berücksichtigung der Anordnung.*

3.) [4.] Werden in a) [2.] Kombinationen, welche dieselben Elemente in verschiedener Anordnung enthalten, als äquivalent aufgefasst, so heißen die einzelnen Äquivalenzklassen *Kombinationen k-ter Ordnung aus A mit [ohne] Wiederholung und ohne Berücksichtigung der Anordnung*. ✕

Interpretation (mit $A = \{1, \dots, n\}$): Aufteilung von k Kugeln auf n Zellen, wobei in 1.) [2.] die Zellen mehrfach [nicht mehrfach] besetzt werden dürfen und die Kugeln unterscheidbar sind (a_i ... Nummer der Zelle, in der die i -te Kugel liegt), in 3.) [4.] die Zellen mehrfach [nicht mehrfach] besetzt werden dürfen und die Kugeln nicht unterscheidbar sind (ein Repräsentant der Äquivalenzklasse ist gegeben durch ein k -tupel von Zahlen aus $\{1, \dots, n\}$, in dem die Zellennummern so oft auftreten, wie die zugehörige Zelle besetzt ist).

2.3 **Satz** Die Anzahl der Kombinationen k -ter Ordnung aus der n -elementigen Menge A - mit [ohne] Wiederholung und mit [ohne] Berücksichtigung der Anordnung - ist gegeben durch

| | <i>m. Wiederholung</i> | <i>o. Wiederholung</i> ($1 \leq k \leq n$) |
|------------------------------|------------------------|--|
| <i>m. Ber. der Anordnung</i> | n^k | $n(n-1)\dots(n-k+1)$ |
| <i>o. Ber. der Anordnung</i> | $\binom{n+k-1}{k}$ | $\binom{n}{k}$ |

Bestimmung von Fakultäten durch Tabellen für $\log n!$ und - bei großem n - durch die *Stirlingsche Formel*

$$n! \cong \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \quad (n \rightarrow \infty)$$

und ihre Verschärfung

$$\exp \frac{1}{12n+1} < \frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} < \exp \frac{1}{12n}, \quad n \in \mathbb{N}. \quad \times$$

» Wir beweisen exemplarisch die Formel für die Anzahl der Kombinationen ohne Berücksichtigung der Anordnung und mit Wiederholungen. Betrachte dazu k Kugeln, die auf n Zellen verteilt werden sollen.



2.1 Zur Platzierung von k Kugeln in n Zellen

Ohne die äußeren Trennwände gibt es also k Kugeln und $n - 1$ Trennwände, d.h. es gibt insgesamt $(n - 1 + k)!$ Permutationen. Berücksichtigen wir nun noch die Ununterscheidbarkeit der Kugeln, so erhalten wir

$$\binom{n + k - 1}{k}. \ll$$

BSP Anwendung in der Physik. Man beschreibt das makroskopische Verhalten von k Teilchen in der Weise, dass man den (der Darstellung des Momentanzustandes dienenden) Phasenraum in n kongruente würfelförmige Zellen zerlegt und die Wahrscheinlichkeit $p(k_1, \dots, k_n)$ bestimmt, dass sich genau k_i der Teilchen in der i -ten Zelle befinden ($i \in \{1, \dots, n\}$). Sei Ω' die Menge aller n -Tupel $(k_1, \dots, k_n) \in \mathbb{N}_0^n$ von Besetzungszahlen mit $\sum_{i=1}^n k_i = k$. W-Maße auf $\mathcal{P}(\Omega')$ (charakterisiert durch p):

a.) **Maxwell-Boltzmann-Statistik** (Unterscheidbarkeit der Teilchen, Mehrfachbesetzbarkeit von Zellen)

$$p(k_1, \dots, k_n) = \frac{k!}{k_1! \dots k_n!} \frac{1}{n^k}.$$

b.) **Bose-Einstein-Statistik** zur Beschreibung des Verhaltens von Photonen (keine Unterscheidbarkeit der Teilchen, jedoch Mehrfachbesetzbarkeit von Zellen)

$$p(k_1, \dots, k_n) = \binom{n + k - 1}{k}^{-1}.$$

c.) **Fermi-Dirac-Statistik** zur Beschreibung des Verhaltens von Elektronen, Protonen und Neutronen (keine Unterscheidbarkeit der Teilchen, keine Mehrfachbesetzbarkeit von Zellen)

$$p(k_1, \dots, k_n) = \begin{cases} \binom{n}{k}^{-1}, & k_i \in \{0, 1\} \\ 0, & \text{sonst.} \quad \blacksquare \end{cases}$$

BSP Die *Binomialverteilung* $b(n, p)$ mit Parametern $n \in \mathbb{N}$, $p \in (0, 1)$ ist ein W-Maß auf \mathcal{B} mit der Zähldichte

$$k \mapsto \begin{cases} \binom{n}{k} p^k (1 - p)^{n-k}, & k = 0, 1, \dots, n \\ 0, & k = n + 1, n + 2, \dots; \end{cases}$$

durch diese wird für n "unabhängige" Versuche mit jeweiliger Erfolgswahrscheinlichkeit p die Wahrscheinlichkeit von k Erfolgen angegeben. \blacksquare

3 Zufallsvariablen und Verteilungen

3-A Meßbare Abbildungen und Zufallsvariablen

Wir betrachten folgende Zufallsexperimente:

- 1.) Zufällige Anzahl der Erfolge bei n Bernoulli-Versuchen.
- 2.) Zufällige Anzahl der an einem Schalter in einem bestimmten Zeitintervall $[0, T]$ eintreffenden Kunden.
- 3.) Zufällige Wartezeit (von 0 an gemessen) bis zum Eintreffen des nächsten Kunden.
- 4.) Zufälliger Abstand eines aus dem Intervall $[-1/2, 1/2]$ herausgegriffenen Punktes von 0.
- 5.) Zufällige Lage eines Partikels im \mathbb{R}^3 zu einem bestimmten Zeitpunkt.

Die zufälligen Größen können durch sogenannte “Zufallsvariablen” modelliert werden.

BSP 1 Zu 4.): Betrachte den W -Raum (Ω, \mathcal{A}, P) mit $\Omega = [-1/2, 1/2]$, $\mathcal{A} = \mathcal{B} \cap \Omega$ und $P = \lambda|_{\mathcal{A}}$ (Restriktion des LB-Maßes auf Mengen aus \mathcal{A}). Ein Punkt $\omega \in \Omega$ hat Abstand $Y(\omega) := |\omega|$ von Null. Y nennen wir Zufallsvariable.

Die Wahrscheinlichkeit, dass $Y \leq x$ mit $x \in [0, 1/2]$ fest, ist gegeben durch

$$P[Y \leq x] := P(\{\omega \in \Omega : Y(\omega) \leq x\}) = [-x, x] \in \mathcal{A}.$$

Allgemeiner ist die Wahrscheinlichkeit, dass $Y \in B$ für $B \in \mathcal{B}$, gegeben durch

$$P[Y \in B] := P(\{\omega \in \Omega : Y(\omega) \in B\}).$$

Damit $P[Y \in B]$ überhaupt definiert ist, muss $[Y \in B]$ eine messbare Menge sein, d.h. im Definitionsbereich von P liegen. Im Folgenden erarbeiten wir die notwendigen Voraussetzungen, dass dem so ist. ■

3.1 **Definition** Seien Ω, Ω' zwei nichtleere Mengen, $A' \subset \Omega'$ und $X : \Omega \rightarrow \Omega'$ eine Abbildung. Die Menge

$$\{\omega \in \Omega : X(\omega) \in A'\} =: X^{-1}(A') =: [X \in A']$$

(in Ω) heißt das **Urbild** von A' bezüglich der Abbildung X ; die somit definierte Abbildung $X^{-1} : \mathcal{P}(\Omega') \rightarrow \mathcal{P}(\Omega)$ heißt **Urbildfunktion** (zu unterscheiden von einer inversen Funktion!). ✕

In vielen Fällen untersuchen wir nicht nur einzelne Mengen sondern ganze Mengensysteme und verwenden daher folgende Bezeichnung als Abkürzung.

Bezeichnung. Sei $X : \Omega \rightarrow \Omega'$ und C' Mengensystem in Ω' . So ist

$$X^{-1}(C') =: \{X^{-1}(A') : A' \in C'\},$$

wobei $X^{-1}(C')$ ein Mengensystem in Ω ist. →

Um zu klären, unter welchen Voraussetzungen $[X \in B]$ messbar ist, wenn B messbar ist, müssen wir untersuchen, wie sich die Urbildfunktion X^{-1} auf σ -Algebren im Bildraum auswirkt.

3.1 **Satz** Sei $X : \Omega \rightarrow \Omega'$.

a) X^{-1} und Mengenoperationen $\cup, \sum, \cap, ^c, \setminus$ sind vertauschbar.

$$\text{z.B. ist } X^{-1}\left(\bigcup_{\alpha \in I} A'_\alpha\right) = \bigcup_{\alpha \in I} X^{-1}(A'_\alpha) \quad \text{für } A'_\alpha \subset \Omega', \alpha \in \text{Indexbereich } I.$$

b) $X^{-1}(\emptyset) = \emptyset$; $X^{-1}(\Omega') = \Omega$.

$$A' \subset B' \subset \Omega' \Rightarrow X^{-1}(A') \subset X^{-1}(B').$$

c) Ist \mathcal{A}' σ -Algebra in Ω' , so ist $X^{-1}(\mathcal{A}')$ σ -Algebra in Ω .

d) $C' \subset \mathcal{P}(\Omega') \Rightarrow X^{-1}(\mathcal{F}_{\Omega'}(C')) = \mathcal{F}_{\Omega}(X^{-1}(C'))$. ✕

» a)-d): Der Beweis sei als Übungsaufgabe überlassen. «

Sei $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$, so ist $X^{-1}(\mathcal{A}')$ stets eine σ -Algebra. Damit die Urbilder messbarer Mengen auch tatsächlich messbar sind, müssen wir also fordern, dass $X^{-1}(\mathcal{A}') \subseteq \mathcal{A}$.

3.2 **Definition** Seien $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ Messräume. Die Abbildung $X : \Omega \rightarrow \Omega'$ heißt \mathcal{A} - \mathcal{A}' -messbar [kurz: *messbar*; *measurable*], wenn gilt:

$$\forall A' \in \mathcal{A}' : X^{-1}(A') \in \mathcal{A}, \quad \text{d.h. } X^{-1}(\mathcal{A}') \subseteq \mathcal{A},$$

d.h. Urbilder von messbaren Mengen in Ω' sind messbare Mengen in Ω .

In diesem Falle verwenden wir die Schreibweise $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$. \times

Wie wir noch sehen werden, ist die Messbarkeit einer Abbildung eine *wesentlich* schwächere Voraussetzung als beispielsweise Stetigkeit oder gar Differenzierbarkeit. Es erfordert sogar einiges an Aufwand, eine nicht messbare Abbildung zu konstruieren.

3.1 **Bemerkung.** In Satz 3.1 c) ist $X^{-1}(\mathcal{A}')$ die kleinste der σ -Algebren \mathcal{A} in Ω mit \mathcal{A} - \mathcal{A}' -Messbarkeit von X . \rightarrow

Wir fassen nun alle nötigen Voraussetzungen in der folgenden Definition zusammen.

3.3 **Definition** Sei (Ω, \mathcal{A}, P) ein W-Raum, (Ω', \mathcal{A}') ein Messraum. Die Abbildung

$$X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$$

heißt *Zufallsvariable (ZV)* auf (Ω, \mathcal{A}, P) [mit Zustandsraum Ω'] (ausführlich: (Ω', \mathcal{A}') -Zufallsvariable auf (Ω, \mathcal{A}, P) ; *random variable*).

1.) $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ heißt *reelle Zufallsvariable* (oft auch nur ZV).

2.) $X : (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ heißt *erweitert-reelle Zufallsvariable*.

3.) $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$ heißt *n-dimensionaler Zufallsvektor*.

4.) $X(\omega)$ für ein $\omega \in \Omega$ heißt eine *Realisierung* der Zufallsvariable X .

Bezeichnung. $\overline{\mathcal{B}} := \{B, B \cup \{+\infty\}, B \cup \{-\infty\}, B \cup \{-\infty, +\infty\} : B \in \mathcal{B}\}$. \times

Zufallsvariablen ermöglichen es uns für ein Experiment lediglich einmal einen W-Raum (Ω, \mathcal{A}, P) zu modellieren und dann für jeden Aspekt, der uns interessiert, eine Zufallsvariable zu konstruieren. $X(\omega)$ kann man als Messung interpretieren, $X(\omega)$ ist der von ω abhängige Messwert.

Außerdem lassen sich Zufallsvariablen verknüpfen, wir können sie addieren, multiplizieren, hintereinanderausführen, ...

Messbarkeit für alle Elemente der Bild- σ -Algebra nachzuweisen, kann sich als äußerst delikat herausstellen. Der folgende Satz besagt jedoch, dass es genügt sich auf ein Erzeugersystem der Bild- σ -Algebra zurückzuziehen.

3.2 Satz Seien (Ω, \mathcal{A}) , (Ω', \mathcal{A}') Messräume, $X : \Omega \rightarrow \Omega'$ und \mathcal{E}' Erzeugersystem von \mathcal{A}' . Dann ist X genau dann messbar, wenn

$$\forall M \in \mathcal{E}' : X^{-1}(M) \in \mathcal{A}, \quad \text{d.h. } X^{-1}(\mathcal{E}') \subseteq \mathcal{A}. \quad \times$$

» \Rightarrow : Gilt *per definitionem*.

\Leftarrow : Sei $X^{-1}(\mathcal{E}') \subseteq \mathcal{A}$, zu zeigen ist nun, dass $X^{-1}(\mathcal{A}') \subseteq \mathcal{A}$. Es gilt

$$X^{-1}(\mathcal{A}') = X^{-1}(\mathcal{F}_{\Omega'}(\mathcal{E}')) = \mathcal{F}_{\Omega}(X^{-1}(\mathcal{E}')) \subseteq \mathcal{A},$$

denn $X^{-1}(\mathcal{E}') \subseteq \mathcal{A}$ und $\mathcal{F}_{\Omega}(X^{-1}(\mathcal{E}'))$ ist die kleinste σ -Algebra, die $X^{-1}(\mathcal{E}')$ enthält und \mathcal{A} ist σ -Algebra. \ll

Ein einfaches aber äußerst nützliches Korollar ergibt sich, wenn wir uns auf reellwertige messbare Abbildungen beschränken.

3.3 Korollar Sei (Ω, \mathcal{A}) ein Messraum und $X : \Omega \rightarrow \overline{\mathbb{R}}$ Abbildung. Dann sind folgende Aussagen äquivalent:

(i) X \mathcal{A} - $\overline{\mathbb{B}}$ -messbar.

(ii) $\forall \alpha \in \mathbb{R} : [X \leq \alpha] \in \mathcal{A}$.

(iii) $\forall \alpha \in \mathbb{R} : [X < \alpha] \in \mathcal{A}$.

(iv) $\forall \alpha \in \mathbb{R} : [X \geq \alpha] \in \mathcal{A}$. \times

» Wird geführt mit Satz 3.2 und der Tatsache, dass $\{(-\infty, \alpha] : \alpha \in \mathbb{R}\}$ ein Erzeugersystem von $\overline{\mathbb{B}}$ ist. \ll

Insbesondere sind somit die Mengen $[X \leq \alpha]$, $[X < \alpha]$, ... für jede reelle Zufallsvariable X messbar.

3.1 **Korollar** Für zwei Abbildungen $X, Y : (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ gilt

$$[X < Y], [X \leq Y], [X = Y], [X \neq Y] \in \mathcal{A},$$

wobei $[X < Y] := \{\omega \in \Omega : X(\omega) < Y(\omega)\}$. \times

» Nach dem Prinzip von Archimedes gilt $[X < Y] = \bigcup_{\alpha \in \mathbb{R}} [X < \alpha < Y]$. Außerdem ist

$$\bigcup_{\alpha \in \mathbb{R}} [X < \alpha < Y] = \bigcup_{\alpha \in \mathbb{Q}} [X < \alpha < Y] = \bigcup_{\alpha \in \mathbb{Q}} [X < \alpha] \cap [Y \leq \alpha]^c.$$

Für jedes $\alpha \in \mathbb{Q}$ ist $[X < \alpha] \cap [Y \leq \alpha]^c$ messbar und die abzählbare Vereinigung messbarer Mengen ist messbar.

Die übrigen Fälle folgen sofort, denn

$$[X \leq Y] = [X > Y]^c \in \mathcal{A},$$

$$[X = Y] = [X \leq Y] \cap [X \geq Y] \in \mathcal{A}$$

$$[X \neq Y] = [X = Y]^c \in \mathcal{A}. \quad \ll$$

3.4 **Satz** Sei $\emptyset \neq A \subset \mathbb{R}^m$ und $A \cap \mathcal{B}_m := \{A \cap B : B \in \mathcal{B}_m\}$. Jede stetige Abbildung $g : A \rightarrow \mathbb{R}^n$ ist $A \cap \mathcal{B}_m$ - \mathcal{B}_n -messbar. \times

» \mathcal{O}_n , das System der offenen Mengen auf \mathbb{R}^n , ist ein Erzeugersystem von \mathcal{B}_n . Nach 3.2 genügt es zu zeigen, dass

$$g^{-1}(\mathcal{O}_n) \subseteq A \cap \mathcal{B}_m.$$

g ist stetig, daher sind die Urbilder offener Mengen offene Mengen $\cap A$. \ll

Für Abbildungen $\mathbb{R}^m \rightarrow \mathbb{R}^n$ ist somit Messbarkeit eine *wesentlich* schwächere Voraussetzung als Stetigkeit.

BSP 2 Die **Dirichletfunktion** $f : \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch,

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & x \in \mathbb{R} \setminus \mathbb{Q}, \end{cases}$$

ist messbar aber sicher *nicht* stetig. \blacksquare

3.5 **Satz** Seien $X : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$, $Y : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_3, \mathcal{A}_3)$. Die zusammengesetzte Abbildung $Y \circ X : \Omega_1 \rightarrow \Omega_3$ ist dann \mathcal{A}_1 - \mathcal{A}_3 -messbar. \times

» Offensichtlich gilt $(Y \circ X)^{-1}(\mathcal{A}_3) = X^{-1} \circ Y^{-1}(\mathcal{A}_3) \subseteq \mathcal{A}_1$. \ll

3.2 **Korollar** Für $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^m, \mathcal{B}_m)$ und eine stetige Abbildung $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ist $g \circ X : \Omega \rightarrow \mathbb{R}^n$ \mathcal{A} - \mathcal{B}_n -messbar. \times

Insbesondere ist die Komposition von Zufallsvariablen bzw. von Zufallsvariablen mit stetigen Funktionen wieder eine Zufallsvariable. Für eine reelle Zufallsvariable X sind somit auch $|X|$, X^2 , $\sqrt{|X|}$, ... Zufallsvariablen.

3.6 **Satz** Sei (Ω, \mathcal{A}) Messraum.

a) Seien $X_n : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ für $n = 1, \dots, m$. Dann ist

$$Y : \Omega \rightarrow \mathbb{R}^m, \omega \mapsto Y(\omega) = (X_1(\omega), \dots, X_m(\omega)), \quad \omega \in \Omega,$$

\mathcal{A} - \mathcal{B}_m -messbar. Für $g : (\mathbb{R}^m, \mathcal{B}_m) \rightarrow (\mathbb{R}, \mathcal{B})$ ist

$$g \circ Y : \Omega \rightarrow \mathbb{R}$$

\mathcal{A} - \mathcal{B} -messbar.

b) Seien $X_{1,2} : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$. Dann sind auch die Abbildungen

(a) $\alpha X_1 + \beta X_2 \quad (\alpha, \beta \in \mathbb{R})'$,

(b) $X_1 X_2$,

(c) $\frac{X_1}{X_2}$ (falls existent)

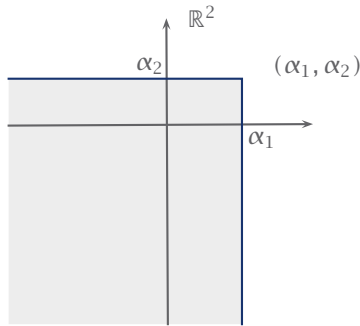
\mathcal{A} - \mathcal{B} -messbar. \times

» a) Ein Erzeugersystem \mathcal{C} von \mathcal{B}_m ist gegeben durch

$$(-\infty, \alpha_1] \times \dots \times (-\infty, \alpha_m],$$

wobei $(\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$. Es genügt zu zeigen, dass $Y^{-1}(\mathcal{C}) \subseteq \mathcal{A}$. Nun gilt

$$\begin{aligned} Y^{-1}((-\infty, \alpha_1] \times \dots \times (-\infty, \alpha_m]) \\ &= \{\omega \in \Omega : x_1(\omega) \leq \alpha_1, \dots, x_m(\omega) \leq \alpha_m\} \\ &= [X_1 \leq \alpha_1] \cap \dots \cap [X_m \leq \alpha_m] \in \mathcal{A}. \quad \ll \end{aligned}$$



3.1 Erzeuger der \mathcal{B}_n .

- b) Die Messbarkeit folgt mit Korollar 3.2 und der Stetigkeit von Summen- und Produktabbildung $(x, y) \mapsto x + y$, $(x, y) \mapsto x \cdot y$. «

Eine Abbildung $X : \Omega \rightarrow \mathbb{R}^n$ ist also genau dann messbar, wenn jede Komponente $X_n : \Omega \rightarrow \mathbb{R}$ messbar ist. Insbesondere sind Produkte und Summen von Zufallsvariablen stets messbar.

3.7 **Satz** Sei (Ω, \mathcal{A}) Messraum, $X_n : (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ für $n = 1, 2, \dots$. Dann sind

- (a) $\inf_n X_n$,
- (b) $\sup_n X_n$,
- (c) $\limsup_n X_n$,
- (d) $\liminf_n X_n$,
- (e) $\lim_n X_n$ (falls existent)

\mathcal{A} - $\overline{\mathcal{B}}$ -messbar. \times

- » (1) $\forall \alpha \in \mathbb{R} : [\inf_n X_n < \alpha] = \bigcup_{n \in \mathbb{N}} [X_n < \alpha] \in \mathcal{A}$, d.h. $\inf_n X_n$ ist \mathcal{A} - $\overline{\mathcal{B}}$ -messbar.
- (2) $\sup_n X_n = -\inf_n (-X_n)$ ist \mathcal{A} - $\overline{\mathcal{B}}$ -messbar.
- (3) $\limsup_n X_n = \inf_n \sup_{k \geq n} X_k$ ist \mathcal{A} - $\overline{\mathcal{B}}$ -messbar.
- (4) $\liminf_n X_n = \sup_n \inf_{k \geq n} X_k$ ist \mathcal{A} - $\overline{\mathcal{B}}$ -messbar.

(5) $\lim X_n = \limsup X_n$, falls $\lim_n X_n$ existiert. «

Da Messbarkeit eine so schwache Voraussetzung ist, ist sie auch ein sehr stabiler Begriff, da sie auch unter Grenzwertbildung und Komposition erhalten bleibt. Probleme treten erst bei überabzählbar vielen Operationen auf.

3-B Bildmaße und Verteilungen

In diesem Abschnitt werden wir eine Beziehung zwischen Zufallsvariablen und Maßen herstellen. Jeder Zufallsvariablen

$$X : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$$

lässt sich eindeutig ein Maß

$$P_X : \mathcal{A}' \rightarrow \mathbb{R}, \quad A' \mapsto P_X(A') := P(X^{-1}(A'))$$

das sogenannte **Bildmaß** zuordnen. Die Verteilungsfunktion des Bildmaßes

$$F(t) = P_X((-\infty, t])$$

nennen wir auch **Verteilungsfunktion von X** . F lässt sich oft leichter handhaben als X , und aus den Eigenschaften von F lassen sich Rückschlüsse auf X machen.

Analog lässt sich zu jedem Maß

$$\mu : \mathcal{A}' \rightarrow \mathbb{R}$$

über Ω' eine Zufallsvariable

$$Y : (\Omega, \mathcal{A}, Q) \rightarrow (\Omega', \mathcal{A}')$$

finden, so dass $P_Y = \mu$.

Zunächst ist natürlich zu klären, ob das Bildmaß überhaupt ein Maß ist.

3.8 **Satz** Seien (Ω, \mathcal{A}) und (Ω', \mathcal{A}') Messräume und $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ eine Abbildung. Sei μ ein Maß auf \mathcal{A} . Durch

$$\begin{aligned}\mu_X(A') &:= \mu(X^{-1}(A')) \\ &= \mu(\{\omega \in \Omega : X(\omega) \in A'\}) =: \mu[X \in A']; \quad A' \in \mathcal{A}'\end{aligned}$$

wird ein Maß (das sogenannte **Bildmaß**) μ_X auf \mathcal{A}' definiert.

Ist μ ein W-Maß auf \mathcal{A} , dann ist μ_X ebenfalls ein W-Maß auf \mathcal{A}' . \times

» (i) $\mu_X \geq 0$ ist klar.

(ii) $\mu_X(\emptyset) = \mu(X^{-1}(\emptyset)) = \mu(\emptyset) = 0$.

(iii) Seien $A'_1, A'_2, \dots \in \mathcal{A}'$ paarweise disjunkt, so gilt

$$\begin{aligned}\mu_X\left(\sum_{i \geq 1} A'_i\right) &= \mu\left(X^{-1}\left(\sum_{i \geq 1} A'_i\right)\right) = \mu\left(\sum_{i \geq 1} (X^{-1}(A'_i))\right) \\ &= \sum_{i \geq 1} \mu(X^{-1}(A'_i)).\end{aligned}$$

(iv) Sei μ ein Wahrscheinlichkeitsmaß so ist

$$\mu_X(\Omega') = \mu(X^{-1}(\Omega')) = \mu(\Omega) = 1. \quad \ll$$

Es genügt also, dass X messbar ist, damit P_X tatsächlich ein Maß ist. Insbesondere ist für jede Zufallsvariable P_X ein Maß.

3.4 **Definition** Sei X eine (Ω', \mathcal{A}') -Zufallsvariable auf dem W-Raum (Ω, \mathcal{A}, P) . Das W-Maß P_X im Bild-W-Raum $(\Omega', \mathcal{A}', P_X)$ heißt **Verteilung** der Zufallsvariable X .

Sprechweise:

- Die Zufallsvariable X liegt in $A' \in \mathcal{A}'$.
- X nimmt Werte in $A' \in \mathcal{A}'$ an mit Wahrscheinlichkeit $P_X(A') = P[X \in A']$.
- Wenn $P[X \in A'] = 1$, sagt man X liegt P -fast sicher (P -f.s.) in A' . \times

Betrachten wir die Verteilung P_X einer reellen Zufallsvariablen X , so besitzt diese eine Verteilungsfunktion

$$F(t) = P_X((-\infty, t]) = P(X^{-1}(-\infty, t]).$$

Da es sich bei der Verteilungsfunktion um eine reelle Funktion handelt, kann man oft viel leichter mit ihr Rechnen, als mit der Zufallsvariablen selbst.

Die Eigenschaften der Verteilungsfunktion charakterisieren die Zufallsvariable, man klassifiziert Zufallsvariablen daher nach Verteilung (binomial-, poisson-, exponentialverteilt, ...).

Definition Besitzen zwei Zufallsvariablen dieselbe Verteilung (also dieselbe Verteilungsfunktion) heißen sie *gleichverteilt*. \times

BSP 3 Wir betrachten n Bernoulli-Versuche mit jeweiliger Erfolgswahrscheinlichkeit p . Ω besteht aus der Menge der Elementarereignisse ω (= n -Tupel aus Nullen und Einsen), $\mathcal{A} = \mathcal{P}(\Omega)$.

Das Wahrscheinlichkeitsmaß P auf \mathcal{A} ist gegeben durch

$$P(\{\omega\}) = p^k(1-p)^{n-k},$$

falls ω aus k Einsen und $n-k$ Nullen besteht.

Wir definieren uns eine Zufallsvariable X durch:

$$X: \Omega \rightarrow \mathbb{R}, \quad X(\omega) = \text{Anzahl der Einsen in } \omega.$$

X ist \mathcal{A} - \mathcal{B} -messbar, denn $X^{-1}(B) \subseteq \mathcal{P}(\Omega) = \mathcal{A}$.

$X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ gibt also die zufällige Anzahl der Erfolge in n Bernoulli-Versuchen an.

Das Bildmaß zu X ist gegeben durch

$$P_X(\{k\}) = P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} =: b(n, p; k).$$

Um das Bildmaß auf allgemeine Mengen in \mathcal{B} fortzusetzen, setzen wir zu $B \in \mathcal{B}$

$$P_X(B) = \sum_{k \in \mathbb{N}_0 \cap B} b(n, p; k).$$

P_X ist die sogenannte **Binominalverteilung** $b(n, p)$. X ist $b(n, p)$ -verteilt. ■

- 3.2 *Bemerkung.* Seien (Ω, \mathcal{A}, P) ein W-Raum und (Ω', \mathcal{A}') , $(\Omega'', \mathcal{A}'')$ Messräume. Seien $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ und $Y : (\Omega', \mathcal{A}') \rightarrow (\Omega'', \mathcal{A}'')$, dann gilt

$$P_{Y \circ X} = (P_X)_Y. \quad \rightarrow$$

» $Y \circ X$ ist wieder messbar. Für $A'' \in \mathcal{A}''$ gilt somit,

$$\begin{aligned} P_{Y \circ X}(A'') &= P((Y \circ X)^{-1}(A'')) = P(X^{-1}(Y^{-1}(A''))) \\ &= P_X(Y^{-1}(A'')) = (P_X)_Y(A''). \quad \leftarrow \end{aligned}$$

■ Produktmessräume

Betrachten wir die Vektorräume \mathbb{R}^n und \mathbb{R}^m , so können wir diese durch das karthesische Produkt verknüpfen zu $V = \mathbb{R}^n \times \mathbb{R}^m$. V ist dann wieder ein Vektorraum und jedes Element $v \in V$ lässt sich darstellen als $v = (x, y)$, wobei $x \in \mathbb{R}^n$ und $y \in \mathbb{R}^m$.

Eine solche Verknüpfung lässt sich auch für Messräume definieren.

- 3.5 **Definition** Seien $(\Omega_i, \mathcal{A}_i)$ Messräume für $i = 1, \dots, n$. Die **Produkt- σ -Algebra** $\otimes_{i=1}^n \mathcal{A}_i$ wird definiert als die von dem Mengensystem

$$\left\{ \prod_{i=1}^n A_i : A_i \in \mathcal{A}_i, \quad i = 1, \dots, n \right\}$$

erzeugte σ -Algebra.

$\otimes_{i=1}^n (\Omega_i, \mathcal{A}_i) := \left(\prod_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i \right)$ heißt **Produkt-Messraum**. \times

Man bildet also das karthesische Produkt der Ω_i und die Produkt- σ -Algebra der \mathcal{A}_i und erhält so wieder einen Messraum $\otimes_{i=1}^n (\Omega_i, \mathcal{A}_i)$.

BSP 4 $(\mathbb{R}^n \times \mathbb{R}^m, \mathcal{B}_n \otimes \mathcal{B}_m) = (\mathbb{R}^{n+m}, \mathcal{B}_{n+m})$. ■

- 3.3 *Bemerkung.* Sei $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i)$ messbar für $i = 1, \dots, n$. Die Abbildung

$$X : \Omega \rightarrow \prod_{i=1}^n \Omega_i$$

mit

$$X(\omega) := (X_1(\omega), \dots, X_n(\omega)), \quad \omega \in \Omega$$

ist dann $\mathcal{A} \otimes \otimes_{i=1}^n \mathcal{A}_i$ -messbar. \rightarrow

» Betrachte ein Erzeugendensystem von $\otimes_{i=1}^n A_i$, z.B. die Quader

$$\mathcal{C} = \left\{ \prod_{i=1}^n A_i : A_i \in \mathcal{A}_i \right\}.$$

Der Beweis wird dann wie im eindimensionalen Fall geführt. «

Insbesondere ist eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto (x_1, \dots, x_n)$$

genau dann messbar, wenn es jede Komponente $X_i : \Omega \rightarrow \mathbb{R}$ ist.

3.4 **Bemerkung.** Seien $(\Omega_i, \mathcal{A}_i)$ Messräume für $i = 1, \dots, n$. Ein W-Maß auf $\otimes_{i=1}^n \mathcal{A}_i$ ist eindeutig festgelegt durch seine Werte auf dem Mengensystem

$$\left\{ \prod_{i=1}^n A_i : A_i \in \mathcal{A}_i, i = 1, \dots, n \right\}. \quad \rightarrow$$

» Fortsetzungs- und Eindeutigkeitsatz 1.4 mit der σ -Algebra gegeben durch die endliche Summe von Quadern der Form

$$\prod_{i=1}^n A_i, \quad A_i \in \mathcal{A}_i. \quad \leftarrow$$

3.6 **Definition** Seien $X_i : (\Omega, \mathcal{A}, P) \rightarrow (\Omega_i, \mathcal{A}_i)$ Zufallsvariablen für $i = 1, \dots, n$ und $X := (X_1, \dots, X_n)$ Zufallsvariable auf (Ω, \mathcal{A}, P) .

1.) Die Verteilung P_X der Zufallsvariablen X - erklärt durch

$$P_X(A) := P[(X_1, \dots, X_n) \in A]$$

für $A \in \otimes_{i=1}^n \mathcal{A}_i$ oder auch nur für $A = \prod_{i=1}^n A_i$ ($A_i \in \mathcal{A}_i, i = 1, \dots, n$) - heißt die **gemeinsame Verteilung** der Zufallsvariablen X_i .

2.) Die Verteilungen P_{X_i} - erklärt durch

$$\begin{aligned} P_{X_i}(A_i) &:= P[X_i \in A_i] \\ &= P[X \in \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n] \end{aligned}$$

für $A_i \in \mathcal{A}_i$ - heißen **Randverteilungen** von P_X ($i = 1, \dots, n$). \times

Wichtiger Spezialfall. $(\Omega_i, \mathcal{A}_i) = (\mathbb{R}, \mathcal{B})$ für $i = 1, \dots, n$. Dann sind die X_i reelle Zufallsvariablen und X ein Zufallsvektor. \rightarrow

3.5 *Bemerkungen.* Seien die Bezeichnungen wie in Definition 3.6.

- A. Die Verteilung P_X ist ohne Zusatzvoraussetzung durch ihre Randverteilungen *nicht* eindeutig festgelegt.
- B. Die Projektionsabbildung

$$\pi_i : \prod_{k=1}^n \Omega_k \rightarrow \Omega_i, \quad (\omega_1, \dots, \omega_n) \rightarrow \omega_i$$

ist messbar und somit gilt

$$P_{X_i} = (P_X)\pi_i, \quad i = 1, \dots, n. \quad \rightarrow$$

Aufgrund der Messbarkeit der Projektionen sind die Randverteilungen einer gemeinsamen Verteilung eindeutig bestimmt. Die Umkehrung, dass die Randverteilungen auch die gemeinsame Verteilung eindeutig bestimmen ist im Allgemeinen falsch. Wir werden uns in Kapitel 5 ausführlicher damit beschäftigen.

3.6 *Bemerkungen.* A. Sei Q ein W-Maß auf $(\mathbb{R}^n, \mathcal{B}_n)$. Dann existieren reelle Zufallsvariablen X_1, \dots, X_n auf einem geeigneten W-Raum (Ω, \mathcal{A}, P) so, dass ihre gemeinsame Verteilung mit Q übereinstimmt:

Auf dem W-Raum $(\Omega, \mathcal{A}, P) := (\mathbb{R}^n, \mathcal{B}_n, Q)$, wird $X_i : \Omega \rightarrow \mathbb{R}$ definiert als die Projektionsabbildung

$$(x_1, \dots, x_n) \rightarrow x_i \quad (i = 1, \dots, n);$$

hierbei ist also $X := (X_1, \dots, X_n)$ die auf \mathbb{R}^n definierte identische Abbildung.

- B. Wir können die Aussage aus a) unmittelbar auf einen beliebigen Produkt-Meßraum $\left(\prod_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i \right)$ statt $(\mathbb{R}^n, \mathcal{B}_n)$ verallgemeinern.
- C. Sonderfall zu b): Ist (Ω, \mathcal{A}, Q) ein W-Raum, $X : \Omega \rightarrow \Omega$ die identische Abbildung, so gilt $P_X = Q$. Jedes W-Maß lässt sich somit als eine Verteilung auffassen (und - definitionsgemäß - umgekehrt). \rightarrow

Dieser Zusammenhang zwischen W-Maß und Verteilung hat eine große Bedeutung, wie wir mit der Einführung des Maßintegrals im folgenden Kapitel sehen werden.

4 Erwartungswerte und Dichten

4-A Erwartungswert, Varianz und Lebesgue Integral

Gegeben seien ein W-Raum (Ω, \mathcal{A}, P) und eine reelle Zufallsvariable X auf (Ω, \mathcal{A}, P) mit Verteilung P_X (auf \mathcal{B}) und Verteilungsfunktion $F: \mathbb{R} \rightarrow \mathbb{R}$.

Der Erwartungswert EX der Zufallsvariable X gibt einen “mittleren Wert” von X bezüglich P an.

BSP 1 Wir betrachten einen fairen Würfel. X gebe die zufällige Augenzahl an. Den Erwartungswert von X erhalten wir, indem wir die möglichen Werte von X mit ihrer Wahrscheinlichkeit multiplizieren und aufsummieren,

$$EX = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5.$$

Für **diskrete** Zufallsvariablen, das sind Zufallsvariablen deren Verteilung auf \mathbb{N}_0 konzentriert ist, erhalten wir somit,

$$EX = \sum_{k=1}^{\infty} k p_k, \quad \text{mit } p_k = P[X = k]. \quad \blacksquare$$

Diese Definition lässt sich allerdings nicht auf den Fall einer Zufallsvariablen mit einer auf einem Kontinuum konzentrierten Verteilung übertragen. Um diesen Fall einzuschließen, gehen wir von der diskreten Summe zum kontinuierlichen Integral über, wir wählen also Integralansatz, um den Erwartungswert einzuführen.

■ Erwartungswert mittels Riemann-Stieltjes-Integral

Bezeichnungen.

$$X^+(\omega) := \begin{cases} X(\omega), & \text{falls } X(\omega) > 0, \\ 0, & \text{sonst} \end{cases}$$

$$X^-(\omega) := \begin{cases} -X(\omega), & \text{falls } X(\omega) < 0, \\ 0, & \text{sonst.} \end{cases}$$

Wir wählen zunächst den naiven Ansatz, den Erwartungswert von X als Integral über den Wertebereich von X zu definieren.

1. *Schritt*: Sei zunächst X positiv ($X \geq 0$), d.h. $P_X(\mathbb{R}_+) = 1$ und $F(x) = 0$ für $x < 0$. Wir ersetzen nun die Summe aus dem Beispiel durch ein Integral,

$$\mathbf{E}X := \int_{\mathbb{R}_+} x P_X(dx) := \int_{\mathbb{R}_+} x dF(x) := \lim_{a \rightarrow \infty} \int_{[0,a]} x dF(x),$$

wobei es sich hier um ein *Riemann-Stieltjes-Integral* von x bezüglich F handelt. Da der Integrand x auf \mathbb{R}_+ positiv, gilt $0 \leq \mathbf{E}X \leq \infty$.

Für $F(x) = x$ entspricht das Riemann-Stieltjes-Integral dem gewöhnlichen Riemann-Integral,

$$\int_{\mathbb{R}_+} x dF(x) = \int_0^\infty x dx.$$

Allgemeiner entspricht das Riemann-Stieltjes-Integral für stetig differenzierbares F dem Riemann-Integral,

$$\int_{\mathbb{R}_+} x dF(x) = \int_0^\infty x F'(x) dx.$$

2. *Schritt*: Sei X beliebig und $\mathbf{E}X^+$, $\mathbf{E}X^-$ nicht beide ∞ , so ist der Erwartungswert von X definiert als

$$\mathbf{E}X := \int_{\mathbb{R}} x P_X(dx) := \mathbf{E}X^+ - \mathbf{E}X^-. \quad \times$$

4.1 **Erster Versuch** des *Erwartungswertes* $\mathbf{E}X$ von X als ein Integral auf dem Wertebereich von X :

$$\mathbf{E}X := \int_{\mathbb{R}} x P_X(dx). \quad \times$$

Für den Erwartungswert ist diese Definition ausreichend und historisch ist man auch genau so vorgegangen. Wir integrieren hier jedoch lediglich die stetige Funktion $x \mapsto x$. In der Wahrscheinlichkeitstheorie arbeitet man aber meist mit unstetigen Funktionen und für diese ist das Riemann-Stieltjes-Integral *nicht* ausreichend.

■ Erwartungswert mittels Maß-Integral

Wir benötigen einen allgemeineren Integrationsbegriff, der es uns erlaubt, lediglich messbare (also insbesondere auch unstetige) Funktionen zu integrieren.

Dazu machen wir einen neuen Integralansatz, wobei wir diesmal über den Definitionsbereich von X also Ω integrieren.

Sei also $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ messbar.

o. Schritt: Sei X positiv und nehme nur endlich viele Werte an. Dann existiert eine Darstellung

$$X = \sum_{i=1}^N \alpha_i \mathbf{1}_{A_i}, \quad \alpha_i \in \mathbb{R}_+, \quad A_i \in \mathcal{A}$$

mit A_i paarweise disjunkt und $\sum_{i=1}^N A_i = \Omega$.

$\mathbf{1}_A$ bezeichnet die sogenannte **Indikatorfunktion** einer Menge A ,

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

Somit können wir den Erwartungswert definieren als

$$EX := \int_{\Omega} X \, dP := \sum_{i=1}^N \alpha_i P(A_i).$$

Diese Definition schließt z.B. den Fall des Würfels ein, jedoch sind wir beispielsweise noch nicht in der Lage, den Erwartungswert eines zufälligen Temperaturwertes anzugeben.

1. Schritt: Sei nun X lediglich positiv. In der Maßtheorie wird gezeigt, dass dann eine Folge von Zufallsvariablen $X_n \geq 0$ existiert, wobei X_n jeweils nur endlich viele Werte annimmt und die Folge (X_n) monoton gegen X konvergiert, $X_n(\omega) \uparrow X(\omega)$ für $n \rightarrow \infty$ und $\omega \in \Omega$.

Somit können wir den Erwartungswert von X definieren als

$$EX := \int_{\Omega} X \, dP := \lim_{n \rightarrow \infty} EX_n,$$

wobei der (uneigentliche) Grenzwert existiert, da die EX_n monoton wachsen.

2. Schritt: Seien EX^+ , EX^- nicht beide ∞ .

$$EX := \int_{\Omega} X \, dP := EX^+ - EX^-.$$

Somit erhalten wir die endgültige Definition des Erwartungswerts.

4.2 **Definition** des Erwartungswertes EX von X als ein Integral auf dem Definitionsbereich von X :

$$EX := \int_{\Omega} X(\omega)P(d\omega) =: \int_{\Omega} X dP. \quad \times$$

4.1 **Bemerkungen zu Definition 4.2** A. Die einzelnen Schritte - insbesondere bei der 2. Definition über das Maß-Integral - sind sinnvoll. Für Interessierte werden die Details in der Maßtheorie, einem Teilgebiet der Analysis, behandelt.

B. Die beiden Definitionen sind äquivalent.

C. Existiert das Integral

$$\int_{\mathbb{R}} x P_X(dx)$$

in 1. Definition, so ändert es seinen Wert nicht, wenn man es gemäß 2. Definition erklärt. Insbesondere gilt dann,

$$\int_{\Omega} X dP = \int_{\mathbb{R}} x P_X(dx) = \int_{\mathbb{R}} x dF(x).$$

D. Der Begriff $\int_{\Omega} X dP$ in Definition 4.2 lässt sich unmittelbar verallgemeinern auf den Fall eines Maßraumes $(\Omega, \mathcal{A}, \mu)$ (anstatt eines W-Raumes) und liefert

$$\int_{\Omega} X(\omega) \mu(d\omega) =: \int_{\Omega} X d\mu.$$

Wichtige Spezialfälle:

1.) Sei $(\Omega, \mathcal{A}) = (\mathbb{R}^n, \mathcal{B}_n)$ oder etwas allgemeiner $\Omega = B$ mit $B \in \mathcal{B}_n$, $\mathcal{A} = B \cap \mathcal{B}_n$ und $\mu = \lambda|_{\mathcal{A}}$ Restriktion des L-B-Maß(es) auf \mathcal{A} .

$$\int_{\mathbb{R}^n} X d\lambda \quad \text{bzw.} \quad \int_B X d\mu$$

wird als **Lebesgue-Integral** bezeichnet. Im Falle $n = 1$ schreiben wir auch

$$\int_{\mathbb{R}} X(x) dx \quad \text{bzw.} \quad \int_B X(x) dx.$$

- 2.) Sei $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$ und $H : \mathbb{R} \rightarrow \mathbb{R}$ eine maßdefinierende Funktion mit zugehörigem Maß μ , d.h.

$$H(b) - H(a) = \mu((a, b]), \quad -\infty < a < b < \infty,$$

so schreiben wir

$$\int_{\mathbb{R}} X(x) dH(x) := \int_{\mathbb{R}} X(x) H'(dx) := \int_{\mathbb{R}} X d\mu$$

Die Verallgemeinerung auf $\Omega = B \in \mathcal{B}$ folgt analog.

- 3.) Sei $(\Omega, \mathcal{A}) = (\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ und μ das Zählmaß, d.h.

$$\mu(A) = \text{Anzahl der Elemente in } A,$$

so gilt für eine Funktion $f : \mathbb{N}_0 \rightarrow \mathbb{R}$,

$$\int_{\mathbb{N}_0} f d\mu = \sum_{k=0}^{\infty} f(k).$$

Reihen sind also lediglich ein Spezialfall des Maß-Integrals.

- E. Sei X eine erweitert-reelle Zufallsvariable auf einem W-Raum (Ω, \mathcal{A}, P) mit $P[|X| = \infty] = 0$ und

$$Y(\omega) := \begin{cases} X(\omega), & \text{falls } |X(\omega)| < \infty, \omega \in \Omega, \\ 0 & \text{sonst,} \end{cases}$$

so definiert man $EX := EY$, falls EX existiert. \rightarrow

Für das Maß- bzw. spezieller das Lebesgue-Integral existieren zahlreiche sehr allgemeine Sätze und elegante Beweisstrategien, die uns für das Riemann-Stieltjes-Integral nicht zur Verfügung stehen, weshalb wir im Folgenden fast ausschließlich von dieser Definition ausgehen werden.

Wir haben bisher nicht geklärt, wie man ein Maß- bzw. ein Lebesgue-Integral konkret berechnet, wenn X nicht nur endlich viele Werte annimmt. Ist die Verteilungsfunktion stetig differenzierbar, so gilt

$$\int_{\Omega} X dP = \int_{\mathbb{R}} x F'(x) dx,$$

wobei wir letzteres Integral durchaus als Lebesgue-Integral mit all seinen angenehmen Eigenschaften auffassen können, wir können damit aber auch rechnen wie mit dem Riemann-Integral (Substitution, partielle Integration, ...). Für allgemeineres F müssen wir uns darauf beschränken, abstrakt mit dem Integral arbeiten zu können.

4.2 **Bemerkung.** Sei (Ω, \mathcal{A}, P) ein W-Raum. Für $A \in \mathcal{A}$ gilt $P(A) = \mathbf{E} \mathbf{1}_A$. \rightarrow

4.3 **Definition** Existiert für die reelle Zufallsvariable X auf dem W-Raum (Ω, \mathcal{A}, P) ein endlicher Erwartungswert $\mathbf{E}X$, so heißt X **integrierbar** (bezüglich P). Analog für $\int_{\Omega} X \, d\mu$ in Bemerkung 4.1 D.. \times

Ist X eine positive Zufallsvariable, so können wir den Erwartungswert direkt mit Hilfe der Verteilungsfunktion berechnen.

4.1 **Lemma** Für eine reelle Zufallsvariable $X \geq 0$ mit Verteilungsfunktion F gilt

$$\mathbf{E}X = \int_0^{\infty} (1 - F(x)) \, dx. \quad \times$$

» Wir verwenden Definition 4.1 mithilfe des R-S-Integrals,

$$\begin{aligned} \mathbf{E}X &= \int_{\mathbb{R}_+} x \, dF(x) = \lim_{a \rightarrow \infty} \int_{[0, a]} x \, dF(x) \\ &\stackrel{\text{part.int.}}{=} \lim_{a \rightarrow \infty} \left(x F(x) \Big|_{x=0}^a - \int_0^a 1 F(x) \, dx \right) \stackrel{(*)}{=} \lim_{a \rightarrow \infty} \left(a - \int_0^a F(x) \, dx \right) \\ &= \lim_{a \rightarrow \infty} \left(\int_0^a (1 - F(x)) \, dx \right) = \int_0^{\infty} (1 - F(x)) \, dx. \end{aligned}$$

1.) Zu (*). Sei $\mathbf{E}X < \infty$

$$\begin{aligned} \infty > \mathbf{E}X &= \int_{\mathbb{R}_+} x \, dF(x) \geq \int_{[a, \infty]} x \, dF(x) \geq \int_{[a, \infty]} a \, dF(x) \\ &\geq a \int_{(a, \infty)} dF(x) = a(1 - F(a)) \geq 0. \end{aligned}$$

2.) Außerdem:

$$\int_{[a, \infty]} x \, dF(x) \xrightarrow{a \rightarrow \infty} 0,$$

da $\int_{[0, a]} x \, dF(x) \rightarrow \int_{\mathbb{R}_+} x \, dF(x)$, also

$$a(1 - F(a)) \xrightarrow{a \rightarrow \infty} 0.$$

Der Fall $\mathbf{E}X = \infty$ wird gesonderd behandelt. \leftarrow

■ Eigenschaften des Erwartungswerts

Der Erwartungswert einer reellen Zufallsvariablen X ,

$$EX = \int_{\Omega} X dP = \int_{\mathbb{R}} x dP_X,$$

ist ein spezielles Maß-Integral. Daher übertragen sich alle Eigenschaften dieses Integrals auf den Erwartungswert.

4.1 **Satz** Sei X eine reelle Zufallsvariable auf einem W -Raum (Ω, \mathcal{A}, P) .

a) X integrierbar $\Leftrightarrow X^+$ und X^- integrierbar $\Leftrightarrow |X|$ integrierbar.

b) Existiert eine reelle integrierbare Zufallsvariable $Y \geq 0$ mit

$$|X| \leq Y \text{ P-f.s., so ist } X \text{ integrierbar.}$$

c) Ist X integrierbar und existiert eine reelle Zufallsvariable Y mit $Y = X$ P-f.s., dann existiert $EY = EX$. \times

» Der Beweis wird in der Maßtheorie geführt. «

Im Gegensatz zum klassischen uneigentlichen Riemann-Integral ist es beim Maß-Integral nicht möglich, dass sich positive und negative Anteile gegenseitig eliminieren und so eine Funktion deren Positiv- oder Negativanteil alleine nicht integrierbar sind, als ganzes integrierbar wird. Es gibt also durchaus Funktionen die (uneigentlich) Riemann- aber nicht Lebesgue-integrierbar sind. Für "viel mehr Funktionen" gilt jedoch das Gegenteil.

Beim Riemann-Integral ändert eine Abänderung einer Funktion an endlich vielen Punkten den Integralwert nicht, beim Lebesgue-Integral hingegen ändert eine Abänderung einer Zufallsvariablen auf einer Menge vom Maß Null ihren Erwartungswert nicht.

BSP 2 Die Dirichletfunktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

ist *nicht* Riemann- aber Lebesgue-integrierbar. Das Lebesgue-Integral lässt sich auch sehr leicht berechnen, denn $f \geq 0$, also gilt nach Definition

$$\int_{\mathbb{R}} f(x) dx = 1 \cdot \lambda(\mathbb{Q}) + 0 \cdot \lambda(\mathbb{R} \setminus \mathbb{Q}) = 1 \cdot \underbrace{\lambda(\mathbb{Q})}_{=0},$$

denn \mathbb{Q} ist abzählbare Teilmenge von \mathbb{R} . ■

Ferner ist der Erwartungswert linear, monoton und erfüllt die Dreiecksungleichung.

4.2 **Satz** Seien X, Y reelle integrierbare Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) .

a) Es existiert $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$.

b) Es existiert $\mathbf{E}(\alpha X) = \alpha \mathbf{E}X$ für $\alpha \in \mathbb{R}$.

c) $X \geq Y \Rightarrow \mathbf{E}X \geq \mathbf{E}Y$.

d) $|\mathbf{E}X| \leq \mathbf{E}|X|$.

e) $X \geq 0, \mathbf{E}X = 0 \Rightarrow X = 0$ P-f.s. ✕

» *Beweisidee.* Wir beweisen lediglich die erste Behauptung, der Rest folgt analog. Dazu bedienen wir uns dem *Standardtrick* für Beweise in der Maßtheorie.

1. *Schritt.* Reduktion auf $X \geq 0$ und $Y \geq 0$, dann Reduktion auf einfache Funktionen

$$X = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}, \quad Y = \sum_{j=1}^m \beta_j \mathbf{1}_{B_j}$$

mit $\sum_{i=1}^n A_i = \Omega = \sum_{j=1}^m B_j$. Sei $C_{ij} = A_i \cap B_j \in \mathcal{A}$, so gilt

$$\begin{aligned} X &= \sum_{i,j} \gamma_{ij} \mathbf{1}_{C_{ij}}, & Y &= \sum_{i,j} \delta_{ij} \mathbf{1}_{C_{ij}}, \\ \Rightarrow X + Y &= \sum_{i,j} (\gamma_{ij} + \delta_{ij}) \mathbf{1}_{C_{ij}}. \end{aligned}$$

2. *Schritt.* Nun seien $X \geq 0$ und $Y \geq 0$ beliebig und X_n, Y_n Folgen von einfachen Funktionen mit $X_n \uparrow X, Y_n \uparrow Y$ und $X_n + Y_n = Z_n \uparrow Z = X + Y$.

Für jedes $n \in \mathbb{N}$ sind X_n und Y_n einfach und damit \mathbf{E} linear. Es gilt also

$$\mathbf{E}(X_n + Y_n) = \mathbf{E}X_n + \mathbf{E}Y_n \leq \mathbf{E}X + \mathbf{E}Y$$

sowie

$$\mathbf{E}X_n + \mathbf{E}Y_n = \mathbf{E}(X_n + Y_n) = \mathbf{E}Z_n \leq \mathbf{E}Z = \mathbf{E}(X + Y).$$

Nach dem Grenzübergang für $n \rightarrow \infty$ erhalten wir somit

$$\mathbf{E}X + \mathbf{E}Y = \mathbf{E}(X + Y).$$

3. *Schritt.* Für X und Y integrierbar betrachten wir

$$X = X_+ - X_-, \quad Y = Y_+ - Y_-.$$

Nach dem eben gezeigten, können wir die Linearität von \mathbf{E} für Positiv und Negativteil verwenden.

$$\begin{aligned} \mathbf{E}X + \mathbf{E}Y &= (\mathbf{E}X_+ + \mathbf{E}Y_+) - (\mathbf{E}X_- + \mathbf{E}Y_-) = \mathbf{E}(X_+ + Y_+) - \mathbf{E}(X_- + Y_-) \\ &= \mathbf{E}(X + Y). \quad \ll \end{aligned}$$

BSP 3 Wir betrachten n Bernoulli-Versuche mit jeweiliger Erfolgswahrscheinlichkeit p . Die Zufallsvariable X_i nehme Werte 0 bzw. 1 an, falls im i -ten Versuch ein Misserfolg bzw. Erfolg aufgetreten ist ($i = 1, \dots, n$).

$X = X_1 + \dots + X_n$ gibt die Anzahl der Erfolge an.

$$\mathbf{E}X = \mathbf{E}X_1 + \dots + \mathbf{E}X_n \stackrel{!}{=} n\mathbf{E}X_1 = np,$$

da alle X_i dieselbe Verteilungsfunktion besitzen.

Der Erwartungswert einer $b(n, p)$ -verteilten Zufallsvariablen ist also np . Ein alternativer (evtl. ungeschickterer) Rechenweg ist

$$\mathbf{E}X = \sum_{k=0}^n kP[X = k] = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \dots = np. \quad \blacksquare$$

Für das Riemann-Integral ist für die Vertauschbarkeit von Integration und Grenzwertbildung

$$\lim_{n \rightarrow \infty} \int_{[a,b]} f_n(x) \, dx = \int_{[a,b]} \lim_{n \rightarrow \infty} f_n(x) \, dx$$

im Allgemeinen die gleichmäßige Konvergenz der f_n erforderlich. Ein entscheidender Vorteil des Lebesgue-Integrals ist die Existenz von Konvergenzsätzen, die wesentlich schwächere Voraussetzungen für die Vertauschbarkeit haben.

- 4.3 **Satz von der monotonen Konvergenz (B. Levi)** Für reelle Zufallsvariablen X_n auf (Ω, \mathcal{A}, P) mit $X_n \geq 0$ ($n \in \mathbb{N}$), $X_n \uparrow X$ ($n \rightarrow \infty$) existiert

$$EX = \lim_n EX_n.$$

Hierbei $EX := \infty$, falls $P[X = \infty] > 0$. Entsprechend für Reihen von nichtnegativen Zufallsvariablen. - Analog für $\int_{\Omega} X_n d\mu$ gemäß Bemerkung 4.1.d. \times

» Die Messbarkeit von X folgt aus Satz 3.7. Die X_n sind monoton wachsend und positiv, also ist $(\int_{\Omega} X_n dP)_n$ monotone Folge, somit existiert ihr Grenzwert. Setze

$$c := \lim_{n \rightarrow \infty} \int_{\Omega} X_n dP \leq \int_{\Omega} X dP.$$

Es genügt nun für jede einfache Funktion Y mit $Y \leq X$ zu zeigen, dass

$$c \geq \int_{\Omega} Y dP.$$

Sei dazu

$$Y = \sum_{i=1}^m \alpha_i \mathbf{1}_{[Y=\alpha_i]}, \quad \alpha_i \geq 0.$$

Sei $\delta < 1$ beliebig aber fest und

$$A_n := [X_n \geq \delta Y].$$

Wegen $X_n \uparrow X$ folgt $A_n \uparrow \Omega$ also auch $A_n \cap [Y = \alpha_j] \uparrow [Y = \alpha_j]$ für $j = 1, \dots, m$.

$$\begin{aligned} \delta \int_{\Omega} Y dP &= \delta \sum_{j=1}^m \alpha_j P[Y = \alpha_j] = \delta \sum_{j=1}^m \alpha_j \lim_{n \rightarrow \infty} P(A_n \cap [Y = \alpha_j]) \\ &= \lim_{n \rightarrow \infty} \sum_{j=1}^m \delta \alpha_j P(A_n \cap [Y = \alpha_j]) = \lim_{n \rightarrow \infty} \int_{\Omega} \underbrace{\delta Y \mathbf{1}_{A_n}}_{\leq X_n} dP \\ &\leq \lim_{n \rightarrow \infty} \int_{\Omega} X_n dP = c. \end{aligned}$$

Da die Ungleichung für jedes $\delta < 1$ gilt, folgt

$$\int_{\Omega} Y dP \leq c. \quad \llcorner$$

4-B Dichtefunktion und Zähldichte

4.4 **Definition** Sei X ein n -dimensionaler Zufallsvektor mit Verteilung P_X auf \mathcal{B}_n und Verteilungsfunktion $F : \mathbb{R}^n \rightarrow \mathbb{R}$.

1.) Nimmt X (eventuell nach Vernachlässigung einer P -Nullmenge in Ω) höchstens abzählbar viele Werte an, so ist P_X eine sogenannte **diskrete W -Verteilung**.

Ist X eine reelle Zufallsvariable und P_X auf \mathbb{N}_0 konzentriert, so heißt die Folge $(p_k)_k$ mit

$$p_k := P_X(\{k\}) = P[X = k], \quad k \in \mathbb{N}_0,$$

(wobei $\sum_{k=0}^{\infty} p_k = 1$) **Zähldichte** von X bzw. P_X .

2.) Gibt es eine Funktion

$$f : (\mathbb{R}^n, \mathcal{B}_n) \rightarrow (\mathbb{R}_+, \mathcal{B}_+)$$

für die das Lebesgue-Integral

$$\int_{\prod_{i=1}^n (-\infty, x_i]} f \, d\lambda$$

existiert und mit

$$P_X \left(\prod_{i=1}^n (-\infty, x_i] \right) = F(x_1, \dots, x_n), \quad (x_1, \dots, x_n) \in \mathbb{R}^n$$

übereinstimmt, so heißt f **Dichte(funktion)** von X bzw. P_X bzw. F .

P_X und F heißen **totalstetig**, falls sie eine Dichtefunktion besitzen. \times

4.3 **Bemerkungen.** A. Eine totalstetige Verteilungsfunktion ist stetig.

B. Besitzt die n -dimensionale Verteilungsfunktion

$$F : \mathbb{R}^n \rightarrow \mathbb{R}$$

eine Dichtefunktion

$$f : \mathbb{R}^n \rightarrow \mathbb{R}_+,$$

so existiert L-f.ü.

$$\frac{\partial^n F}{\partial x_1 \dots \partial x_n},$$

und L-f.ü., d.h. insbesondere an den Stetigkeitsstellen von f , gilt

$$\frac{\partial^n F}{\partial x_1 \dots \partial x_n} = f.$$

- c. Ein uneigentliches Riemann-Integral einer nichtnegativen \mathcal{B}_n - \mathcal{B} -messbaren Funktion lässt sich als Lebesgue-Integral deuten. \rightarrow

Sind Dichtefunktion bzw. Zähldichte bekannt, so können wir den abstrakten Ausdruck P_X durch ein Integral über eine konkrete Funktion bzw. durch eine Reihe ersetzen,

$$P_X(A) = \int_A f \, d\lambda \quad \text{bzw.} \quad \sum_{k \in \mathbb{N}_0 \cap A} p_k.$$

Zufallsvariablen mit Dichtefunktion bzw. Zähldichte, sind für uns somit sehr "zugänglich". Es lassen sich jedoch nicht für alle Zufallsvariablen Dichten finden, denn dazu müsste jede Verteilungsfunktion L-f.ü. differenzierbar und Differentiation mit der Integration über \mathbb{R}^n vertauschbar sein.

4.4 **Satz** Sei X eine reelle Zufallsvariable mit Verteilung P_X .

a) Ist P_X auf \mathbb{N}_0 konzentriert mit Zähldichte (p_k) , dann gilt

$$\mathbf{E}X = \sum_{k=1}^{\infty} k p_k.$$

b) Besitzt X eine Dichtefunktion $f : \mathbb{R} \rightarrow \mathbb{R}_+$, so existiert das Lebesgue-Integral

$$\int_{\mathbb{R}} x f(x) \, dx =: \int_{\mathbb{R}} x f(x) \lambda(dx)$$

genau dann, wenn $\mathbf{E}X$ existiert, und es gilt hierbei

$$\mathbf{E}X = \int_{\mathbb{R}} x f(x) \, dx. \quad \times$$

» *Beweisidee zu Satz 4.4a*) Für den Erwartungswert gilt nach 4.2,

$$EX = \int_{\Omega} X dP.$$

X nimmt höchstens abzählbar viele verschiedene Werte in \mathbb{N}_0 an, wir können das Integral also schreiben als,

$$\int_{\Omega} X dP = \lim_{n \rightarrow \infty} \sum_{k=0}^n k \cdot P[X = k] = \sum_{k=0}^{\infty} k p_k.$$

Spezialfall zu 4.4b). Sei $X \geq 0$ und F stetig differenzierbar auf $(0, \infty)$. Wir verwenden die Definition des Erwartungswerts als Riemann-Stieltjes-Integral

$$EX = \lim_{x \rightarrow \infty} \int_{[0, x]} t dF(t) \stackrel{!}{=} \lim_{x \rightarrow \infty} \int_{[0, x]} t F'(t) dt = \int_{[0, \infty)} x f(x) dx.$$

(!) gilt, da F stetig differenzierbar. Da f nur auf der positiven reellen Achse Werte $\neq 0$ annimmt, gilt somit

$$EX = \int_{\mathbb{R}} x f(x) dx.$$

Mit Hilfe des Transformationssatzes werden wir später einen vollständigen Beweis geben können. «

Für eine Zufallsvariable X auf (Ω, \mathcal{A}, P) mit Dichtefunktion bzw. Zähldichte haben wir mit Satz 4.4 eine konkrete Möglichkeit den Erwartungswert zu berechnen. Wir betrachten dazu nun einige Beispiele.

BSP 4 Sei X $b(n, p)$ -verteilt, d.h. X hat die Zähldichte

$$p_k = P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}, \quad (k = 0, 1, \dots, n)$$

mit $n \in \mathbb{N}$ und $p \in [0, 1]$. Für den Erwartungswert gilt $EX = np$. ■

BSP 5 Sei X $\pi(\lambda)$ -verteilt, d.h.

$$p_k = P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0,$$

mit $\lambda > 0$, so gilt $EX = \lambda$.

» Zu berechnen ist der Wert folgender Reihe,

$$\mathbf{E}X = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!}.$$

Wir verwenden dazu folgende Hilfsfunktion,

$$g(s) := \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda} e^{\lambda s}, \quad s \in [0, 1].$$

g bezeichnet man auch als **erzeugende Funktion** der Poisson-Verteilung. Wir werden in Kapitel 6 genauer darauf eingehen. Die Reihe konvergiert auf $[0, 1]$ gleichmäßig, wir können also gliedweise differenzieren und erhalten somit,

$$g'(s) = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} s^{k-1} = \lambda e^{-\lambda} e^{\lambda s}.$$

Setzen wir $s = 1$, so erhalten wir

$$\mathbf{E}X = \sum_{k=1}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = g'(1) = \lambda. \quad \ll \blacksquare$$

BSP 6 Sei $X N(a, \sigma^2)$ -verteilt, d.h. X besitzt die Dichtefunktion $f: \mathbb{R} \rightarrow \mathbb{R}_+$ mit

$$f(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad a \in \mathbb{R}, \quad \sigma > 0.$$

Es gilt $\mathbf{E}X = a$.

» Wir wenden Satz 4.4b) an und erhalten,

$$\mathbf{E}X = \int_{\mathbb{R}} x f(x) dx = \int_{\mathbb{R}} (x - a) f(x) dx + \int_{\mathbb{R}} a f(x) dx.$$

$f(x)$ ist symmetrisch bezüglich $x = a$, d.h. der linke Integrand ist asymmetrisch bezüglich $x = a$ und damit verschwindet das Integral.

$$\Rightarrow \mathbf{E}X = a \int_{\mathbb{R}} f(x) dx = a,$$

da f Dichte. $\ll \blacksquare$

BSP 7 Sei $X \exp(\lambda)$ verteilt mit $\lambda > 0$, so besitzt X die Dichtefunktion f mit,

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

Es gilt $EX = \frac{1}{\lambda}$.

» Eine Anwendung von Satz 4.4b) ergibt,

$$\begin{aligned} EX &= \int_{\mathbb{R}} x f(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx \stackrel{\text{part.int.}}{=} -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda}. \quad \ll \blacksquare \end{aligned}$$

BSP 8 X sei auf $[a, b]$ gleichverteilt, d.h. mit Dichtefunktion $f : \mathbb{R} \rightarrow \mathbb{R}_+$,

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

Dann ist $EX = \frac{a+b}{2}$.

» f ist eine Rechteckfunktion, ihr Erwartungswert ist die Intervallmitte. « ■

■ Der Transformationssatz

4.5 **Transformationssatz** Sei $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ eine reelle Zufallsvariable mit Verteilung P_X und Verteilungsfunktion F , sowie

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

$E(g \circ X)$ existiert genau dann, wenn $\int_{\mathbb{R}} g dP_X$ existiert, und es gilt hierbei nach dem Transformationssatz für Integrale

$$E(g \circ X) = \int_{\mathbb{R}} g dP_X =: \int_{\mathbb{R}} g dF = \int_{\mathbb{R}} g(x) dF(x).$$

Spezialfälle sind:

$$E(g \circ X) = \begin{cases} \sum_{k \geq 0} g(k) p_k, & \text{unter der Voraussetzung von Satz 4.4a,} \\ \int_{\mathbb{R}} g(x) f(x) d\lambda(x), & \text{unter der Voraussetzung von Satz 4.4b.} \end{cases}$$

Insbesondere gilt (mit $g = \mathbf{1}_A$, wobei $A \in \mathcal{B}$)

$$P[X \in A] = P_X(A) = \begin{cases} \sum_{k \in A \cap \mathbb{N}_0} p_k & \text{u. V. von Satz 4.4a} \\ \int_{\mathbb{R}} \mathbf{1}_A(x) f(x) d\lambda(x) = \int_A f(x) dx & \text{u. V. von Satz 4.4b. } \times \end{cases}$$

» Wir führen den Beweis nach der STANDARDPROZEDUR.

o. Schritt: Beweis der Behauptung für $g = \mathbf{1}_A$ und $A \in \mathcal{B}$.

$$\int_{\mathbb{R}} \mathbf{1}_A(x) dP_X = P_X(A) = P(X^{-1}(A)) = \int_{\Omega} X^{-1}(A) dP = \int_{\Omega} \mathbf{1}_A(X) dP.$$

1. Schritt: $g \geq 0$, wobei g Linearkombination einfacher Funktionen.

2. Schritt: $g \geq 0$. Verwende den Satz von der monotonen Konvergenz.

3. Schritt: $g = g^+ - g^-$. «

Wir können somit einen vollständigen Beweis von 4.4 b) geben.

» Sei X reelle Zufallsvariable und $g = \text{id}$, d.h. $g(x) = x$, so gilt

$$EX = \mathbf{E}(g \circ X) = \int_{\mathbb{R}} g dP_X = \int_{\mathbb{R}} \text{id} dP_X.$$

Die Verteilungsfunktion F erzeugt das Maß P_X , somit gilt

$$\int_{\mathbb{R}} \text{id} dP_X = \int_{\mathbb{R}} x dF(x).$$

Besitzt X die Dichte f , d.h. $F' = f$ L-f.ü., so gilt

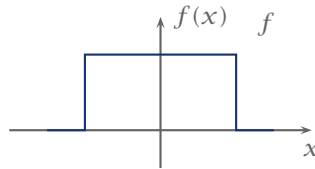
$$\int_{\mathbb{R}} x dF(x) = \int_{\mathbb{R}} x f(x) d\lambda(x). \quad \ll$$

Wir betrachten einige Beispiele zu Satz 4.5

BSP 9 Ein Schütze trifft einen Punkt des Intervalls $[-1, 1]$ gemäß einer Gleichverteilung und erhält den Gewinn $\frac{1}{\sqrt{|x|}}$, wenn er den Punkt x trifft, wobei wir $\frac{1}{0} := \infty$ gewichten.

Gesucht ist der mittlere Gewinn des Schützen. Dazu konstruieren wir eine Zufallsvariable X , die den zufällig getroffenen Punkt angibt. X ist nach Voraussetzung gleichverteilt, d.h. die zugehörige Dichte gegeben durch,

$$f(x) = \begin{cases} \frac{1}{2}, & x \in [-1, 1], \\ 0, & \text{sonst.} \end{cases}$$

4.1 Dichtefunktion zu X .

Den mittleren Gewinn erhalten wir durch den Erwartungswert,

$$\begin{aligned} \mathbf{E}(\text{Gewinn}) &= \mathbf{E} \frac{1}{\sqrt{X}} = \int_{\mathbb{R}} \frac{1}{\sqrt{|x|}} f(x) \, dx = \frac{1}{2} \int_{[-1,1]} \frac{1}{\sqrt{|x|}} \, dx \\ &= \int_0^1 \frac{1}{\sqrt{x}} \, dx = 2\sqrt{x} \Big|_0^1 = 2. \quad \blacksquare \end{aligned}$$

BSP 10 Ein Punkt wird rein zufällig aus der Einheitskreisscheibe K von 0 ausgewählt. Wie groß ist der zufällige Abstand dieses Punktes von 0?

Um diese Fragestellung zu modellieren, konstruieren wir einen zweidimensionalen Zufallsvektor $X = (X_1, X_2)$, der auf K gleichverteilt ist. Die Dichtefunktion $f: \mathbb{R}^2 \rightarrow \mathbb{R}_+$ ist dann gegeben durch,

$$f(x_1, x_2) = \begin{cases} \frac{1}{\pi}, & (x_1, x_2) \in K, \\ 0, & \text{sonst.} \end{cases}$$

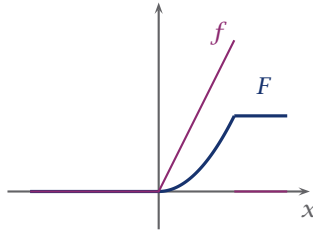
Gesucht ist nun $\mathbf{E}\sqrt{X_1^2 + X_2^2}$. Man arbeitet üblicherweise nicht auf Ω .

1. Weg (üblich). Anwendung des Transformationssatzes,

$$\begin{aligned} \mathbf{E}\sqrt{X_1^2 + X_2^2} &\stackrel{4.5}{=} \int_{\mathbb{R}^2} \sqrt{x_1^2 + x_2^2} f(x_1, x_2) \, d\lambda(x_1, x_2) \\ &\stackrel{\text{pol.koord.}}{=} \frac{1}{\pi} \int_{\varphi=0}^{2\pi} \int_{r=0}^1 r \cdot r \, d\lambda(r, \varphi) = \frac{2}{3}. \end{aligned}$$

2. Weg. Betrachte die Zufallsvariable $Y = \sqrt{X_1^2 + X_2^2}$ mit der zugehörigen Verteilungsfunktion

$$F(y) := P[Y \leq y] = \begin{cases} 0, & \text{falls } y \leq 0, \\ \frac{1}{\pi} \pi y^2 = y^2, & \text{falls } y \in (0, 1), \\ 1, & \text{falls } y \geq 1. \end{cases}$$



4.2 Dichtefunktion zu X .

F ist bis auf den Punkt $y = 1$ stetig differenzierbar, es gilt also

$$EY = \int_0^1 yF'(y) dy = 2 \int_0^1 y^2 dy = \frac{2}{3}.$$

Welcher Weg (schneller) zum Ziel führt, hängt vom Problem ab. ■

Zusatz Sei X ein n -dimensionaler Zufallsvektor auf (Ω, \mathcal{A}, P) und

$$g : (\mathbb{R}^n, \mathcal{B}_n) \rightarrow (\mathbb{R}, \mathcal{B}),$$

so existiert $E(g \circ X)$ genau dann, wenn $\int_{\mathbb{R}^n} g dP_X$ existiert. In diesem Fall gilt

$$E(g \circ X) = \int_{\Omega} g \circ X dP = \int_{\mathbb{R}^n} g dP_X.$$

Besitzt X außerdem eine Dichtefunktion $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$, so gilt

$$E(g \circ X) = \int_{\mathbb{R}^n} g(x) f(x) d\lambda(x).$$

Insbesondere gilt für $A \in \mathcal{B}_n$,

$$P[X \in A] = P_X(A) = \int_{\mathbb{R}^n} \mathbf{1}_A(x) f(x) dx = \int_A f(x) dx.$$

[Falls das entsprechende Riemann-Integral existiert, so stimmen Riemann- und Lebesgue-Integral überein.] ✕

» Der Beweis erfolgt wie im eindimensionalen Fall. Die letzte Behauptung folgt sofort für $g = \mathbf{1}_A$. «

- 4.4 *Bemerkung.* Das Integral $\int_A f(x) dx$ im Zusatz zu Satz 4.4 ist entspricht dem λ -Maß der Ordinatenmenge

$$\{(x, y) \in \mathbb{R}^{n+1} : x \in A, 0 \leq y \leq f(x)\}$$

von $f|_A$ in \mathbb{R}^{n+1} . \rightarrow

4-C Momente von Zufallsvariablen

- 4.5 **Definition** Sei X eine reelle Zufallsvariable und $k \in \mathbb{N}$. Im Falle der Existenz heißt $\mathbf{E}X^k$ das *k-te Moment* von X und $\mathbf{E}(X - \mathbf{E}X)^k$ das *k-te zentrale Moment* von X . Ist X integrierbar, dann heißt

$$\mathbf{V}(X) := \mathbf{E}(X - \mathbf{E}X)^2$$

die *Varianz* von X und

$$\sigma(X) := \sqrt{\mathbf{V}(X)}$$

die *Streuung* von X . \times

Das erste Moment einer Zufallsvariablen haben wir bereits als Erwartungswert kennengelernt. Die Varianz $\mathbf{V}(X)$, gibt die "Schwankung" der Zufallsvariablen X als "mittleren Wert" ihrer quadratischen Abweichung von $\mathbf{E}X$ an.

Für die Existenz der Varianz ist es nicht notwendig, die quadratische Integrierbarkeit von X explizit zu fordern, da durch die Integrierbarkeit von X gesichert ist, dass $(X - \mathbf{E}X)^2$ eine nicht-negative reelle Zufallsvariable und damit $\mathbf{E}(X - \mathbf{E}X)^2$ definiert ist (wenn auch möglicherweise ∞).

- 4.2 **Lemma** Sei X eine reelle Zufallsvariable und $0 \leq \alpha < \beta < \infty$. Dann gilt

$$\mathbf{E}|X|^\beta < \infty \Rightarrow \mathbf{E}|X|^\alpha < \infty. \quad \times$$

$$\gg \mathbf{E}|X|^\alpha \leq \mathbf{E}\left(\max\{1, |X|^\beta\}\right) \leq 1 + \mathbf{E}|X|^\beta < \infty. \quad \ll$$

Da $\mathbf{E}X^k$ genau dann existiert, wenn $\mathbf{E}|X|^k$ existiert, impliziert die Existenz des k -ten Moments somit die Existenz aller niedrigeren Momente $\mathbf{E}X^l$ mit $l \leq k$.

- 4.6 **Satz** Sei X eine reelle integrierbare Zufallsvariable. Dann gilt

$$a) \mathbf{V}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2.$$

$$b) \mathbf{V}(aX + b) = a^2\mathbf{V}(X), \quad a, b \in \mathbb{R}. \quad \times$$

Im Gegensatz zum Erwartungswert ist die Varianz ist also insbesondere *keine* lineare Operation.

» Den Fall $\mathbf{E}X^2 = \infty$ behandeln wir durch Stützung von X in Höhe c und gehen dann zum Grenzwert für $c \rightarrow \infty$ über.

Betrachten wir also den Fall $\mathbf{E}X^2 < \infty$.

a) $\mathbf{V}(X) = \mathbf{E}(X - \mathbf{E}(X))^2 = \mathbf{E}(X^2 - 2(\mathbf{E}X)X + (\mathbf{E}X)^2)$. Nach Voraussetzung ist $\mathbf{E}X < \infty$ und somit,

$$\mathbf{V}(X) = \mathbf{E}X^2 - 2(\mathbf{E}X)(\mathbf{E}X) + (\mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2.$$

b) Unter Verwendung der Linearität von \mathbf{E} erhalten wir sofort,

$$\begin{aligned} \mathbf{V}(aX + b) &= \mathbf{E}(aX + b - \mathbf{E}(aX + b))^2 = \mathbf{E}(aX + b - a\mathbf{E}X - b)^2 \\ &= a^2\mathbf{E}(X - \mathbf{E}X)^2 = a^2\mathbf{V}(X). \quad \ll \end{aligned}$$

BSP 11 X sei $b(n, p)$ -verteilt mit $n \in \mathbb{N}$, $p \in [0, 1]$. $\mathbf{V}(x) = np(1 - p)$. Wir werden dafür später einen einfachen Beweis gegeben können. ■

BSP 12 X sei $\pi(\lambda)$ -verteilt mit $\lambda > 0$. Die Berechnung der Varianz erfolgt mittels der Erzeugendenfunktion,

$$\begin{aligned} g(s) &= \sum_{k=0}^n e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda} e^{\lambda s}, \\ g''(s) &= \sum_{k=2}^n k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} s^{k-2} = \lambda^2 e^{-\lambda} e^{\lambda s}, \quad s \in [0, 1]. \end{aligned}$$

Auswerten ergibt

$$g''(1) = \underbrace{\sum_{k=2}^n k(k-1) e^{-\lambda} \frac{\lambda^k}{k!}}_{4.5 = \mathbf{E}(X(X-1))} = \lambda^2.$$

Somit erhalten wir,

$$\mathbf{E}X^2 = \mathbf{E}(X(X-1)) + \mathbf{E}X = \lambda^2 + \lambda,$$

$$\mathbf{V}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda. \quad \blacksquare$$

BSP 13 X sei $N(a, \sigma^2)$ -verteilt mit $a \in \mathbb{R}$, $\sigma > 0$, dann ist $Y = \frac{1}{\sigma}(X - a) \sim N(0, 1)$ verteilt. Wir können uns also auf den Fall $a = 0$, $\sigma = 1$ zurückziehen, wobei man leicht zeigen kann, dass $\mathbf{E}|X|^n < \infty$, $\forall n \in \mathbb{N}$.

Weiterhin gilt

$$\mathbf{E}X^{2k+1} = 0, \quad k \in \mathbb{N}_0,$$

da der Integrand eine ungerade Funktion ist und daher das Integral verschwindet. Insbesondere ist $\mathbf{E}X = 0$.

Für geraden Exponenten erhalten wir

$$\begin{aligned} \mathbf{E}X^{2k} &\stackrel{4.5}{=} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^{2k} e^{-\frac{x^2}{2}} dx \\ &\stackrel{\text{part. int.}}{=} \underbrace{\frac{2}{\sqrt{2\pi}} \frac{1}{2k+1} x^{2k+1} e^{-\frac{x^2}{2}} \Big|_0^\infty}_{=0} - \frac{2}{\sqrt{2\pi}} \int_0^\infty \frac{1}{2k+1} x^{2k+1} (-x) e^{-\frac{x^2}{2}} dx \\ &= \frac{2}{\sqrt{2\pi}} \frac{1}{2k+1} \int_0^\infty x^{2k+2} e^{-\frac{x^2}{2}} dx = \frac{1}{2k+1} \mathbf{E}X^{2k+2}. \end{aligned}$$

Somit gilt für jedes $k \in \mathbb{N}_0$, $\mathbf{E}X^{2k+2} = (2k+1)\mathbf{E}X^{2k}$. Insbesondere erhalten wir

$$\mathbf{E}X^0 = \mathbf{E}1 = 1 \quad \Rightarrow \quad \mathbf{E}X^2 = 1 \quad \Rightarrow \quad \mathbf{V}(X) = 1.$$

Für allgemeines $N(a, \sigma^2)$ verteiltes X erhalten wir somit,

$$\begin{aligned} \mathbf{E}\left(\frac{1}{\sigma}(X - a)\right) &= 0 \Rightarrow \mathbf{E}X = a, \\ \mathbf{V}\left(\frac{1}{\sigma}(X - a)\right) &= 1 \Rightarrow \mathbf{V}X = \sigma^2. \quad \blacksquare \end{aligned}$$

Die folgenden Ungleichungen haben zahlreiche Anwendungen und gehören zu den wichtigsten Hilfsmitteln der Stochastik.

4.7 **Satz** Sei X eine reelle Zufallsvariable auf einem W -Raum (Ω, \mathcal{A}, P) . Dann gilt für jedes $\varepsilon > 0$, $r > 0$:

$$P[|X| \geq \varepsilon] \leq \varepsilon^{-r} \mathbf{E}|X|^r, \quad \text{Markoffsche Ungleichung.}$$

Für $r = 2$ erhält man die sog. *Tschebyschevsche Ungleichung*, die bei integrierbarem X auch in der Variante

$$P[|X - \mathbf{E}X| \geq \varepsilon] \leq \varepsilon^{-2} \mathbf{V}(X)$$

angegeben wird. \times

» Zum Beweis verwenden wir eine Zufallsvariable Y , die X “stutzt”,

$$Y := \begin{cases} 1, & \text{falls } |X| \geq \varepsilon, \\ 0, & \text{sonst.} \end{cases}$$

Offensichtlich ist $Y \leq \frac{|X|^r}{\varepsilon^r}$, d.h. $EY \leq \varepsilon^{-r} E|X|^r$, wobei $EY = P[|X| \geq \varepsilon]$. «

Als Anwendung beweisen wir das Bernoullische schwache Gesetz der großen Zahlen.

BSP 14 Sei Y_n eine Zufallsvariable, die die Anzahl der Erfolge in n Bernoulli-Versuchen mit jeweiliger Erfolgswahrscheinlichkeit $p \in [0, 1]$ (fest) angibt. Y_n ist also $b(n, p)$ -verteilt und es gilt

$$EY_n = np, \quad \mathbf{V}Y_n = np(1-p).$$

Die relative Häufigkeit der Anzahl der Erfolge ist gegeben durch $\frac{Y_n}{n}$.

Sei nun $\varepsilon > 0$ beliebig aber fest, so gilt

$$P\left[\left|\frac{Y_n}{n} - p\right| \geq \varepsilon\right] \leq \frac{\mathbf{V}\left(\frac{Y_n}{n}\right)}{\varepsilon^2} = \frac{1}{\varepsilon^2 n^2} \mathbf{V}(Y_n) = \frac{p(1-p)}{\varepsilon^2 n} \rightarrow 0, \quad n \rightarrow \infty.$$

Zusammenfassend erhalten wir das **Bernoullische schwache Gesetz der großen Zahlen**,

$$\forall \varepsilon > 0 : P\left[\left|\frac{Y_n}{n} - p\right| \geq \varepsilon\right] \rightarrow 0, \quad n \rightarrow \infty.$$

Eines der Ziele dieser Vorlesung ist die Verallgemeinerung dieses Gesetzes auf das starke Kolmogorovsche Gesetz der großen Zahlen in Kapitel 9. ■

5 Unabhängigkeit

Es ist eine zentrale Fragestellung der Wahrscheinlichkeitstheorie, inwiefern sich zufällige Experimente gegenseitig beeinflussen. In 1-D haben wir bereits die Bedingte Wahrscheinlichkeit definiert und festgestellt, dass sich Zufallsexperimente tatsächlich beeinflussen können und damit die Wahrscheinlichkeit für das Eintreten eines Ereignisses vom Eintreten eines vorangegangenen Ereignisses abhängen kann.

In diesem Kapitel wollen wir die Eigenschaften von unabhängigen Ereignissen und insbesondere unabhängigen Zufallsvariablen studieren. Solche Zufallsvariablen haben viele angenehme Eigenschaften und lassen sich besonders “leicht” handhaben — Die wirklich interessanten Experimente sind jedoch gerade *nicht* unabhängig.

5-A Unabhängige Ereignisse und Zufallsvariablen

Wir betrachten einige Beispiele zum Einstieg.

BSP 1 *Zwei Ausspielungen eines Würfels "ohne gegenseitige Beeinflussung".*

Betrachte dazu einen laplaceschen W-Raum (Ω, \mathcal{A}, P) mit $\Omega = \{1, \dots, 6\}^2$ und $\mathcal{A} = \mathcal{P}(\Omega)$.

Sei A das Ereignis, dass im ersten Wurf eine 1 erzielt wird, B das Ereignis, dass im zweiten Wurf eine gerade Zahl erscheint. Einfaches Abzählen ergibt,

$$P(A) = \frac{6}{36} = \frac{1}{6},$$

$$P(B) = \frac{18}{36} = \frac{1}{2}.$$

Die Wahrscheinlichkeit, dass beide Ereignisse gleichzeitig eintreten ist,

$$P(A \cap B) = \frac{3}{36} = \frac{1}{12} = \frac{1}{6} \cdot \frac{1}{2} = P(A) \cdot P(B).$$

Eine solche Situation ($P(A \cap B) = P(A) \cdot P(B)$) ist typisch bei Ereignissen, bei denen es physikalisch keine gegenseitige Beeinflussung gibt. ■

Wir werden den Fall, dass die Wahrscheinlichkeit des gemeinsamen Eintretens mit dem Produkt der einzelnen Wahrscheinlichkeiten übereinstimmt als Grundlage für die Definition der stochastischen Unabhängigkeit verwenden. Es ist wichtig, zwischen "stochastisch unabhängig" und "physikalisch unabhängig" zu unterscheiden, denn im Allgemeinen lassen sich aus der stochastischen Unabhängigkeit keine Rückschlüsse auf die physikalische machen.

Sei (Ω, \mathcal{A}, P) ein W-Raum, A und $B \in \mathcal{A}$ mit $P(B) > 0$, dann gilt

$$P(A \cap B) = P(A | B)P(B).$$

Falls A von B "unabhängig" ist, muss gelten $P(A | B) = P(A)$ und damit

$$P(A \cap B) = P(A)P(B).$$

- 5.1 **Definition** Sei (Ω, \mathcal{A}, P) ein W-Raum. Eine Familie $\{A_i : i \in I\}$ von Ereignissen $A_i \in \mathcal{A}$ heißt **unabhängig** (ausführlich: stochastisch unabhängig bzgl. P), falls für jede nichtleere endliche Menge $K \subset I$ gilt

$$P\left(\bigcap_{k \in K} A_k\right) = \prod_{k \in K} P(A_k). \quad \times$$

In dieser Definition werden keine Anforderungen an die Indexmenge I gemacht, unendliche (z.B. überabzählbare) Indexmengen sind also durchaus zugelassen.

Im Folgenden sei mit "unabhängig" stets stochastisch unabhängig gemeint. Auf logische oder physikalische Unabhängigkeit lassen sich aus der stochastischen Unabhängigkeit im Allgemeinen keine Rückschlüsse machen.

- 5.2 **Definition** Sei (Ω, \mathcal{A}, P) ein W-Raum. Eine Familie $\{X_i : i \in I\}$ von $(\Omega_i, \mathcal{A}_i)$ - Zufallsvariablen auf (Ω, \mathcal{A}, P) heißt **unabhängig**, wenn gilt: Für jede nichtleere endliche Indexmenge $\{i_1, \dots, i_n\} \subset I$ und jede Wahl von Mengen $A_{i_\nu} \in \mathcal{A}_{i_\nu}$ ($\nu = 1, \dots, n$) ist

$$P[X_{i_1} \in A_{i_1}, \dots, X_{i_n} \in A_{i_n}] = \prod_{\nu=1}^n P[X_{i_\nu} \in A_{i_\nu}].$$

Sprechweise: Unabhängigkeit der Zufallsvariablen statt Unabhängigkeit der Familie der Zufallsvariablen. ✕

Für die Unabhängigkeit von Zufallsvariablen ist es somit nicht notwendig explizit zu fordern, dass alle Zufallsvariablen den selben Wertebereich teilen. Die Voraussetzung, dass alle auf dem selben W-Raum (Ω, \mathcal{A}, P) definiert sind, lässt sich jedoch nicht abschwächen.

- 5.1 **Lemma** Eine Familie von Ereignissen ist genau dann unabhängig, wenn jede endliche Teilfamilie unabhängig ist. Entsprechendes gilt für Zufallsvariablen. ✕

» Die Aussage folgt direkt aus den Definitionen 5.1 und 5.2. «

- 5.1 **Bemerkung zu Definition 5.1 bzw. Definition 5.2.** Die paarweise Unabhängigkeit impliziert im Allgemeinen *nicht* die Unabhängigkeit. ∞

» *Beweis durch Gegenbeispiel.* Betrachte zwei Auspielungen eines echten Würfels ohne gegenseitige Beeinflussung. Dieses Experiment können wir durch einen Laplaceschen W-Raum (Ω, \mathcal{A}, P) mit $|\Omega| = 36$ modellieren.

Sei A_i das Ereignis, dass im i -ten Wurf eine ungerade Zahl auftritt $P(A_i) = \frac{1}{2}$ und B das Ereignis, dass die Summe der Augenzahlen ungerade ist $P(B) = \frac{1}{2}$.

$$P(A_1 \cap A_2) = \frac{9}{36} = \frac{1}{4}, \quad P(A_1 \cap B) = \frac{9}{36} = \frac{1}{4}, \quad P(A_2 \cap B) = \frac{9}{36} = \frac{1}{4},$$

somit sind die Ereignisse paarweise unabhängig aber

$$P(A_1 \cap A_2 \cap B) = P(\emptyset) = 0 \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}. \quad \ll$$

- 5.2 *Bemerkung.* Sei (Ω, \mathcal{A}, P) ein W-Raum und $A_i \in \mathcal{A}$, $i \in I$ Ereignisse, mit (reellen) Zufallsvariablen und Indikatorfunktionen $\mathbf{1}_{A_i}$ auf (Ω, \mathcal{A}, P) .

$$\{A_i : i \in I\} \text{ unabhängig} \Leftrightarrow \{\mathbf{1}_{A_i} : i \in I\} \text{ unabhängig} . \quad \rightarrow$$

■ Unabhängige Zufallsvariablen

- 5.1 **Satz** Gegeben seien $(\Omega_i, \mathcal{A}_i)$ - Zufallsvariablen X_i auf einem W-Raum (Ω, \mathcal{A}, P) und $g_i : (\Omega_i, \mathcal{A}_i) \rightarrow (\Omega'_i, \mathcal{A}'_i)$, $i \in I$ Abbildungen. Dann gilt:

$$\{X_i : i \in I\} \text{ unabhängig} \Rightarrow \{g_i \circ X_i : i \in I\} \text{ unabhängig} \quad \times$$

» Aufgrund von Lemma 5.1 genügt es den Beweis auf einer endlichen Indexmenge $I = \{1, \dots, n\}$ zu führen.

$$\begin{aligned} P[g_1 \circ X_1 \in A'_1, \dots, g_n \circ X_n \in A'_n] &= P\left[\forall i \in \{1, \dots, n\} : X_i \in g_i^{-1}(A'_i)\right] \\ &= \prod_{i=1, \dots, n} P\left[X_i \in g_i^{-1}(A'_i)\right] = \prod_{i=1, \dots, n} P\left[g_i \circ X_i \in A'_i\right], \end{aligned}$$

da die X_i nach Voraussetzung unabhängig sind. «

Sind also die Zufallsvariablen $\{X_i\}$ unabhängig, so sind es insbesondere auch $\{X_i^2\}$, $\{\sqrt{X_i}\}$, ...

BSP 2 Betrachte erneut zwei Ausspielungen eines echten Würfels ohne gegenseitige Beeinflussung mit W-Raum (Ω, \mathcal{A}, P) .

Sei X_1, X_2 die Augenzahl im 1. bzw. 2. Wurf. Da $\{X_1, X_2\}$ unabhängig, ist auch $\{X_1^2, X_2^3\}$ unabhängig. ■

- 5.2 **Satz** Sei eine unabhängige Familie $\{X_i : i \in I\}$ reeller Zufallsvariablen auf einem W-Raum (Ω, \mathcal{A}, P) und eine Zerlegung $I = \sum_{j \in J} I_j$ mit $|I_j| < \infty$ gegeben. Das $|I_j|$ -Tupel $\{X_k : k \in I_j\}$ sei mit Y_j bezeichnet für $j \in J$.

a) Die Familie $\{Y_j : j \in J\}$ ist unabhängig.

b) Sind Abbildungen $g_j : (\mathbb{R}^{|I_j|}, \mathcal{B}_{|I_j|}) \rightarrow (\mathbb{R}, \mathcal{B})$, $j \in J$, gegeben, so ist die Familie $\{g_j \circ Y_j : j \in J\}$ unabhängig. \times

- » a) *Beweisidee.* Betrachte den Spezialfall $Y_1 = X_1$, $Y_2 = (X_2, X_3)$. Nach Annahme gilt, dass die X_i unabhängige reelle Zufallsvariablen sind. Zu zeigen ist nun für $B_1 \in \mathcal{B}, B_2 \in \mathcal{B}_2$

$$P[Y_1 \in B_1, Y_2 \in B_2] = P[Y_1 \in B_1] \cdot P[Y_2 \in B_2].$$

Dabei genügt es nach Satz 1.4, die Aussage auf Intervallen $I_1 \in \mathcal{B}_1$ bzw. Rechtecken $I_2 \times I_3 \in \mathcal{B}_2$ nachzuweisen. Also

$$\begin{aligned} P[X_1 \in I_1, (X_2, X_3) \in I_2 \times I_3] &= P[X_1 \in I_1, X_2 \in I_2, X_3 \in I_3] \\ &= P[X_1 \in I_1] P[X_2 \in I_2, X_3 \in I_3] \\ &= P[X_1 \in I_1] P[(X_2, X_3) \in I_2 \times I_3]. \end{aligned}$$

- b) Anwendung von Satz 5.1. «

5.3 *Bemerkung.* Satz 5.2 lässt sich auf $(\Omega_i, \mathcal{A}_i)$ - Zufallsvariablen und Abbildungen $g_j : \otimes_{i \in I_j} (\Omega_i, \mathcal{A}_i) \rightarrow (\Omega'_j, \mathcal{A}'_j)$ verallgemeinern. «

5.3 **Satz** Seien X_i ($i = 1, \dots, n$) reelle Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) mit Verteilungsfunktionen F_i . Der Zufallsvektor $X := (X_1, \dots, X_n)$ habe die n -dimensionale Verteilungsfunktion F .

- a) $\{X_1, \dots, X_n\}$ ist genau dann unabhängig, wenn gilt

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n : F(x_1, \dots, x_n) = \prod_{i=1}^n F_i(x_i) \quad (*)$$

- b) Existieren zu F_i ($i = 1, \dots, n$) bzw. F , Dichten f_i bzw. f , dann gilt

$$(*) \Leftrightarrow f(x_1, \dots, x_n) = \prod_{i=1}^n f_i(x_i) \text{ für } \lambda\text{-fast alle } (x_1, \dots, x_n) \in \mathbb{R}^n. \quad \times$$

- » a) Seien $\{X_1, \dots, X_n\}$ unabhängig, so ist Definition 5.2 äquivalent zu

$$\Leftrightarrow \forall B_1, \dots, B_n \in \mathcal{B} : \begin{aligned} P[X_1 \in B_1, \dots, X_n \in B_n] \\ = P[X_1 \in B_1] \cdots P[X_n \in B_n], \end{aligned}$$

$$\Leftrightarrow \forall x_1, \dots, x_n \in \mathbb{R} : \begin{aligned} P[X_1 \leq x_1, \dots, X_n \leq x_n] \\ = P[X_1 \leq x_1] \cdots P[X_n \leq x_n] \end{aligned}$$

$$\Leftrightarrow \forall x_1, \dots, x_n \in \mathbb{R} : F(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n). \quad \llcorner$$

Unabhängige Zufallsvariablen haben somit die angenehme Eigenschaft, dass die **gemeinsame Dichte** (die Dichte der gemeinsamen Verteilung) dem Produkt der **Randverteilungsdichten** (der Dichten der Randverteilungen) entspricht. Insbesondere ist für unabhängige Zufallsvariablen die gemeinsame Verteilung *eindeutig* durch die Randverteilungen bestimmt.

BSP 3 Seien X, Y unabhängig, wobei $X \exp(\lambda)$ und $Y \exp(\tau)$ verteilt, so besitzen sie die Dichten

$$f_X(x) = \lambda e^{-\lambda x}, \quad f_Y(y) = \tau e^{-\tau y}.$$

Die gemeinsame Dichte ist nach Satz 5.3 gegeben durch

$$f(x, y) = \lambda \tau e^{-\lambda x} e^{-\tau y}.$$

Betrachten wir andererseits die gemeinsame Verteilung, so gilt

$$F(s, t) = P[X \leq s, Y \leq t] \stackrel{!}{=} P[X \leq s]P[Y \leq t]$$

da X und Y unabhängig. Weiterhin ist

$$\begin{aligned} P[X \leq s]P[Y \leq t] &= F_X(s)F_Y(t) = \left(\int_0^s \lambda e^{-\lambda x} dx \right) \left(\int_0^t \tau e^{-\tau y} dy \right) \\ &= \int_0^s \int_0^t \underbrace{\lambda \tau e^{-\lambda x} e^{-\tau y}}_{f(x, y)} dx dy. \quad \blacksquare \end{aligned}$$

5.4 **Satz** Seien X_1, \dots, X_n unabhängige reelle Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) mit endlichen Erwartungswerten. Dann existiert

$$\mathbf{E}(X_1 \cdot \dots \cdot X_n) = \mathbf{E}X_1 \cdot \dots \cdot \mathbf{E}X_n.$$

— Entsprechendes gilt für komplexe Zufallsvariablen, wobei wir \mathbb{C} mit dem \mathbb{R}^2 identifizieren und komplexe Zufallsvariablen als $X_k : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^2, \mathcal{B}_2)$ mit

$$\mathbf{E}X_k := \mathbf{E} \operatorname{Re} X_k + i \mathbf{E} \operatorname{Im} X_k \quad (k = 1, \dots, n). \quad \times$$

» Wir beschränken uns zunächst auf reelle Zufallsvariablen. Mit Satz 5.2 ist eine Reduktion auf den Fall $n = 2$ möglich. Zu zeigen ist also, dass $\mathbf{E}(X_1 X_2) = (\mathbf{E}X_1)(\mathbf{E}X_2)$.

Ohne Einschränkung sind $X_1, X_2 \geq 0$, ansonsten betrachten wir eine Zerlegung in $X_1 = X_{1,+} - X_{1,-}$. Mit Hilfe des des Satzes von der monotonen Konvergenz können wir uns auf einfache Funktionen der Form

$$x \mapsto \sum_{j=1}^n \lambda_j \mathbf{1}_{A_j}(x)$$

zurückziehen. Wir beweisen die Aussage nun für Indikatorfunktionen $X_1 = \mathbf{1}_{A_1}$, $X_2 = \mathbf{1}_{A_2}$ mit $A_1, A_2 \in \mathcal{A}$.

$$\begin{aligned} \mathbf{E}(X_1 X_2) &= \mathbf{E}(\mathbf{1}_{A_1} \mathbf{1}_{A_2}) = \mathbf{E}(\mathbf{1}_{A_1 \cap A_2}) = P(A_1 \cap A_2) = P(\mathbf{1}_{A_1} = 1, \mathbf{1}_{A_2} = 1) \\ &= P(\mathbf{1}_{A_1} = 1)P(\mathbf{1}_{A_2} = 1) = P(A_1)P(A_2) = (\mathbf{E}\mathbf{1}_{A_1})(\mathbf{E}\mathbf{1}_{A_2}) \\ &= (\mathbf{E}X_1)(\mathbf{E}X_2). \quad \ll \end{aligned}$$

Der Erwartungswert ist somit linear und vertauscht mit der Multiplikation.

5.5 **Satz von Bienaymé** Seien X_1, \dots, X_n paarweise unabhängige reelle Zufallsvariablen mit endlichen Erwartungswerten. Dann gilt

$$\mathbf{V}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \mathbf{V}(X_j). \quad \times$$

» Addieren wir zum Erwartungswert eine Konstante, ändert sich die Varianz nicht, wir können also ohne Einschränkung annehmen $\mathbf{E}X_j = 0$ ($j = 1, \dots, n$).

$$\begin{aligned} \mathbf{V}\left(\sum_{j=1}^n X_j\right) &= \mathbf{E}\left(\sum_{j=1}^n X_j\right)^2 = \sum_{i=1}^n \mathbf{E}X_i^2 + 2 \sum_{i < j} \mathbf{E}(X_i X_j) \\ &= \sum_{i=1}^n \mathbf{E}X_i^2 + 2 \underbrace{\sum_{i < j} \mathbf{E}(X_i) \mathbf{E}(X_j)}_{=0} = \sum_{i=1}^n \mathbf{E}X_i^2 = \sum_{i=1}^n \mathbf{V}X_i. \quad \ll \end{aligned}$$

Für unabhängige Zufallsvariablen vertauscht die Varianz mit *endlichen* Summen, wobei für reelle Konstanten $a, b \in \mathbb{R}$ und eine Zufallsvariable X gilt

$$\mathbf{V}(aX + b) = a^2 \mathbf{V}(X).$$

Dies ist kein Widerspruch, denn setzen wir $Y \equiv b$, so ist Y Zufallsvariable und $\mathbf{E}Y = b$. Somit ist $\mathbf{V}Y = \mathbf{E}Y^2 - (\mathbf{E}Y)^2 = b^2 - b^2 = 0$, also

$$\mathbf{V}(aX + b) = \mathbf{V}(aX + Y) = \mathbf{V}(aX) + \mathbf{V}(Y) = a^2 \mathbf{V}(X).$$

- 5.4 *Bemerkung.* In Satz 5.5 genügt statt der Unabhängigkeit von $\{X_1, \dots, X_n\}$ die sog. **paarweise Unkorreliertheit** zu fordern, d.h.

$$\forall j \neq k : \mathbf{E}((X_j - \mathbf{E}X_j)(X_k - \mathbf{E}X_k)) = 0.$$

Dies ist tatsächlich eine schwächere Voraussetzung, denn sind X und Y unabhängig, so gilt

$$\begin{aligned} \mathbf{E}((X - \mathbf{E}X)(Y - \mathbf{E}Y)) &= \mathbf{E}(X \cdot Y - (\mathbf{E}X)Y - (\mathbf{E}Y)X + (\mathbf{E}X)(\mathbf{E}Y)) \\ &= \mathbf{E}(X \cdot Y) - (\mathbf{E}X)(\mathbf{E}Y) - (\mathbf{E}Y)(\mathbf{E}X) + (\mathbf{E}X)(\mathbf{E}Y) \\ &= \mathbf{E}(X)(\mathbf{E}Y) - (\mathbf{E}X)(\mathbf{E}Y) = 0. \quad \rightarrow \end{aligned}$$

■ Faltungen

Sind X, Y zwei unabhängige reelle Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) mit Verteilungen P_X, P_Y , so ist nach Satz 5.3 und Satz 2.1 die Verteilung $P_{(X,Y)}$ des 2-dimensionalen Zufallsvektors durch P_X und P_Y eindeutig festgelegt.

Setzen wir $g : (x, y) \mapsto x + y$, so ist g messbar und damit auch die Verteilung $P_{X+Y} = P_{g \circ (X,Y)}$ von $X + Y$ eindeutig durch P_X und P_Y festgelegt.

Wie ermittelt man nun diese Verteilung?

- 5.3 **Definition** Sind X, Y zwei unabhängige reelle Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) mit Verteilungen P_X, P_Y , so wird die Verteilung $P_{X+Y} = P_X * P_Y$ der Zufallsvariable $X + Y$ als **Faltung** von P_X und P_Y bezeichnet. \times

- 5.5 *Bemerkung.* Die Faltungsoperation ist kommutativ und assoziativ. \rightarrow

- 5.6 **Satz** Seien X, Y zwei unabhängige reelle Zufallsvariablen.

- a) Besitzen X und Y Dichten f bzw. g , so besitzt $X + Y$ eine [als Faltung von f und g bezeichnete] Dichte h mit

$$h(t) = \int_{\mathbb{R}} f(t - y)g(y) dy = \int_{\mathbb{R}} g(t - x)f(x) dx, \quad t \in \mathbb{R}.$$

- b) Besitzen X und Y Zähldichten $(p_k)_{k \in \mathbb{N}_0}$ bzw. $(q_k)_{k \in \mathbb{N}_0}$, so besitzt $X + Y$ eine [als Faltung von (p_k) und (q_k) bezeichnete] Zähldichte $(r_k)_{k \in \mathbb{N}_0}$ mit

$$r_k = \sum_{j=0}^k p_{k-j}q_j = \sum_{i=0}^k q_{k-i}p_i, \quad k \in \mathbb{N}_0. \quad \times$$

Den Beweis verschieben wir auf das nächste Kapitel, denn mit Hilfe der charakteristischen Funktionen werden wir über die für einen einfachen Beweis nötigen Mittel verfügen.

BSP 4 Sei X_1 eine $b(n_1, p)$ -verteilte und X_2 eine $b(n_2, p)$ -verteilte Zufallsvariable. Sind X_1 und X_2 unabhängig, dann ist $X_1 + X_2$ eine $b(n_1 + n_2, p)$ -verteilte Zufallsvariable.

» *Einfacher Beweis.* Betrachte $n_1 + n_2$ Bernoulli-Versuche mit jeweiliger Erfolgswahrscheinlichkeit p . Sei nun Y_1 die Zufallsvariable, die die Anzahl der Erfolge der ersten n_1 Versuche angibt und Y_2 die Zufallsvariable, die die Anzahl der Erfolge in den letzten n_2 Versuchen angibt. Somit sind Y_1 und Y_2 unabhängig und $Y_1 + Y_2$ gibt die Anzahl der Erfolge in $n_1 + n_2$ Bernoulli-Versuchen an.

Scharfes Hinsehen liefert $P_{X_1} = P_{Y_1}$ und $P_{X_2} = P_{Y_2}$, dann gilt auch

$$P_{X_1+X_2} = P_{Y_1+Y_2} = b(n_1 + n_2, p). \quad \ll$$

Somit folgt für unabhängige Zufallsvariablen X_1, \dots, X_n , die $b(1, p)$ -verteilt sind, dass $X_1 + \dots + X_n$ eine $b(n, p)$ -verteilte Zufallsvariable ist. ■

BSP 5 Wir können jetzt einen einfachen Beweis dafür geben, dass eine $b(n, p)$ -verteilte Zufallsvariable die Varianz $np(1 - p)$ besitzt. Ohne Einschränkung lässt sich diese Zufallsvariable als Summe von n unabhängigen $b(1, p)$ -verteilten Zufallsvariablen X_1, \dots, X_n darstellen,

$$\mathbf{V}(X_1 + \dots + X_n) \stackrel{5.5}{=} n\mathbf{V}(X_1) = np(1 - p),$$

da $\mathbf{V}(X_1) = \mathbf{E}X_1^2 - (\mathbf{E}X_1)^2 = 0^2 \cdot (1 - p) + 1^2 \cdot p - p^2 = p(1 - p)$. ■

Die Unabhängigkeit der Zufallsvariablen ist für die Aussage des Satzes 5.6 notwendig, denn er basiert darauf, dass für unabhängige Zufallsvariablen die gemeinsame Dichte das Produkt der Randverteilungsdichten ist. Für allgemeine (nicht unabhängige) Zufallsvariablen wird diese Aussage und damit der Satz falsch.

5.6 *Bemerkung (Satz von Andersen und Jessen).* Gegeben seien Messräume $(\Omega_i, \mathcal{A}_i)$ und W -Maße Q_i auf \mathcal{A}_i , $i \in I$. Dann existiert ein W -Raum (Ω, \mathcal{A}, P) und $(\Omega_i, \mathcal{A}_i)$ -Zufallsvariablen X_i auf (Ω, \mathcal{A}, P) , $i \in I$, mit Unabhängigkeit von $\{X_i : i \in I\}$ und $P_{X_i} = Q_i$. \rightarrow

5-B Null-Eins-Gesetz

In Kapitel 1-B haben wir das 1. Lemma von Borel und Cantelli bewiesen, welches für einen W -Raum (Ω, \mathcal{A}, P) und $A_n \in \mathcal{A}$ die Aussage macht

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(\limsup A_n) = 0.$$

Das 2. Lemma von Borel und Cantelli ist für unabhängige Ereignisse das Gegenstück zu dieser Aussage.

5.7 **2. Lemma von Borel und Cantelli** Sei (Ω, \mathcal{A}, P) ein W -Raum. Die Familie $\{A_n : n \in \mathbb{N}\}$ von Ereignissen $(\in \mathcal{A})$ sei unabhängig. Dann gilt:

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P(\limsup A_n) = 1.$$

In diesem Fall tritt A_n P -f.s. unendlich oft auf. \times

» Sei $B = \limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k$. Zu zeigen ist nun, dass $P(B) = 1$, d.h. $P(B^c) = 0$. Mit den De-Morganschen Regeln folgt,

$$P(B^c) = P\left(\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} A_k^c\right).$$

Aufgrund der σ -Additivität genügt es nun zu zeigen, dass die Mengen $\bigcap_{k \geq n} A_k^c$ das P -Maß Null haben.

Sei also $n \in \mathbb{N}$ fest.

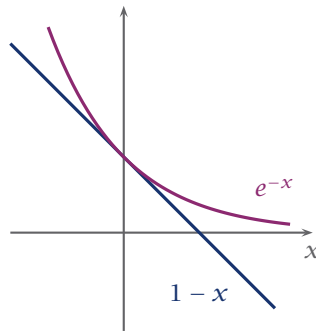
$$\bigcap_{k=n}^N A_k^c \downarrow \bigcap_{k \geq n} A_k^c, \quad N \rightarrow \infty.$$

Somit genügt es wegen der Stetigkeit der W -Maße von oben zu zeigen, dass

$$P\left(\bigcap_{k=n}^N A_k^c\right) \downarrow 0.$$

Aufgrund der Unabhängigkeit der A_1, A_2, \dots gilt

$$P\left(\bigcap_{k=n}^N A_k^c\right) = \prod_{k=n}^N P(A_k^c) = \prod_{k=n}^N (1 - P(A_k)).$$



5.1 e^{-x} mit Tangente $1-x$ in $(0, 1)$.

Eine geometrische Überlegung zeigt $1-x \leq e^{-x}$ (siehe Skizze), also gilt auch $1 - P(A_k) \leq e^{-P(A_k)}$ und somit

$$P\left(\bigcap_{k=n}^N A_k^c\right) \leq e^{-\sum_{k=n}^N P(A_k)} \rightarrow 0, \quad N \rightarrow \infty. \quad \ll$$

Korollar Sind die Ereignisse A_1, A_2, \dots unabhängig, gilt

$$P(\limsup A_n) = 0 \text{ oder } 1,$$

abhängig davon ob

$$\sum_{n \geq 1} P(A_n) < \infty \text{ oder } \sum_{n \geq 1} P(A_n) = \infty. \quad \times$$

Die Voraussetzung der Unabhängigkeit der (A_n) im 2. Lemma von Borel-Cantelli ist notwendig, damit die Aussage gilt, wie folgendes Beispiel zeigt.

BSP 6 Sei $A \in \mathcal{A}$ mit $P(A) = \frac{1}{2}$ und $\forall n \in \mathbb{N} : A_n = A$. Dann gilt $\limsup A_n = A$ und $\sum_{n \geq 1} P(A_n) = \infty$. Dennoch ist

$$P(\limsup A_n) = P(A) = \frac{1}{2}. \quad \blacksquare$$

BSP 7 Betrachte eine unabhängige Folge von Versuchen A_n mit Misserfolgswahrscheinlichkeit $\frac{1}{n}$. Da

$$\sum_{n \geq 1} P(A_n) = \infty,$$

folgt, dass die Wahrscheinlichkeit für unendlich viele Misserfolge gleich 1 ist (obwohl $P(A_n) \rightarrow 0$). ■

5.4 **Definition** Sei (Ω, \mathcal{A}) ein Messraum und $\mathcal{A}_n \subset \mathcal{A}$ σ -Algebren für $n \in \mathbb{N}$. Mit

$$\mathcal{T}_n := \mathcal{F} \left(\bigcup_{k=n}^{\infty} \mathcal{A}_k \right)$$

bezeichnen wir die von $\mathcal{A}_n, \mathcal{A}_{n+1}, \dots$ erzeugte σ -Algebra.

$$\mathcal{T}_\infty := \bigcap_{n=1}^{\infty} \mathcal{T}_n$$

heißt die σ -Algebra der *terminalen Ereignisse (tail events)* der Folge (\mathcal{A}_n) . ✕

BSP 8 Seien $X_n : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ und $\mathcal{A}_n := \mathcal{F}(X_n) = X_n^{-1}(\mathcal{B})$ für $n \in \mathbb{N}$. \mathcal{T}_n ist somit die von den Ereignissen, die lediglich von X_n, X_{n+1}, \dots abhängen, erzeugte σ -Algebra.

a.) $[(X_n) \text{ konvergiert}] \in \mathcal{T}_\infty$, denn es gilt

$$\begin{aligned} [(X_n) \text{ konvergiert}] &= \left[\lim_{n \rightarrow \infty} X_n = X \in \mathbb{R} \right] = \left[\lim_{n \rightarrow \infty} |X_n - X| = 0 \right] \\ &= [\forall \varepsilon > 0 \exists N \in \mathbb{N} \forall n \geq N : |X_n - X| < \varepsilon] \end{aligned}$$

\mathbb{Q} liegt dicht in \mathbb{R} , wir können uns also für die Wahl von ε auf \mathbb{Q} zurückziehen,

$$[\forall \varepsilon \in \mathbb{Q} > 0 \exists N \in \mathbb{N} \forall n \geq N : |X_n - X| < \varepsilon]$$

da alle $\varepsilon > 1$ uninteressant sind, können wir auch schreiben

$$\begin{aligned} & \left[\forall k \in \mathbb{N} \exists N \in \mathbb{N} \forall n \geq N : |X_n - X| < \frac{1}{k} \right] \\ &= \bigcap_{k \in \mathbb{N}} \left[\exists N \in \mathbb{N} \forall n \geq N : |X_n - X| < \frac{1}{k} \right] \end{aligned}$$

und da es auf endlich viele N nicht ankommt,

$$\begin{aligned} &= \bigcap_{k \in \mathbb{N}} \bigcup_{N \geq k} \left[\forall n \geq N : |X_n - X| < \frac{1}{k} \right] \\ &= \bigcap_{k \in \mathbb{N}} \underbrace{\bigcup_{N \geq k} \left[\sup_{n \geq N} |X_n - X| < \frac{1}{k} \right]}_{\in \mathcal{T}_k}. \end{aligned}$$

Also handelt es sich um ein terminales Ereignis.

b.) $[\sum_n X_n \text{ konvergiert}] \in \mathcal{T}_\infty$, denn

$$\left[\sum_n X_n \text{ konvergiert} \right] = \left[\sum_{n \geq k} X_n \text{ konvergiert} \right], \quad \forall k \in \mathbb{N}$$

da es auf endlich viele nicht ankommt also $[\sum_n X_n \text{ konvergiert}] \in \mathcal{T}_k$ für $k \in \mathbb{N}$.

■

BSP 9 Sei (Ω, \mathcal{A}) ein Messraum, $A_n \in \mathcal{A}$ Ereignisse ($n \in \mathbb{N}$) und weiterhin $\mathcal{A}_n := \{\emptyset, A_n, A_n^c, \Omega\}$ ($n \in \mathbb{N}$).

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \text{ für unendliche viele } n\} \in \mathcal{T}_\infty,$$

$$\liminf_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \{\omega \in \Omega : \omega \in A_n \text{ von einem Index an}\} \in \mathcal{T}_\infty. \quad \blacksquare$$

5.8 **Null-Eins-Gesetz von Kolmogorov** Sei (Ω, \mathcal{A}, P) ein W-Raum und (\mathcal{A}_n) eine unabhängige Folge von σ -Algebren $\mathcal{A}_n \subset \mathcal{A}$. Dann gilt für jedes terminale Ereignis A von (\mathcal{A}_n) entweder $P(A) = 0$ oder $P(A) = 1$. \times

Für den Beweis benötigen wir das folgende Resultat.

Lemma Seien \mathcal{E}_1 und \mathcal{E}_2 *schnittstabile* Mengensysteme von Ereignissen, d.h.

$$A, B \in \mathcal{E}_1, \mathcal{E}_2 \Rightarrow A \cap B \in \mathcal{E}_1, \mathcal{E}_2.$$

Sind \mathcal{E}_1 und \mathcal{E}_2 unabhängig, so sind es auch $\mathcal{F}(\mathcal{E}_1)$ und $\mathcal{F}(\mathcal{E}_2)$. \times

» Den Beweis findet man z.B. in [Wengenroth]. «

» *Beweis von Satz 5.8.* Sei also $A \in \mathcal{T}_\infty$. Es genügt zu zeigen, dass A von sich selbst unabhängig ist, d.h.

$$P(A \cap A) = P(A)P(A),$$

denn dann ist $P(A) = 0$ oder 1. Betrachte dazu die Mengen

$$\mathcal{T}_{n+1} = \mathcal{F}\left(\bigcup_{k \geq n+1} \mathcal{A}_k\right),$$

$$\mathcal{D}_n := \mathcal{F}\left(\bigcup_{k \leq n} \mathcal{A}_k\right),$$

und wende das vorige Lemma an, so sind \mathcal{T}_{n+1} und \mathcal{D}_n unabhängig.

Weiterhin sind aus dem selben Grund \mathcal{T}_∞ und \mathcal{D}_n unabhängig für alle $n \in \mathbb{N}$.
Daher ist auch

$$\mathcal{G} = \mathcal{F} \left(\bigcup_{n \in \mathbb{N}} \mathcal{D}_n \right)$$

unabhängig von \mathcal{T}_∞ , wobei gerade $\mathcal{G} = \mathcal{T}_1 \supset \mathcal{T}_\infty$. Somit ist \mathcal{T}_∞ unabhängig von jeder Teilmenge von \mathcal{T}_1 insbesondere von sich selbst. Somit folgt die Behauptung.

«

Wir betrachten nun einige Anwendungen des 0-1-Gesetzes.

BSP 10 Sei $\{A_n : n \in \mathbb{N}\}$ unabhängig, $\mathcal{A}_n = \{\emptyset, A_n, A_n^c, \Omega\}$, somit sind

$$\mathcal{A}_1, \mathcal{A}_2, \dots$$

unabhängig. Mit dem 0-1-Gesetz gilt daher,

$$P \left(\underbrace{\limsup A_n}_{\in \mathcal{T}_\infty} \right) = 0 \text{ oder } 1.$$

Diese Aussage erhalten wir auch mit dem 1. und 2. Lemma von Borel und Cantelli. Insofern kann man das 0-1-Gesetz als Verallgemeinerung dieser Lemmata betrachten. Die Aussagen von Borel und Cantelli liefern jedoch darüber hinaus noch ein Kriterium, wann die Voraussetzungen des 0-1-Gesetzes erfüllt sind. ■

BSP 11 Sei $\{X_n : n \in \mathbb{N}\}$ eine Folge reeller unabhängiger Zufallsvariablen, dann gilt

$$P[(X_n) \text{ konvergiert}] = 0 \text{ oder } 1,$$

$$P \left[\sum_n X_n \text{ konvergiert} \right] = 0 \text{ oder } 1. \quad \blacksquare$$

6 Erzeugende und charakteristische Funktionen, Faltungen

Die erzeugenden und charakteristischen Funktionen sind ein methodisches Hilfsmittel zur Untersuchung von W-Maßen.

Bisher haben wir gesehen, dass Verteilungen und Verteilungsfunktionen manchmal schwierig “zu fassen” sind. Erzeugende und charakteristische Funktionen ermöglichen den Übergang von Verteilungen definiert auf einer σ -Algebra zu Funktionen definiert auf \mathbb{R} . Dadurch erhalten wir die Möglichkeit, Konzepte aus der Analysis wie Integration, Differentiation und Grenzwertbildung anzuwenden.

6-A Erzeugende Funktionen

- 6.1 **Definition** *Es sei $\mu : \mathcal{B} \rightarrow \mathbb{R}$ ein auf \mathbb{N}_0 konzentriertes W-Maß mit Zähldichte $(b_k)_{k \in \mathbb{N}_0}$ bzw. X eine reelle Zufallsvariable auf einem W-Raum (Ω, \mathcal{A}, P) , deren Verteilung P_X auf \mathbb{N}_0 konzentriert ist, mit Zähldichte $(b_k)_{k \in \mathbb{N}_0}$. Dann heißt die auf $[0, 1]$ (oder auch $\{s \in \mathbb{C} \mid |s| \leq 1\}$) definierte Funktion g mit*

$$g(s) := \sum_{k=0}^{\infty} b_k s^k = \begin{cases} \sum \mu(\{k\}) s^k, \\ \text{bzw.} \\ \sum P[X = k] s^k = \sum P_X(\{k\}) s^k = \mathbf{E}s^X. \end{cases}$$

die *erzeugende Funktion* von μ bzw. X . \times

- 6.1 **Bemerkung.** In Definition 6.1 gilt $|g(s)| \leq \sum b_k |s|^k \leq \sum b_k = 1$ für $|s| \leq 1$, da es sich bei (b_k) um eine Zähldichte handelt. ∞

Bei der erzeugenden Funktion handelt es sich also um die formale Potenzreihe

$$g(s) = \sum_{k=0}^{\infty} b_k s^k,$$

die nach obiger Bemerkung auf dem Intervall $[0, 1]$ bzw. der abgeschlossenen Kreisscheibe $\{s \in \mathbb{C} : |s| \leq 1\}$ absolut und gleichmäßig bezüglich s konvergiert.

Von besonderem Interesse sind die erzeugenden Funktionen, die neben der Darstellung als Potenzreihe auch über eine explizite Darstellung verfügen, denn dann lassen sich $g(s)$, $g'(s)$, ... meist leicht berechnen und man erspart sich kombinatorische Tricks, um die Reihen auf eine Form mit bekanntem Wert zu bringen.

6.2 *Bemerkungen.* A. Die Binomialverteilung $b(n, p)$ mit $n \in \mathbb{N}$ und $p \in [0, 1]$ hat mit $q := 1 - p$ die erzeugende Funktion g ,

$$\begin{aligned} g(s) &:= \sum_{k=0}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} s^k = \sum_{k=0}^n \binom{n}{k} (p \cdot s)^k (1-p)^{n-k} \\ &= (p \cdot s + (1-p))^n = (ps + q)^n. \end{aligned}$$

Die Reihe lässt sich im Gegensatz zu dem Ausdruck $(ps + q)^n$ nur mit einigem Aufwand direkt berechnen.

B. Die Poissonverteilung $\pi(\lambda)$ mit $\lambda \in (0, \infty)$ hat die erzeugende Funktion g mit

$$g(s) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k = e^{-\lambda} \sum_{k=0}^{\infty} e^{-\lambda} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{s\lambda} = e^{-\lambda(1-s)}.$$

C. Als **negative Binomialverteilung** oder **Pascal-Verteilung** $Nb(r, p)$ mit Parametern $r \in \mathbb{N}$, $p \in (0, 1)$ wird ein W-Maß auf \mathcal{B} (oder auch $\overline{\mathcal{B}}$) bezeichnet, das auf \mathbb{N}_0 konzentriert ist und — mit $q := 1 - p$ — die Zähldichte

$$k \rightarrow Nb(r, p; k) := \binom{r+k-1}{k} p^r q^k, \quad k \in \mathbb{N}_0$$

besitzt. Speziell wird $Nb(1, p)$ — mit Zähldichte

$$k \rightarrow p(1-p)^k, \quad k \in \mathbb{N}_0$$

— als **geometrische Verteilung** mit Parameter $p \in (0, 1)$ bezeichnet.

Ist $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Folge von $b(1, p)$ -verteilten Zufallsvariablen, so ist die erweitert-reelle Zufallsvariable

$$X := \inf \left\{ n \in \mathbb{N} : \sum_{k=1}^n X_k = r \right\} - r,$$

mit der Konvention $\inf \emptyset := \infty$, $Nb(r, p)$ -verteilt. X gibt die Anzahl der Misserfolge bis zum r -ten Erfolg bei der zu (X_n) gehörigen Folge von Bernoulli-Versuchen an.

$$\mathbf{E}X = \frac{rq}{p}, \mathbf{V}(X) = \frac{rq}{p^2}.$$

$Nb(r, p)$ hat die erzeugende Funktion g mit $g(s) = p^r (1 - qs)^{-r}$. \rightarrow

6.1 Eindeutigkeitssatz für erzeugende Funktionen *Besitzen zwei auf \mathcal{B} definierte, aber auf \mathbb{N}_0 konzentrierte W -Maße bzw. Verteilungen dieselbe erzeugende Funktion, so stimmen sie überein. \times*

» Dies folgt aus dem Identitätssatz für Potenzreihen, der besagt, dass zwei auf einer nichtleeren offenen Menge identische Potenzreihen dieselben Koeffizienten haben. «

Es besteht also eine Bijektion zwischen diskreten Verteilungen und erzeugenden Funktionen. Insbesondere haben zwei diskrete Zufallsvariablen mit derselben erzeugenden Funktion auch dieselbe Verteilung.

6.2 Satz *Sei X eine reelle Zufallsvariable, deren Verteilung auf \mathbb{N}_0 konzentriert ist, mit erzeugender Funktion g .*

a) *g ist auf der offenen Kreisscheibe $\{s \in \mathbb{C} : |s| < 1\}$ unendlich oft differenzierbar, sogar analytisch und für $j \in \mathbb{N}$ gilt*

$$g^{(j)}(1-) := \lim_{0 \leq s \uparrow 1} g^{(j)}(s) = \mathbf{E}[X(X-1) \dots (X-j+1)] \quad (\leq \infty),$$

insbesondere $g'(1-) = \mathbf{E}X$.

b) *Falls $\mathbf{E}X < \infty$, so gilt $\mathbf{V}(X) = g''(1-) + g'(1-) - (g'(1-))^2$. \times*

- » a) Die Differenzierbarkeit bzw. die Analytizität auf der offenen Kreisscheibe wird in der Funktionentheorie bewiesen. Auf dieser Kreisscheibe können wir gliedweise differenzieren und erhalten somit,

$$g^{(j)}(s) = \sum_{k=j}^{\infty} b_k k(k-1) \cdots (k-j+n) s^{k-j}, \quad 0 \leq s < 1.$$

Auf dem Rand gehen wir gegebenenfalls zum Grenzwert über und erhalten mit dem Satz von der monotonen Konvergenz für $s \uparrow 1$,

$$g^{(j)}(s) \rightarrow \sum_{k=j}^{\infty} b_k k(k-1) \cdots (k-j+n).$$

Diese Reihe können wir auch als Erwartungswert der Zufallsvariablen

$$X(X-1) \cdots (X-j+1)$$

betrachten (siehe Satz 4.5), d.h.

$$\lim_{s \uparrow 1} g^{(j)}(s) = \mathbf{E}(X(X-1) \cdots (X-j+1)).$$

Insbesondere erhalten wir $g'(1-) = \mathbf{E}X$.

- b) Da X integrierbar, können wir die Varianz darstellen als

$$\begin{aligned} \mathbf{V}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 = \mathbf{E}(X(X-1)) + \mathbf{E}X - (\mathbf{E}X)^2 \\ &= g''(1-) + g'(1-) - (g'(1-))^2. \quad \ll \end{aligned}$$

Kennen wir eine explizite Darstellung der erzeugenden Funktion einer diskreten Zufallsvariablen, lassen sich mit Satz 6.2 ihre Momente oftmals leicht berechnen.

- 6.3 **Satz** Seien X_1, \dots, X_n unabhängige reelle Zufallsvariablen, deren Verteilungen jeweils auf \mathbb{N}_0 konzentriert sind, mit erzeugenden Funktionen g_1, \dots, g_n . Für die erzeugende Funktion g der Summe $X_1 + \dots + X_n$ gilt dann $g = \prod_{j=1}^n g_j$. \times

- » Sei $|s| < 1$, dann gilt *per definitionem*

$$g(s) = \sum_{k=0}^{\infty} P[X_1 + \dots + X_n = k] s^k.$$

Nun interpretieren wir die erzeugende Funktion als Erwartungswert wie in 6.1, d.h.

$$g(s) = \mathbf{E} s^{X_1 + \dots + X_n} = \mathbf{E} \prod_{i=1}^n s^{X_i}.$$

Da X_1, \dots, X_n unabhängig, sind nach Satz 5.4 auch s^{X_1}, \dots, s^{X_n} unabhängig, d.h.

$$g(s) = \prod_{i=1}^n \mathbf{E} s^{X_i} = \prod_{i=1}^n g_i(s). \quad \ll$$

BSP 1 Seien X_1, X_2, \dots unabhängige $b(1, p)$ verteilte Zufallsvariablen, mit

$$P[X_i = 1] = p, \quad P[X_i = 0] = \underbrace{1 - p}_q,$$

so können wir sie als Folge von Bernoulli-Versuchen interpretieren mit jeweiliger Erfolgswahrscheinlichkeit p ,

$$[X_i = 1], \quad \text{Erfolg im } i\text{-ten Versuch.}$$

Die erweiterte reellwertige Zufallsvariable Z_1 gebe die Anzahl der Misserfolge bis zum 1. Erfolg an. Z_1 ist erweitert reellwertig, da die Möglichkeit besteht, dass Z_1 unendlich wird.

Nach dem 1. Erfolg wiederholt sich die stochastische Situation. Z_2 gebe die Anzahl der Misserfolge nach dem 1. bis zum 2. Erfolg an.

Betrachte nun eine Realisierung

$$\underbrace{0, 0, 0, 0}_Z, \underbrace{1, 0, 0}_Z, \underbrace{1}_Z, 1, 0, \dots$$

Die Zufallsvariablen Z_1, Z_2, \dots sind unabhängig und identisch verteilt.

» *Nachweis der Unabhängigkeit und der identischen Verteilung.* Wir zeigen zuerst die identische Verteilung,

$$P[Z_1 = k] = q^k p, \quad k \in \mathbb{N}_0$$

$$P[Z_1 = \infty] = 1 - \sum_{k=0}^{\infty} q^k p = 1 - p \frac{1}{1 - q} = 0, \quad \text{für } p > 0,$$

$$\begin{aligned} P[Z_2 = k] &= \sum_{j=1}^{\infty} P[Z_1 = j, Z_2 = k] = \sum_{j=1}^{\infty} q^j p q^k p \\ &= p^2 q^k \frac{1}{1 - q} = q^k p, \end{aligned}$$

$$P[Z_2 = \infty] = \dots = 0.$$

Induktiv erhalten wir

$$\forall j, k \in \mathbb{N}_0 : P[Z_1 = j, Z_2 = k] = q^j p q^k p = P[Z_1 = j]P[Z_2 = k]$$

insbesondere sind Z_1 und Z_2 unabhängig. «

Z_1 hat die erzeugende Funktion,

$$\begin{aligned} h(s) &= \sum_{k=0}^{\infty} q^k p s^k = p \frac{1}{1 - qs} = p(1 - qs)^{-1}, \quad |s| < 1, \\ h'(s) &= pq(1 - qs)^{-2}, \quad h'(1-) = \frac{pq}{(1 - q)^2} = \frac{q}{p}, \\ h''(s) &= 2pq^2(1 - qs)^{-3}, \quad h''(1-) = \frac{2pq^2}{(1 - q)^3} = \frac{2q^2}{p^2}. \end{aligned}$$

Nach Satz 6.2 gilt somit

$$\mathbf{E}Z_1 = \frac{q}{p}, \quad \mathbf{V}Z_1 = \frac{2q^2}{p^2} + \frac{q}{p} - \left(\frac{q}{p}\right)^2 = \frac{q(q+p)}{p^2} = \frac{q}{p^2}.$$

Sei nun $r \in \mathbb{N}$ fest, so hat die Zufallsvariable

$$X := \sum_{i=1}^r Z_i$$

nach Satz 6.3 die erzeugende Funktion $g = h^r$, also

$$g(s) = h(s)^r = p^r (1 - qs)^{-r}.$$

Aus der Analysis kennen wir folgenden Spezialfall der binomischen Reihe,

$$(1 + x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k, \quad |x| < 1, \quad \binom{\alpha}{k} = \frac{\alpha(\alpha - 1) \cdots (\alpha - k + 1)}{k!}.$$

Diese Definition ist auch für $\alpha \in \mathbb{R}$ sinnvoll. Wir können somit g darstellen als

$$\begin{aligned} g(s) &= \sum_{k=0}^{\infty} p^r \binom{-r}{k} (-1)^k q^k s^k = \sum_{k=0}^{\infty} p^r \binom{r+k-1}{k} q^k s^k, \\ \stackrel{6.1}{\Rightarrow} P[X = k] &= \binom{r+k-1}{k} p^r q^k, \quad k \in \mathbb{N}_0. \end{aligned}$$

Die Zufallsvariable X gibt die Anzahl der Misserfolge bis zum r -ten Erfolg an. Also ist eine äquivalente Definition von X gegeben durch

$$X := \inf \left\{ n \in \mathbb{N} : \sum_{k=1}^n X_k = r \right\} - r$$

wobei X_k angibt, ob im k -ten Versuch ein Erfolg aufgetreten ist.

Eine Zufallsvariable, welche die Anzahl der Misserfolge bis zum r -ten Erfolg bei Bernoulli-Versuchen mit Erfolgswahrscheinlichkeit $p \in (0, 1)$ angibt, hat somit die Zähldichte

$$\binom{r+k-1}{k} p^r q^k, \quad k \in \mathbb{N}_0, r \in \mathbb{N}, p \in (0, 1),$$

wobei $N_b(r, p; k) := \binom{-r}{k} (-1)^k p^r q^k = \binom{r+k-1}{k} p^r q^k$. $N_b(r, p)$ heißt **negative Binomial-Verteilung** oder **Pascal-Verteilung** mit Parametern $r \in \mathbb{N}$, $p \in (0, 1)$.

Für $r = 1$ heißt $N_b(1, p)$ mit Zähldichte $(pq^k)_{k \in \mathbb{N}}$ **geometrische Verteilung** mit Parameter $p \in (0, 1)$. Wir können nun die negative Binomial-Verteilung darstellen als

$$N_b(r, p) = \underbrace{N_b(1, p) * \dots * N_b(1, p)}_{r\text{-mal}}. \quad \blacksquare$$

■ Konvergenzsätze der Maßtheorie

Wir haben bereits den Satz von der monotonen Konvergenz kennengelernt, der eine Aussage über Vertauschbarkeit von Grenzwertbildung und Integration für positive, monotone Funktionenfolgen macht. Im Allgemeinen sind die Funktionenfolgen, mit denen wir in der Wahrscheinlichkeitstheorie arbeiten weder monoton noch positiv, wir benötigen also allgemeinere Aussagen.

Lemma von Fatou Sei (Ω, \mathcal{A}, P) ein Maßraum. Für jede Folge von nichtnegativen erweitert reellen messbaren Funktionen f_n gilt,

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

Gilt außerdem $|f_n| \leq g$ mit $\int_{\Omega} g \, d\mu < \infty$, so gilt auch

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

» Wir beweisen lediglich die erste Ungleichung, die andere ist trivial. Setze

$$g_n := \inf_{m \geq n} f_m \uparrow \liminf_n f_n =: g.$$

Auf diese Funktionenfolge können wir den Satz von der monotonen Konvergenz anwenden und erhalten somit,

$$\begin{aligned} \int_{\Omega} \liminf f_n \, d\mu &= \lim_{n \rightarrow \infty} \int_{\Omega} \inf_{m \geq n} f_n \, d\mu \leq \lim_{n \rightarrow \infty} \inf_{m \geq n} \int_{\Omega} f_n \, d\mu \\ &= \liminf_n \int_{\Omega} f_n \, d\mu. \quad \ll \end{aligned}$$

Man kann das Lemma von Fatou als Verallgemeinerung des Satzes von der monotonen Konvergenz von monotonen nichtnegativen auf lediglich nichtnegative Funktionenfolgen betrachten. Dadurch wird jedoch auch die Aussage auf “ \leq ” anstatt “ $=$ ” abgeschwächt.

Satz von der dominierten Konvergenz Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f_n, f, g erweitert reellwertige messbare Funktionen mit $f_n \rightarrow f$ μ -f.ü., $|f_n| \leq g$ μ -f.ü. sowie $\int_{\Omega} g \, d\mu < \infty$. Dann existiert

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

und es gilt

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

Man findet den Satz in der Literatur auch unter dem Namen Satz von der majorisierten Konvergenz bzw. Satz von Lebesgue.

» Setze $N := [f_n \neq f] \cup \bigcup_{n \in \mathbb{N}} [|f_n| > g] \cup [g = \infty] \in \mathcal{A}$, so ist nach Voraussetzung $\mu(N) = 0$. Setze nun

$$\tilde{f}_n := f_n \cdot \mathbf{1}_{N^c}, \quad \tilde{f} := f \cdot \mathbf{1}_{N^c}, \quad \tilde{g} := g \cdot \mathbf{1}_{N^c}.$$

Dann gilt $\tilde{f}_n + \tilde{g} \geq 0$, $\tilde{g} - \tilde{f}_n \geq 0$ auf ganz Ω und für alle $n \in \mathbb{N}$. Mit dem Lemma

von Fatou folgt,

$$\begin{aligned}
 \int_{\Omega} \tilde{f} \, d\mu &= \int_{\Omega} \liminf_n (\tilde{f}_n + \tilde{g}) \, d\mu - \int_{\Omega} \tilde{g} \, d\mu \\
 &\leq \liminf_n \int_{\Omega} (\tilde{f}_n + \tilde{g}) \, d\mu - \int_{\Omega} \tilde{g} \, d\mu = \liminf_n \int_{\Omega} \tilde{f}_n \, d\mu \\
 &\leq \limsup_n \int_{\Omega} \tilde{f}_n \, d\mu = - \liminf_n \int_{\Omega} -\tilde{f}_n \, d\mu \\
 &= \int_{\Omega} \tilde{g} \, d\mu - \liminf_n \int_{\Omega} (\tilde{g} - \tilde{f}_n) \, d\mu \\
 &\leq \int_{\Omega} \tilde{g} \, d\mu - \int_{\Omega} \liminf_n (\tilde{g} - \tilde{f}_n) \, d\mu = \int_{\Omega} \tilde{f} \, d\mu.
 \end{aligned}$$

Insbesondere gilt $\liminf_n \int_{\Omega} \tilde{f}_n \, d\mu = \limsup_n \int_{\Omega} \tilde{f}_n \, d\mu$ und somit ist alles gezeigt.

«

6-B Charakteristische Funktionen

6.2 Definition Sei μ ein W -Maß auf \mathcal{B} bzw. X eine reelle Zufallsvariable mit Verteilungsfunktion F . Dann heißt die auf \mathbb{R} definierte (i.a. komplexwertige) Funktion φ mit

$$\varphi(u) := \int_{\mathbb{R}} e^{iux} \mu(dx)$$

bzw.

$$\varphi(u) := \mathbf{E}e^{iuX} = \int_{\mathbb{R}} e^{iux} P_X(dx) = \int_{\mathbb{R}} e^{iux} dF(x)$$

die *charakteristische* Funktion von μ bzw. X . \times

Diese Definition ist tatsächlich sinnvoll, denn wir können e^{iux} für $u, x \in \mathbb{R}$ stets durch 1 majorisieren. Es gilt also

$$|\varphi(u)| \leq \mathbf{E} |e^{iuX}| \leq \mathbf{E}1 = 1$$

und daher existiert $\varphi(u)$ für jedes $u \in \mathbb{R}$.

6.3 Bemerkungen. In Definition 6.2 gilt

A. $\varphi(0) = 1,$

B. $|\varphi(u)| \leq 1, u \in \mathbb{R},$

C. φ gleichmäßig stetig in \mathbb{R} ,

D. $\varphi(-u) = \overline{\varphi(u)}$, $u \in \mathbb{R}$. \rightarrow

» Nachweis von Bemerkung 6.3 c)

$$\begin{aligned} |\varphi(u+h) - \varphi(u)| &= \left| \int_{\mathbb{R}} e^{i(u+h)x} - e^{iux} \mu(dx) \right| = \left| \int_{\mathbb{R}} (e^{ihx} - 1) e^{iux} \mu(dx) \right| \\ &\leq \int_{\mathbb{R}} |e^{ihx} - 1| \mu(dx) \end{aligned}$$

Nun verschwindet der Integrand für $h \rightarrow 0$, d.h. für jede Nullfolge h_n gilt

$$|e^{ih_n x} - 1| \rightarrow 0, \quad n \rightarrow \infty.$$

Außerdem ist $|e^{ih_n x} - 1| \leq |e^{ih_n x}| + 1 = 2$, und $\int_{\mathbb{R}} 2 d\mu = 2\mu(\mathbb{R}) = 2$, wir können also den Satz von der dominierten Konvergenz anwenden. \leftarrow

Die charakteristische Funktion ist im Wesentlichen die Inverse Fouriertransformierte von P_X , der Verteilung von X . Besitzt X eine Dichte f , so gilt nach dem Transformationssatz

$$\varphi_X(u) = \int_{\mathbb{R}} e^{iux} f(x) dx.$$

6.4 *Bemerkung.* Die Normalverteilung $N(a, \sigma^2)$ mit $a \in \mathbb{R}$ und $\sigma^2 \in (0, \infty)$ besitzt die charakteristische Funktion $\varphi : \mathbb{R} \rightarrow \mathbb{C}$ mit

$$\varphi(u) = e^{iau} e^{-\frac{\sigma^2 u^2}{2}}.$$

Insbesondere besitzt die Standardnormalverteilung $N(0, 1)$ die charakteristische Funktion φ mit $\varphi(u) = e^{-\frac{u^2}{2}}$. \rightarrow

» *Beweisskizze.* 1. Teil. Wir betrachten zunächst $N(0, 1)$. Die Standardnormalverteilung besitzt die Dichte

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

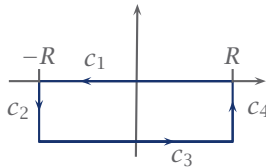
somit gilt für die charakteristische Funktion,

$$\begin{aligned} \varphi(u) &= \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i(x-iu)^2 - \frac{1}{2}u^2} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \int_{\mathbb{R}} e^{\frac{1}{2}(x-iu)^2} dx. \end{aligned}$$

Um das Integral zu berechnen, betrachten wir $z \mapsto e^{\frac{1}{2}z^2}$; diese Funktion einer komplexen Variablen z ist komplex differenzierbar (analytisch, holomorph). Wir verwenden nun den Residuensatz, ein Resultat aus der Funktionentheorie, der besagt, dass das Integral einer holomorphen Funktion über jede geschlossene Kurve c verschwindet,

$$\int_c e^{\frac{1}{2}z^2} dz = 0.$$

Um dies auszunutzen, wählen wir eine spezielle geschlossene Kurve c (siehe Skizze) bestehend aus den Geradenstücken c_1, c_2, c_3 und c_4 .



6.1 Zur Konstruktion der geschlossenen Kurve c .

Für diese Geradenstücke können wir die Integrale leicht berechnen bzw. abschätzen,

$$\begin{aligned} \int_{c_1} e^{-\frac{1}{2}z^2} dz &\stackrel{R \rightarrow \infty}{\rightarrow} -\sqrt{2\pi}, \\ \int_{c_2, c_4} e^{-\frac{1}{2}z^2} dz &= \int_{c_2} e^{-\frac{1}{2}z^2} dz + \underbrace{\int_{c_4} e^{-\frac{1}{2}z^2} dz}_{-\int_{c_2} e^{-\frac{1}{2}z^2} dz} = 0, \\ \int_{c_3} e^{-\frac{1}{2}z^2} dz &\stackrel{R \rightarrow \infty}{\rightarrow} \int_{\mathbb{R}} e^{-\frac{1}{2}(x-iu)^2} dx. \end{aligned}$$

Für ganz c gilt nach Voraussetzung,

$$\begin{aligned} 0 &= \int_c e^{-\frac{1}{2}z^2} dz = \sum_{i=1}^4 \int_{c_i} e^{-\frac{1}{2}z^2} dz = \int_{c_1} e^{-\frac{1}{2}z^2} dz + \int_{c_3} e^{-\frac{1}{2}z^2} dz \\ &\Rightarrow \int_{\mathbb{R}} e^{-\frac{1}{2}(x-iu)^2} dx = \sqrt{2\pi}. \end{aligned}$$

Somit besitzt $N(0, 1)$ die charakteristische Funktion

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \int_{\mathbb{R}} e^{-\frac{1}{2}(x-iu)^2} dx = e^{-\frac{1}{2}u^2}.$$

2. Teil. Für $N(a, \sigma^2)$ allgemein ist die charakteristische Funktion gegeben durch,

$$\varphi(u) = \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} dx$$

Die Substitution $y = \frac{x-a}{\sigma}$ liefert die Behauptung. «

6.5 *Bemerkung.* Als **Exponentialverteilung** $\exp(\lambda)$ mit Parameter $\lambda \in (0, \infty)$ wird ein W-Maß auf \mathcal{B} oder $\overline{\mathcal{B}}$ bezeichnet, das eine Dichtefunktion f mit

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

besitzt.

Die zufällige Lebensdauer eines radioaktiven Atoms wird durch eine exponentialverteilte Zufallsvariable angegeben. Ist X eine erweitert-reelle Zufallsvariable, deren Verteilung auf $\overline{\mathbb{R}}_+$ konzentriert ist, mit $P[0 < X < \infty] > 0$, so gilt

$$\forall s, t \in (0, \infty) : P[X > t + s \mid X > s] = P[X > t].$$

Diese Eigenschaft nennt man **Gedächtnislosigkeit**. Eine auf $\overline{\mathbb{R}}_+$ konzentrierte reell erweiterte Zufallsvariable ist genau dann gedächtnislos, wenn sie exponentialverteilt ist.

» “ \Leftarrow ”: Sei X also $\exp(\lambda)$ verteilt. Per definitionem der bedingten Wahrscheinlichkeit gilt,

$$P[X > t + s \mid X > s] = \frac{P[X > t + s, X > s]}{P[X > s]} = \frac{1 - F(t + s)}{1 - F(s)}.$$

Die Verteilungsfunktion erhält man durch elementare Integration,

$$F(x) = \begin{cases} \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Somit ergibt sich

$$\frac{1 - F(t + s)}{1 - F(s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F(t) = P[X > t].$$

“ \Rightarrow ”: Sei X gedächtnislos, d.h. $P[X > t + s \mid X > s] = P[X > s]$. Unter Verwendung der Definition der bedingten Wahrscheinlichkeit erhalten wir

$$\underbrace{P[X > t + s]}_{H(t+s)} = \underbrace{P[X > t]}_{H(t)} \underbrace{P[X > s]}_{H(s)}.$$

Wir interpretieren diese Gleichung als Funktionalgleichung und suchen nach einer Lösung $H: \mathbb{R}_+ \rightarrow \mathbb{R}_+$, die diese Gleichung erfüllt. Da

$$H(t) = P[X > t] = 1 - F(t)$$

erhalten wir aus den Eigenschaften der Verteilungsfunktion für H außerdem die Randbedingungen $H(0) = 1$ und $\lim_{t \rightarrow \infty} H(t) = 0$. Offensichtlich ist $H(t) = e^{-\lambda t}$ für $\lambda \in (0, \infty)$ eine Lösung.

Man kann zeigen, dass dies in der Klasse auf $\overline{\mathbb{R}_+}$ konzentrierten Verteilungen die einzige Lösung ist (siehe Hinderer § 28). «

Für die Beschreibung von technischen Geräten oder Bauteilen ist es oft notwendig ein “Gedächtnis” miteinzubauen. Dafür eignet sich die sog. Weibull-Verteilung¹, welche eine Verallgemeinerung der Exponential-Verteilung darstellt.

Für eine Zufallsvariable X mit $P_X = \exp(\lambda)$ gilt

$$\mathbf{E}X = \frac{1}{\lambda}, \quad \mathbf{V}(X) = \frac{1}{\lambda^2}.$$

» Der Nachweis folgt durch direktes Rechnen,

$$\begin{aligned} \mathbf{E}X &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} \lambda x e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_{x=0}^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda} \\ \mathbf{V}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 = \int_0^{\infty} \lambda x^2 e^{-\lambda x} dx - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \quad \llcorner \end{aligned}$$

$\exp(\lambda)$ hat die charakteristische Funktion φ mit $\varphi(u) = \frac{\lambda}{\lambda - iu}$.

» $\varphi(u) = \int_0^{\infty} e^{iux} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-(\lambda - iu)x} dx = \frac{\lambda}{\lambda - iu}$. « \dashv

¹Ernst Hjalmar Waloddi Weibull (* 18. Juni 1887; † 12. Oktober 1979 in Annecy) war ein schwedischer Ingenieur und Mathematiker.

6.6 *Bemerkung.* Als **Gamma-Verteilung** $\Gamma_{\lambda, \nu}$ mit Parametern $\lambda, \nu \in (0, \infty)$ wird ein W-Maß auf \mathcal{B} oder $\overline{\mathcal{B}}$ bezeichnet, das eine Dichtefunktion f mit

$$f(x) = \begin{cases} \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

besitzt. Hierbei ist $\Gamma_{\lambda, 1} = \exp(\lambda)$. Die Gamma-Verteilung stellt also ebenfalls eine Verallgemeinerung der Exponential-Verteilung dar. $\Gamma_{\frac{1}{2}, \frac{n}{2}}$ wird als **Chi-Quadrat-Verteilung** χ_n^2 mit n Freiheitsgraden bezeichnet ($n \in \mathbb{N}$).

Für eine Zufallsvariable X mit $P_X = \Gamma_{\lambda, \nu}$ gilt

$$\mathbf{E}X = \frac{\nu}{\lambda}, \quad \mathbf{V}(X) = \frac{\nu}{\lambda^2}.$$

$\Gamma_{\lambda, \nu}$ hat die charakteristische Funktion φ mit $\varphi(u) = \left(\frac{\lambda}{\lambda - iu}\right)^\nu$.

» $\mathbf{E}X$ und $\mathbf{V}X$ folgen durch direktes Integrieren oder unter Verwendung der charakteristischen Funktion.

Die charakteristische Funktion ist gegeben durch folgendes Integral

$$\varphi(u) = \int_0^\infty \frac{\lambda^\nu}{\Gamma(\nu)} e^{iux} x^{\nu-1} e^{-\lambda x} dx.$$

Dieses wollen wir nicht direkt berechnen, sondern in eine Differentialgleichung umformen und so elegant lösen. Unter Verwendung des Satzes über dominierte Konvergenz lässt sich nachweisen, dass Differentiation und Integration hier vertauscht werden können. Somit erhalten wir

$$\begin{aligned} \varphi'(u) &= \int_0^\infty \partial_u \frac{\lambda^\nu}{\Gamma(\nu)} e^{iux} x^{\nu-1} e^{-\lambda x} dx = \frac{\lambda^\nu}{\Gamma(\nu)} \int_0^\infty ix e^{iux} x^{\nu-1} e^{-\lambda x} dx \\ &= \frac{i\lambda^\nu}{\Gamma(\nu)} \int_0^\infty e^{(iu-\lambda)x} x^\nu dx \\ &= \underbrace{\frac{i\lambda^\nu}{\Gamma(\nu)} x^\nu \frac{e^{(iu-\lambda)x}}{iu-\lambda} \Big|_0^\infty}_{=0} - \frac{i\lambda^\nu}{\Gamma(\nu)} \int_0^\infty \nu x^{\nu-1} \frac{e^{(iu-\lambda)x}}{iu-\lambda} dx \\ &= \frac{i\nu}{\lambda - iu} \int_0^\infty \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{(iu-\lambda)x} dx = \frac{i\nu}{\lambda - iu} \varphi(u). \end{aligned}$$

Dies liefert nach Zerlegung der DGL in Real- und Imaginäranteil und anschließendem Lösen der Anfangswertprobleme $\operatorname{Re} \varphi(0) = 1$ und $\operatorname{Im} \varphi(0) = 0$ die eindeutige Lösung

$$\varphi(u) = \left(\frac{\lambda}{\lambda - iu} \right)^{\nu} . \quad \ll \quad \infty$$

■ Eindeutigkeitsatz und Umkehrformeln

Per se ist nicht klar, ob zwei unterschiedliche Verteilungsfunktionen auch unterschiedliche charakteristische Funktionen besitzen bzw. inwiefern eine charakteristische Funktion die Verteilung "charakterisiert". Der folgende Satz beantwortet die Frage.

6.4 **Eindeutigkeitsatz für charakteristische Funktionen** *Besitzen zwei auf \mathcal{B} definierte W -Maße bzw. Verteilungen dieselbe charakteristische Funktion, so stimmen sie überein.*

Ist μ ein W -Maß auf \mathcal{B} bzw. P_X die Verteilung einer reellen Zufallsvariablen X mit zugehöriger Verteilungsfunktion F und zugehöriger charakteristischer Funktion φ und sind a, b mit $a < b$ Stetigkeitsstellen von F , so gilt die Umkehrformel

$$\underbrace{F(b) - F(a)}_{=\mu((a,b]) \text{ bzw. } P_X((a,b])} = \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{-U}^U \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u) \, du . \quad \times$$

Wir erhalten mit Hilfe dieses Satzes zunächst nur die Eindeutigkeit auf den halboffenen Intervallen. Der Fortsetzungssatz besagt jedoch, dass damit auch die Fortsetzung auf die Menge der Borelschen Mengen eindeutig ist.

» Es genügt, die Umkehrformel zu beweisen. Wir verwenden dazu die Dirichlet-Formel

$$\lim_{\substack{A \rightarrow -\infty \\ B \rightarrow \infty}} \frac{1}{\pi} \int_A^B \frac{\sin(\nu)}{\nu} \, d\nu = 1,$$

die man leicht unter Verwendung der Funktionentheorie beweisen kann. Wir verwenden außerdem die Identität,

$$e^{-iua} - e^{-iub} = -e^{-iuy} \Big|_{y=a}^b = iu \int_{y=a}^b e^{-iuy} \, dy ,$$

somit folgt nach Anwendung des Satzes von Fubini

$$\begin{aligned}
 & \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{u=-U}^U \frac{1}{iu} (e^{-iua} - e^{-iub}) \varphi(u) du \\
 &= \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{u=-U}^U \int_{x=-\infty}^{\infty} \frac{1}{iu} (e^{-iua} - e^{-iub}) e^{iux} F(dx) du \\
 &= \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{x=-\infty}^{\infty} \int_{u=-U}^U \int_{y=a}^b e^{iu(x-y)} dy F(dx) du \\
 &= \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{x=-\infty}^{\infty} \int_{y=a}^b \int_{u=-U}^U e^{iu(x-y)} du dy F(dx) \\
 &= \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{x=-\infty}^{\infty} \int_{y=a}^b \frac{e^{iU(x-y)} - e^{-iU(x-y)}}{i(x-y)} dy F(dx) \\
 &= \lim_{U \rightarrow \infty} \frac{1}{\pi} \int_{x=-\infty}^{\infty} \int_{y=a}^b \frac{\sin(U(x-y))}{(x-y)} dy F(dx)
 \end{aligned}$$

Nun substituieren wir $v = U(x - y)$, d.h. " $dx = \frac{1}{U} dv$ ",

$$\begin{aligned}
 &= \int_{x=-\infty}^{\infty} \lim_{U \rightarrow \infty} \frac{1}{\pi} \underbrace{\int_{y=U(x-a)}^{U(x-b)} \frac{\sin(v)}{v} dv}_{=: G(U, x)} F(dx) \\
 &= \int_{-\infty}^{\infty} \lim_{U \rightarrow \infty} G(U, x) F(dx)
 \end{aligned}$$

Für den Integranden erhalten wir unter Verwendung der Formel von Dirichlet,

$$\lim_{U \rightarrow \infty} G(U, x) = \begin{cases} 1, & \text{falls } a < x < b, \\ 0, & \text{falls } x < a, \\ 0, & \text{falls } x > b. \end{cases}$$

Falls wir den Satz von der dominierten Konvergenz anwenden können, erhalten wir für das Integral

$$\int_{-\infty}^{\infty} \lim_{U \rightarrow \infty} G(U, x) F(dx) = F(b) - F(a),$$

denn das Verhalten des Integranden an den Unstetigkeitsstellen ist für den Integralwert nicht von Bedeutung.

Wir müssen also noch nachweisen, dass wir den Satz über die dominierte Konvergenz auch anwenden können. Setze $n = U$ und $f_n = G(U, \cdot)$, $G(U, \cdot)$ ist beschränkt, d.h.

$$|G(U, \cdot)| \leq g < \infty, \quad g \in \mathbb{R}.$$

Nun ist die konstante Funktion g integrierbar bezüglich der durch die Verteilungsfunktion F induzierten Verteilung μ . ($\mu(a, b] = F(b) - F(a)$). «

Mit dem Eindeutigkeitsatz erhält man aus φ die Verteilungsfunktion, ist φ darüber hinaus integrierbar, gilt wesentlich mehr.

6.5 **Satz** Gegeben sei eine Verteilungsfunktion F mit charakteristischer Funktion φ ; es sei $\int_{-\infty}^{\infty} |\varphi(u)| du < \infty$. Dann besitzt F eine Dichte f , die stetig und beschränkt ist, und es gilt die Umkehrformel

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \varphi(u) du, \quad x \in \mathbb{R}. \quad \times$$

» Sei also φ integrierbar, d.h. $\int_{\mathbb{R}} |\varphi|(x) dx = c < \infty$. Wir verwenden die Umkehrformel

$$F(b) - F(a) = \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{-U}^U \frac{e^{-iua} - e^{-iub}}{iu} \varphi(u) du$$

und bilden so den Differenzenquotienten für F , falls F in x stetig,

$$\frac{F(x+h) - F(x)}{h} = \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{-U}^U \left(\frac{e^{-iu(x+h)} - e^{-iux}}{iuh} \right) \varphi(u) du.$$

Diesen Ausdruck können wir durch $\frac{1}{2\pi} \int_{\mathbb{R}} |\varphi|(u) du$ majorisieren, denn

$$\left| \frac{e^{-iu(x+h)} - e^{-iux}}{iuh} \right| = \frac{1}{|uh|} \left| \int_x^{x+h} -iue^{-ius} ds \right| \leq \frac{1}{|h|} \left| \int_x^{x+h} ds \right| = 1.$$

Somit können wir den Satz von der dominierten Konvergenz anwenden und die Grenzwertbildung für $h \rightarrow 0$ mit der Grenzwertbildung für $U \rightarrow \infty$ und der Integration vertauschen.

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} &= \lim_{h \rightarrow 0} \lim_{U \rightarrow \infty} \frac{1}{2\pi} \int_{-U}^U - \left(\frac{e^{-iu(x+h)} - e^{-iux}}{iuh} \right) \varphi(u) \, du \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \varphi(u) \, du =: f(x). \end{aligned}$$

f ist offensichtlich Dichte zu F und beschränkt, denn

$$|f(x)| \leq \frac{1}{2\pi} \int_{\mathbb{R}} |e^{-iux}| |\varphi(u)| \, du = \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi(u)| \, du < \infty,$$

und außerdem stetig, denn

$$\begin{aligned} |f(y) - f(x)| &\leq \int_{\mathbb{R}} |e^{iuy} - e^{iux}| |\varphi(u)| \, du \\ &\leq \sup_{x,y \in U} |e^{iuy} - e^{iux}| \int_{\mathbb{R}} |\varphi(u)| \, du \\ &= c \sup_{x,y \in U} |e^{iuy} - e^{iux}|. \end{aligned}$$

Der rechte Ausdruck wird klein für $|x - y|$ klein, denn $e^{iu \cdot}$ ist stetig. \llcorner

Ist X Zufallsvariable mit integrierbarer charakteristischer Funktion φ , so besitzt X eine Dichte und diese entspricht im Wesentlichen der Fouriertransformierten von φ . Wie wir bereits festgestellt haben, erhalten wir φ als inverse Fouriertransformierte dieser Dichte zurück.

6.6 Satz Sei X eine reelle Zufallsvariable mit Verteilung P_X und charakteristischer Funktion φ . Falls für $j \in \mathbb{N}$ gilt:

$$\mathbf{E}|X|^j < \infty,$$

so besitzt φ eine stetige j -te Ableitung $\varphi^{(j)}$ mit

$$\varphi^{(j)}(u) = i^j \int_{\mathbb{R}} x^j e^{iux} P_X(dx), \quad u \in \mathbb{R}.$$

Insbesondere $\varphi^{(j)}(0) = i^j \mathbf{E}X^j$. \times

» Wir beweisen die Behauptung für $j = 1$, der Rest folgt induktiv.

$$\begin{aligned}\varphi(u) &= \int_{-\infty}^{\infty} e^{iux} P_X(\mathrm{d}x) \\ \varphi'(u) &= \lim_{h \rightarrow 0} \frac{\varphi(u+h) - \varphi(u)}{h} = \lim_{h \rightarrow 0} \int_{\mathbb{R}} \frac{e^{i(u+h)x} - e^{iux}}{h} P_X(\mathrm{d}x) \\ &= \lim_{h \rightarrow 0} \int_{\mathbb{R}} e^{iux} \underbrace{\left(\frac{e^{ihx} - 1}{h} \right)}_{\rightarrow ix e^{iux}} P_X(\mathrm{d}x)\end{aligned}$$

Wir müssen also nachweisen, dass wir den Satz von der dominierten Konvergenz anwenden können,

$$\left| e^{iux} \underbrace{\left(\frac{e^{ihx} - 1}{h} \right)}_{\rightarrow ix e^{iux}} \right| = 1 \left| \frac{e^{ihx} - 1}{h} \right| < \text{const } |x|, \quad x, h \in \mathbb{R}$$

und $\int_{\mathbb{R}} |x| P_X(\mathrm{d}x) = \mathbf{E}|X| < \infty$.

Somit erhalten wir für das Integral,

$$\varphi'(u) = i \int_{\mathbb{R}} x e^{iux} P_X(\mathrm{d}x)$$

φ' ist gleichmäßig stetig, denn

$$|\varphi'(u+h) - \varphi'(u)| \leq \int_{\mathbb{R}} |x| \left| e^{i(u+h)x} - e^{iux} \right| P_X(\mathrm{d}x)$$

Nun ist $\left| e^{i(u+h)x} - e^{iux} \right|$ nach der Dreiecksungleichung durch 2 beschränkt und konvergiert für $h \rightarrow 0$ gegen Null, somit haben wir eine integrierbare Majorante und mit dem Satz von der dominierten Konvergenz folgt,

$$|\varphi'(u+h) - \varphi'(u)| \rightarrow 0, \quad h \rightarrow 0. \quad \llcorner$$

Erinnern wir uns an die erzeugende Funktion g einer auf \mathbb{N}_0 konzentrierten Verteilung so besagt Satz 6.2,

$$g^{(j)}(1-) = \mathbf{E}(X(X-1) \cdots (X-j+1)).$$

Für charakteristische Funktionen reeller Zufallsvariablen gilt nun,

$$\varphi^{(j)}(0) = i^j \mathbf{E}X^j.$$

Erzeugende und charakteristische Funktionen ermöglichen es uns Integrationsaufgaben,

$$\int_{\mathbb{R}} x^j P_X, \quad \sum_{k \geq 0} k^j b_k$$

auf Differentiationsaufgaben $\varphi^j(u)|_{u=0}$, $g^{(j)}(1-)$ zurückzuführen, die sich meist wesentlich leichter lösen lassen.

6.7 **Satz** Seien X_1, \dots, X_n unabhängige reelle Zufallsvariablen mit charakteristischen Funktionen $\varphi_1, \dots, \varphi_n$. Für die charakteristische Funktion der Summe

$$X_1 + \dots + X_n$$

gilt dann

$$\varphi = \prod_{j=1}^n \varphi_j. \quad \times$$

$$\gg \varphi(u) = \mathbf{E}e^{iu(X_1 + \dots + X_n)} = \mathbf{E}(e^{iuX_1} \dots e^{iuX_n}) \stackrel{5.4}{=} (\mathbf{E}e^{iuX_1}) \dots (\mathbf{E}e^{iuX_n}). \quad \ll$$

Für die erzeugende und charakteristische Funktion einer Summe von unabhängigen Zufallsvariablen gelten also analoge Aussagen. Beide beruhen auf der Multiplikatивität des Erwartungswerts für unabhängige Zufallsvariablen.

6-C Faltungen

In Abschnitt 5-A haben wir bereits für zwei reelle unabhängige Zufallsvariablen X und Y die Faltung $P_X * P_Y := P_{X+Y}$ als Verteilung von $X + Y$ definiert. Wir wollen die Definition nun noch etwas Verallgemeinern und die Argumentationslücken aus 5-A schließen.

6.3 **Definition** Zu zwei W -Maßen P, Q auf \mathcal{B} wird das W -Maß $P * Q$ auf \mathcal{B} , die sog. *Faltung* von P und Q , folgendermaßen definiert:

1.) Sei $T : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ mit $T(x, y) = x + y$;

$$P * Q := (P \otimes Q)_T$$

($P \otimes Q$ W-Maß auf \mathcal{B}_2 mit $(P \otimes Q)(B_1 \times B_2) = P(B_1)Q(B_2)$, $B_{1,2} \in \mathcal{B}$, sog. Produkt-W-Maß von P und Q)

oder — äquivalent —

2.) Man wähle ein unabhängiges Paar reeller Zufallsvariablen X, Y mit Verteilungen $P_X = P$, $P_Y = Q$ und setze

$$P * Q := P_{X+Y}. \quad \times$$

» Beweis der Äquivalenz in 6.3. Sei $Z = (X, Y)$ gemäß b), dann gilt

$$P_{X+Y} = P_{T \circ Z} = (P_Z)_T \stackrel{!}{=} (P_X \otimes P_Y)_T.$$

Um (!) nachzuweisen, genügt es zu zeigen, dass

$$\forall B_1, B_2 \in \mathcal{B} : \underbrace{(P_Z)(B_1 \times B_2)}_{P[Z \in B_1 \times B_2]} = \underbrace{(P_X \otimes P_Y)(B_1 \times B_2)}_{P_X(B_1)P_Y(B_2)}. \quad (*)$$

Da X, Y unabhängig sind, gilt

$$\begin{aligned} P[Z \in B_1 \times B_2] &= P[(X, Y) \in B_1 \times B_2] = P[X \in B_1, Y \in B_2] \\ &= P[X \in B_1]P[Y \in B_2] \end{aligned}$$

und somit ist die Identität (*) gezeigt. Mit Hilfe des Fortsetzungssatzes liegt die Identität auf allen Borelschen Mengen aus \mathcal{B}_2 vor. «

6.8 **Satz** a) Die Faltungsoption ist kommutativ und assoziativ.

b) Sei (X, Y) ein unabhängiges Paar reeller Zufallsvariablen. Für X, Y und $X + Y$ seien die Verteilungsfunktionen mit F, G, H , die Dichten — falls existent — mit f, g, h und die Zähldichten — falls P_X, P_Y auf \mathbb{N}_0 konzentriert sind — mit $(p_k), (q_k), (r_k)$ bezeichnet. Dann gilt

$$H(t) = \int_{\mathbb{R}} F(t - y) G(dy) = \int_{\mathbb{R}} G(t - x) F(dx), \quad t \in \mathbb{R}.$$

Falls f existiert, so existiert h mit

$$h(t) = \int_{\mathbb{R}} f(t-y) G(dy) \quad \text{für (L-f.a.) } t \in \mathbb{R};$$

falls zusätzlich g existiert, so gilt

$$h(t) = \int_{\mathbb{R}} f(t-y)g(y) dy = \int_{\mathbb{R}} g(t-x)f(x) dx \quad \text{für (L-f.a.) } t \in \mathbb{R}.$$

Falls $(p_k)_{k \in \mathbb{N}_0}$, $(q_k)_{k \in \mathbb{N}_0}$ existieren, so existiert $(r_k)_{k \in \mathbb{N}_0}$ mit

$$r_k = \sum_{i=0}^k p_{k-i}q_i = \sum_{i=0}^k q_{k-i}p_i \quad (k \in \mathbb{N}_0). \quad \times$$

» a) Assoziativität und Kommutativität der Faltungsoperation folgen direkt aus der Assoziativität und Kommutativität der Addition reeller Zahlen.

b) Siehe auch Satz 5.6. Sei

$$H(t) = P[X + Y \leq t] = P[(X, Y) \in B]$$

mit $B := \{(x, y) : x + y \leq t\}$, also

$$\begin{aligned} H(t) &= P_{(X,Y)}(B) = (P_X \otimes P_Y)(B) = \int_{\mathbb{R}} P_X((-\infty, t-y])P_Y(dy) \\ &= \int_{\mathbb{R}} F(t-y)G(dy) \stackrel{\text{Komm. von } X+Y}{=} \int_{\mathbb{R}} G(t-x)F(dx). \end{aligned}$$

Falls zu F eine Dichte f existiert, dann gilt für $t \in \mathbb{R}$,

$$H(t) = \int_{\mathbb{R}} \left[\int_{-\infty}^{t-y} f(\tau) d\tau \right] G(dy) \stackrel{\text{Fubini}}{=} \int_{-\infty}^t \left[\int_{\mathbb{R}} f(\tau-y)G(dy) \right] d\tau.$$

Falls zusätzlich G eine Dichte g besitzt, dann ist $G(dy) = g(y) dy$ und somit besitzt H eine Dichte der Form

$$h(t) = \int_{\mathbb{R}} f(t-y)g(y) dy = \int_{\mathbb{R}} g(t-x)f(x) dx,$$

wobei $h(t)$ nur λ -f.ü. definiert ist. Falls X und Y Zähldichten (p_k) und (q_k) besitzen, dann ist

$$\begin{aligned} r_k &:= P[X + Y = k] = P[\exists i \in \{0, 1, \dots, k\} : X = k - i, Y = i] \\ &= \sum_{i=0}^k P[X = k - i, Y = i] = \sum_{i=0}^k P[X = k - i]P[Y = i] \\ &= \sum_{i=0}^k p_{k-i}q_i = \sum_{i=0}^k q_{k-i}p_i. \quad \ll \end{aligned}$$

Sind X und Y reelle unabhängige Zufallsvariablen mit Dichten f und g , so ist mit dem eben bewiesenen die Dichte von $X + Y$ gegeben durch

$$\int_{\mathbb{R}} f(x - y)g(y) dy = \int_{\mathbb{R}} g(y - x)f(x) dx.$$

Dieser Ausdruck wird als **Faltung von f mit g** bezeichnet, geschrieben

$$f * g =: f_{X+Y}.$$

6.7 *Bemerkungen.* A. Für $n_{1,2} \in \mathbb{N}$, $p \in [0, 1]$ gilt

$$b(n_1, p) * b(n_2, p) = b(n_1 + n_2, p).$$

Die Summe von n unabhängigen $b(1, p)$ verteilten Zufallsvariablen ist also $b(n, p)$ -verteilt.

B. Für $\lambda_{1,2} > 0$ gilt $\pi(\lambda_1) * \pi(\lambda_2) = \pi(\lambda_1 + \lambda_2)$.

» $\pi(\lambda_1), \pi(\lambda_2)$ haben nach Bemerkung 6.2 die erzeugenden Funktionen

$$g_1(s) = e^{\lambda_1(s-1)}, \quad g_2(s) = e^{\lambda_2(s-1)}.$$

Im Satz 6.3 haben wir gezeigt, dass die Verteilungsfunktion einer Zufallsvariable $X + Y$, wobei X und Y unabhängig und $X \pi(\lambda_1)$ und $Y \pi(\lambda_2)$ verteilt ist, die Form hat

$$\pi(\lambda_1) * \pi(\lambda_2).$$

Diese hat die erzeugende Funktion

$$g_1(s)g_2(s) = e^{(\lambda_1 + \lambda_2)(s-1)},$$

welche ebenfalls die erzeugende Funktion von $\pi(\lambda_1 + \lambda_2)$ ist. Mit Satz 6.1 folgt nun, dass $\pi(\lambda_1 + \lambda_2) = \pi(\lambda_1) * \pi(\lambda_2)$. «

C. Für $r_{1,2} \in \mathbb{N}$, $p \in (0, 1)$ gilt

$$Nb(r_1, p) * Nb(r_2, p) = Nb(r_1 + r_2, p).$$

Die Summe von r unabhängigen $Nb(1, p)$ -verteilten Zufallsvariablen ist also $Nb(r, p)$ -verteilt.

D. Für $a_{1,2} \in \mathbb{R}$, $\sigma_{1,2}^2 \in (0, \infty)$ gilt

$$N(a_1, \sigma_1^2) * N(a_2, \sigma_2^2) = N(a_1 + a_2, \sigma_1^2 + \sigma_2^2).$$

E. Für $\lambda \in (0, \infty)$, $\nu_{1,2} \in (0, \infty)$ gilt

$$\Gamma_{\lambda, \nu_1} * \Gamma_{\lambda, \nu_2} = \Gamma_{\lambda, \nu_1 + \nu_2}.$$

Die Summe von n unabhängigen $\exp(\lambda)$ -verteilten Zufallsvariablen ist $\Gamma_{\lambda, n}$ -verteilt.

Die Summe der Quadrate von n unabhängigen jeweils $N(0, 1)$ -verteilten Zufallsvariablen ist χ_n^2 -verteilt. \rightarrow

Es lässt sich auch für nicht unabhängige reelle Zufallsvariablen X und Y nach der Verteilung von $X + Y$ fragen. Um diese zu berechnen kann man stets folgenden Ansatz wählen,

$$\begin{aligned} F_{X+Y}(t) &= P[X + Y \leq t] = P[g \circ (X, Y) \leq t] = P[(X, Y) \in g^{-1}(-\infty, t]] \\ &= P_{(X, Y)}[g^{-1}(-\infty, t]] = \int_{g^{-1}(-\infty, t]} 1 \, dP_{(X, Y)}. \end{aligned}$$

Besitzen X und Y eine gemeinsame Dichte, so kann man das eigentliche Integrationsgebiet $g^{-1}(-\infty, t] \cap \{f_{(X, Y)} \neq 0\}$ (meißt geometrisch) bestimmen und so das Integral berechnen.

7 Spezielle Verteilungen

Wir wollen in diesem Kapitel die wesentlichen Eigenschaften der bisher vorgestellten Verteilungen zusammenfassen.

7-A Diskrete Verteilungen

Verteilungen, die auf höchstens abzählbare Mengen konzentriert sind, heißen **diskret**. Nimmt eine Zufallsvariable nur endlich oder abzählbar viele Werte an, so ist ihre Verteilung P_X diskret.

- 7.1 **Definition** Als *Binomialverteilung* $b(n, p)$ mit Parametern $n \in \mathbb{N}$, $p \in (0, 1)$ oder auch $[0, 1]$ wird ein W -Maß auf \mathcal{B} bezeichnet, das auf $\{0, 1, \dots, n\}$ konzentriert ist und die Zähldichte

$$k \mapsto b(n, p; k) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \mathbb{N}_0$$

besitzt. \times

- 7.1 **Satz** Sei $n \in \mathbb{N}$, $p \in [0, 1]$; $q := 1 - p$.

a) Ist (X_1, \dots, X_n) ein n -Tupel unabhängiger reeller Zufallsvariablen mit

$$P[X_i = 1] = p, \quad P[X_i = 0] = q, \quad i = 1, \dots, n$$

d.h. mit jeweiliger $b(1, p)$ -Verteilung, so ist $\sum_{i=1}^n X_i$ $b(n, p)$ -verteilt.

b) Für eine Zufallsvariable X mit $P_X = b(n, p)$ gilt

$$\mathbf{E}X = np, \quad \mathbf{V}(X) = npq.$$

c) $b(n, p)$ hat die erzeugende Funktion g mit $g(s) = (q + ps)^n$.

d) Für $n_{1,2} \in \mathbb{N}$ gilt

$$b(n_1, p) * b(n_2, p) = b(n_1 + n_2, p),$$

d.h. für zwei unabhängige Zufallsvariablen X_1, X_2 mit $P_{X_1} = b(n_1, p)$ und $P_{X_2} = b(n_2, p)$ gilt $P_{X_1+X_2} = b(n_1 + n_2, p)$. \times

7.1 **Bemerkung.** Das n -Tupel (X_1, \dots, X_n) von Zufallsvariablen aus Satz 7.1 beschreibt ein n -Tupel von Bernoulli-Versuchen, d.h. Zufallsexperimenten ohne gegenseitige Beeinflussung mit Ausgängen 1 ("Erfolg") oder 0 ("Misserfolg") und jeweiliger Erfolgswahrscheinlichkeit p ; $\sum_{k=1}^n X_i$ gibt die Anzahl der Erfolge in n Bernoulli-Versuchen an. \rightarrow

7.2 **Definition** Als **Poisson-Verteilung** $\pi(\lambda)$ mit Parameter $\lambda > 0$ wird ein W -Maß auf \mathcal{B} bezeichnet, das auf \mathbb{N}_0 konzentriert ist und die Zähldichte

$$k \mapsto \pi(\lambda; k) := e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

besitzt. \times

7.2 **Satz** Sei $\lambda > 0$.

a) Ist $(b(n, p_n))_n$ eine Folge von Binomialverteilungen mit $n p_n \rightarrow \lambda$ für $n \rightarrow \infty$, dann konvergiert

$$b(n, p_n; k) \rightarrow \pi(\lambda; k) \quad (n \rightarrow \infty) \text{ für alle } k \in \mathbb{N}_0.$$

b) Für eine Zufallsvariable X mit $P_X = \pi(\lambda)$ gilt

$$\mathbf{E}X = \mathbf{V}(X) = \lambda.$$

c) $\pi(\lambda)$ hat die erzeugende Funktion g mit $g(s) = e^{-\lambda + \lambda s}$.

d) Für $\lambda_{1,2} > 0$ gilt

$$\pi(\lambda_1) * \pi(\lambda_2) = \pi(\lambda_1 + \lambda_2). \quad \times$$

7.2 **Bemerkung.** Bei einer rein zufälligen Aufteilung von n unterscheidbaren Teilchen auf m gleichgroße mehrfach besetzbare Zellen in einem Euklidischen Raum wird die Anzahl der Teilchen in einer vorgegebenen Zelle durch eine $b(n, \frac{1}{m})$ -verteilte Zufallsvariable angegeben.

Der Grenzübergang $n \rightarrow \infty, m \rightarrow \infty$ mit Belegungsintensität $\frac{n}{m} \rightarrow \lambda > 0$ führt gemäß Satz 7.2a zu einer $\pi(\lambda)$ -verteilten Zufallsvariable. \rightarrow

7.3 **Definition** Als *negative Binomialverteilung* oder *Pascal-Verteilung* $Nb(r, p)$ mit Parametern $r \in \mathbb{N}$, $p \in (0, 1)$ wird ein W -Maß auf \mathcal{B} (oder auch $\overline{\mathcal{B}}$) bezeichnet, das auf \mathbb{N}_0 konzentriert ist und — mit $q := 1 - p$ — die Zähldichte

$$k \mapsto Nb(r, p; k) := \binom{r+k-1}{k} p^r q^k, \quad k \in \mathbb{N}$$

besitzt. Speziell wird $Nb(1, p)$ — mit Zähldichte $k \mapsto p(1-p)^k$, $k \in \mathbb{N}_0$ — als *geometrische Verteilung* mit Parameter $p \in (0, 1)$ bezeichnet. \times

7.3 **Satz** Sei $r \in \mathbb{N}$, $p \in (0, 1)$, $q := 1 - p$.

a) Sei $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Folge von $b(1, p)$ -verteilten Zufallsvariablen. Die erweitert-reelle Zufallsvariable

$$X := \inf \left\{ n \in \mathbb{N} : \sum_{k=1}^n X_k = r \right\} - r$$

mit $\inf \emptyset := \infty$ ist $Nb(r, p)$ -verteilt.

b) Für eine Zufallsvariable X mit $P_X = Nb(r, p)$ gilt

$$EX = \frac{rq}{p}, \quad V(X) = \frac{rq}{p^2}.$$

c) $Nb(r, p)$ hat die erzeugende Funktion g mit $g(s) = p^r (1 - qs)^{-r}$.

d) Für $r_{1,2} \in \mathbb{N}$ gilt

$$Nb(r_1, p) * Nb(r_2, p) = Nb(r_1 + r_2, p).$$

Die Summe von r unabhängigen $Nb(1, p)$ -verteilten Zufallsvariablen ist somit $Nb(r, p)$ -verteilt. \times

7.3 **Bemerkung.** Für die Folge (X_n) in Satz 7.3a wird durch die erweitert-reelle Zufallsvariable

$$T := \inf \left\{ n \in \mathbb{N} : \sum_{k=1}^n X_k = r \right\}$$

mit $\inf \emptyset := \infty$ die *Wartezeit* bis zum r -ten Auftreten der Eins in (X_n) und durch die erweitert-reelle Zufallsvariable $X = T - r$ die Anzahl der Misserfolge bis zum r -ten Erfolg bei der zu (X_n) gehörigen Folge von Bernoulli-Versuchen angegeben. \rightarrow

- 7.4 **Definition** Als *hypergeometrische Verteilung* mit Parametern $n, r, s \in \mathbb{N}$, wobei $n \leq r+s$, wird ein W -Maß auf \mathcal{B} bezeichnet, das auf $\{0, 1, \dots, n\}$ — sogar auf $\{\max(0, n-s), \dots, \min(n, r)\}$ — konzentriert ist und die Zähldichte

$$k \mapsto \begin{cases} \frac{\binom{r}{k} \binom{s}{n-k}}{\binom{r+s}{n}}, & k = 0, 1, \dots, n \\ 0, & k = n+1, n+2, \dots \end{cases}$$

besitzt. \times

- 7.4 **Bemerkung.** Werden aus einer Urne mit r roten und s schwarzen Kugeln rein zufällig n Kugeln ohne Zurücklegen gezogen (n, r, s wie in Definition 7.4), so wird die Anzahl der gezogenen roten Kugeln durch eine Zufallsvariable angegeben, die eine hypergeometrische Verteilung mit Parametern n, r, s besitzt. Anwendung der hypergeometrischen Verteilung in der Qualitätskontrolle. \rightarrow

7-B Totalstetige Verteilungen

Totalstetige Verteilungen sind die Verteilungen, die eine Dichtefunktion

$$f : \mathbb{R} \rightarrow \mathbb{R}_+$$

besitzen. Sie sind dann natürlich auch stetig, denn für ihre Verteilungsfunktion gilt

$$F(t) = \int_{-\infty}^t f(x) dx,$$

aber nicht jede stetige Verteilung besitzt eine Dichtefunktion. Mit Hilfe der Dichtefunktion können wir Randverteilungen, Erwartungswert, Varianz, Momente ... sehr leicht berechnen. Totalstetige Verteilungen sind also sehr angenehm.

- 7.5 **Definition** Als *Gleichverteilung auf (a, b)* mit $-\infty < a < b < \infty$ wird ein W -Maß auf \mathcal{B} bezeichnet, das eine Dichte

$$f = \frac{1}{b-a} \mathbf{1}_{(a,b)}$$

besitzt. \times

- 7.4 **Satz** Für eine Zufallsvariable X mit Gleichverteilung auf (a, b) gilt

$$EX = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}. \quad \times$$

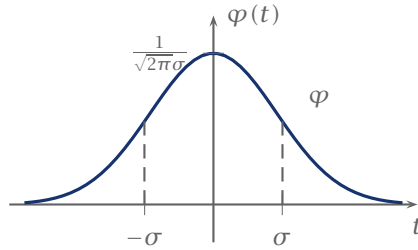
- 7.6 **Definition** Als (eindimensionale) *Normalverteilung* oder *Gauß-Verteilung* $N(a, \sigma^2)$ mit Parametern $a \in \mathbb{R}$, $\sigma > 0$ wird ein W -Maß auf \mathcal{B} bezeichnet, das eine Dichte f mit

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$

besitzt. Speziell heißt $N(0, 1)$ *standardisierte Normalverteilung*. \times

- 7.5 **Bemerkungen.** A. Sei f wie in Definition 7.6. Der Graph von f heißt Gaußsche Glockenkurve. Im Fall $a = 0$ ist $f(0) = \frac{1}{\sqrt{2\pi\sigma}}$ und hat f zwei Wendepunkte $(\pm\sigma; \frac{1}{\sqrt{2\pi e\sigma}})$.

B. Dichte- und die Verteilungsfunktion ϕ von $N(0, 1)$ sind tabelliert.



7.1 Gaußsche Glockenkurve mit Wendepunkten

c. Ist die Zufallsvariable X $N(a, \sigma^2)$ -verteilt, so ist $\frac{X-a}{\sigma}$ $N(0, 1)$ -verteilt. Es gilt hierbei

$$P[a - \sigma \leq X \leq a + \sigma] = 2\phi(1) - 1 \approx 0,683$$

$$P[a - 2\sigma \leq X \leq a + 2\sigma] = 2\phi(2) - 1 \approx 0,954$$

$$P[a - 3\sigma \leq X \leq a + 3\sigma] = 2\phi(3) - 1 \approx 0,997.$$

d. Anwendung der Normalverteilung z.B. in der Theorie der Beobachtungsfehler.

→

7.5 **Satz** Sei $a \in \mathbb{R}$, $\sigma > 0$.

a) Für eine Zufallsvariable X mit Verteilung $N(a, \sigma^2)$ gilt:

$$\mathbf{E}(X - a)^{2k-1} = 0, \quad \mathbf{E}(X - a)^{2k} = \sigma^{2k} \prod_{j=1}^k (2j - 1), \quad k \in \mathbb{N};$$

insbesondere $\mathbf{E}X = a$, $\mathbf{V}(X) = \sigma^2$. Die Zufallsvariable $Y = cX + b$ mit $0 \neq c \in \mathbb{R}$, $b \in \mathbb{R}$ hat die Verteilung $N(ca + b, c^2\sigma^2)$.

b) Die charakteristische Funktion von $N(a, \sigma^2)$ ist φ mit

$$\varphi(u) = e^{iau} e^{-\frac{\sigma^2 u^2}{2}}.$$

c) Für $a_{1,2} \in \mathbb{R}$, $\sigma_{1,2} > 0$ gilt

$$N(a_1, \sigma_1^2) * N(a_2, \sigma_2^2) = N(a_1 + a_2, \sigma_1^2 + \sigma_2^2). \quad \times$$

7.7 **Definition** Als *Exponentialverteilung* $\exp(\lambda)$ mit Parameter $\lambda > 0$ wird ein W -Maß auf \mathcal{B} oder $\overline{\mathcal{B}}$ bezeichnet, das eine Dichtefunktion f mit

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

besitzt. \times

7.6 **Satz** Sei $\lambda > 0$.

a) Sei X eine erweitert-reelle Zufallsvariable, deren Verteilung auf $\overline{\mathbb{R}}_+$ konzentriert sei, mit $P[0 < X < \infty] > 0$. X erfüllt

$$\forall s, t \in (0, \infty) : P[X > t + s \mid X > s] = P[X > t].$$

genau dann, wenn P_X eine Exponentialverteilung ist. Diese Eigenschaft heißt “Gedächtnislosigkeit”.

b) Für eine Zufallsvariable X mit $P_X = \exp(\lambda)$ gilt

$$\mathbf{E}X = \frac{1}{\lambda}, \quad \mathbf{V}(X) = \frac{1}{\lambda^2}.$$

c) $\exp(\lambda)$ hat die charakteristische Funktion φ mit

$$\varphi(u) = \frac{\lambda}{\lambda - iu}. \quad \times$$

7.6 **Bemerkung.** Die zufällige Lebensdauer eines radioaktiven Atoms wird durch eine exponentialverteilte Zufallsvariable angegeben. \rightarrow

7.8 **Definition** Als *Gamma-Verteilung* $\Gamma_{\lambda, \nu}$ mit Parametern $\lambda, \nu > 0$ wird ein W -Maß auf \mathcal{B} oder $\overline{\mathcal{B}}$ bezeichnet, das eine Dichtefunktion f mit

$$f(x) = \begin{cases} \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

besitzt. Hierbei gilt $\Gamma_{\lambda, 1} = \exp(\lambda)$. $\Gamma_{\frac{1}{2}, \frac{n}{2}}$ wird als Chi-Quadrat-Verteilung χ_n^2 mit $n \in \mathbb{N}$) Freiheitsgraden bezeichnet. \times

7.7 **Satz** Seien $\lambda, \nu > 0$, $n \in \mathbb{N}$.

a) Für eine Zufallsvariable X mit $P_X = \Gamma_{\lambda, \nu}$ gilt $\mathbf{E}X = \frac{\nu}{\lambda}, \mathbf{V}(X) = \frac{\nu}{\lambda^2}$.

b) $\Gamma_{\lambda, \nu}$ hat die charakteristische Funktion φ mit

$$\varphi(u) = \left(\frac{\lambda}{\lambda - iu} \right)^\nu.$$

c) Für $\nu_{1,2} > 0$ gilt

$$\Gamma_{\lambda, \nu_1} * \Gamma_{\lambda, \nu_2} = \Gamma_{\lambda, \nu_1 + \nu_2}.$$

Insbesondere ist $\Gamma_{\lambda, n}$ die n -fache Faltung von $\exp(\lambda)$.

d) Die Summe der Quadrate von n unabhängigen jeweils $N(0, 1)$ -verteilten reellen Zufallsvariablen ist χ_n^2 -verteilt. \times

7.7 **Bemerkung.** Sind X, Y zwei unabhängige reelle Zufallsvariablen mit $P_X = N(0, 1)$ und $P_Y = \chi_n^2$, so wird die Verteilung der Zufallsvariablen $\frac{X}{\sqrt{Y/n}}$ als **t -Verteilung** oder **Student-Verteilung** t_n bzw. St_n mit n Freiheitsgraden bezeichnet ($n \in \mathbb{N}$). t_1 bzw. St_1 ist eine sogenannte **Cauchy-Verteilung** — Anwendung von χ_n^2 und St_n in der Statistik. \rightarrow

Sei $\{T_n : n \in \mathbb{N}\}$ eine unabhängige Folge nichtnegativ-reeller Zufallsvariablen (d.h. T_i und T_j haben jeweils dieselbe Verteilung) mit $P\{T_n = 0\} < 1$.

Definiere die Zufallsvariable $N_t := \sup \{n \in \mathbb{N} : T_1 + \dots + T_n \leq t\}$ für $t \in \mathbb{R}_+$ und der Konvention $\sup \emptyset = 0$.

BSP 1 In einem technischen System wird ein Bauelement mit endlicher zufälliger Lebensdauer bei Ausfall durch ein gleichartiges Element ersetzt. Die einzelnen Lebensdauern seien Realisierung der Zufallsvariablen T_n . N_t gibt die Anzahl der Erneuerungen im Zeitintervall $[0, t]$ an. ■

7.8 **Satz** Ist — mit den obigen Bezeichnungen — T_n $\exp(\lambda)$ -verteilt ($\lambda > 0$), so ist N_t $\pi(\lambda t)$ -verteilt, $t \in \mathbb{R}_+$. \times

» Setze $S_k := T_1 + \dots + T_k$ mit $k \in \mathbb{N}$, wobei $S_0 = 0$. Gesucht ist nun

$$P\{N_t = k\} = P\{S_k \leq t, S_{k+1} > t\} = P\{S_k \leq t\} - P\{S_{k+1} \leq t\},$$

denn $[S_{k+1} > t] = [S_{k+1} \leq t]^c$ und somit ist

$$[S_k \leq t] \cap [S_{k+1} \leq t]^c = [S_k \leq t] \setminus [S_{k+1} \leq t].$$

Nun ist $S_k \Gamma_{\lambda,k}$ verteilt und $S_{k+1} \Gamma_{\lambda,k+1}$ verteilt. Für die Dichte erhalten wir somit,

$$\begin{aligned} P[N_t = k] &= \int_0^t \frac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} dx - \int_0^t \frac{\lambda^{k+1}}{k!} x^k e^{-\lambda x} dx \\ &= e^{-\lambda t} \frac{(\lambda t)^k}{k!} \end{aligned}$$

Der letzte Schritt folgt durch Differenzieren der Integrale nach t , Zusammenfassen und anschließendes Integrieren. «

7.8 *Bemerkung.* Die Familie $\{N_t \mid t \in \mathbb{R}_+\}$ aus Satz 7.8 ist ein sogenannter **Poisson-Prozess**. ◊

8 Gesetze der großen Zahlen

8-A Konvergenzbegriffe

8.1 **Definition** Seien X_n, X reelle Zufallsvariablen (für $n \in \mathbb{N}$) auf einem W -Raum (Ω, \mathcal{A}, P) . Die Folge (X_n) heißt gegen X

1.) *vollständig konvergent*, wenn für jedes $\varepsilon > 0$ gilt

$$\sum_{n \in \mathbb{N}} P[|X_n - X| \geq \varepsilon] \rightarrow 0.$$

Schreibweise $X_n \xrightarrow{c} X$.

2.) *konvergent P -fast sicher* (P -f.s., f.s.), wenn

$$P[X_n \rightarrow X] = 1.$$

3.) *konvergent nach Wahrscheinlichkeit* oder *stochastisch konvergent*, wenn für jedes $\varepsilon > 0$ gilt

$$P[|X_n - X| \geq \varepsilon] \rightarrow 0.$$

Schreibweise $X_n \xrightarrow{P} X$,

4.) *konvergent im r -ten Mittel* ($r > 0$), wenn $E|X_n|^r, E|X|^r < \infty$ für $n \in \mathbb{N}$ und $E|X_n - X|^r \rightarrow 0$.

Spezialfall: $r = 2, \dots$ konvergenz im quadratischen Mittel.

5.) *konvergent nach Verteilung*, wenn für jede beschränkte stetige Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ gilt

$$Eg(X_n) \rightarrow Eg(X), \quad n \rightarrow \infty,$$

oder — hier äquivalent — wenn für die Verteilungsfunktion F_n und F von X_n bzw. X gilt

$$F_n(x) \rightarrow F(x), \quad n \rightarrow \infty,$$

in jedem Stetigkeitspunkt x von F . Schreibweise $X_n \xrightarrow{D} X$. \times

- 8.1 **Bemerkungen.** A. Seien F_n Verteilungsfunktionen für $n \in \mathbb{N}$, dann impliziert (F_n) konvergiert *nicht*, dass $\lim_{n \rightarrow \infty} F_n$ eine Verteilungsfunktion ist.
- B. Die Grenzverteilungsfunktion in Definition 8.1 e) ist eindeutig bestimmt.
- C. Die Konvergenz in e.) entspricht einer schwach*-Konvergenz der F_n im Dualraum von $C_b(\mathbb{R})$. \rightarrow

8.1 **Satz** Seien X_n, X reelle Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) .

- a) Falls $X_n \xrightarrow{c} X$, so gilt $X_n \rightarrow X$ f.s.
- b) Falls $X_n \rightarrow X$ f.s., so gilt $X_n \xrightarrow{P} X$.
- c) Falls für $r > 0$, $X_n \rightarrow X$ im r -ten Mittel, so gilt $X_n \xrightarrow{P} X$.
- d) Falls $X_n \xrightarrow{P} X$, so gilt $X_n \xrightarrow{D} X$. \times

» **Vorbetrachtung.** Seien X_n, X reelle Zufallsvariablen, so gilt

$$\begin{aligned} X_n \rightarrow X \text{ P-f.s.} &\Leftrightarrow P[\lim_{n \rightarrow \infty} X_n = X] = 1 \\ &\Leftrightarrow P\left(\bigcap_{0 < \varepsilon \in \mathbb{Q}} \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} [|X_m - X| < \varepsilon]\right) = 1 \end{aligned}$$

Wobei sich die letzte Äquivalenz direkt aus dem $\varepsilon - \delta$ -Kriterium für konvergente Folgen ergibt, wenn man sich für die Wahl von ε auf \mathbb{Q} zurückzieht.

Dies ist natürlich äquivalent zu

$$\begin{aligned} &P\left(\bigcup_{0 < \varepsilon \in \mathbb{Q}} \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} [|X_m - X| \geq \varepsilon]\right) = 0 \\ &\Leftrightarrow \forall \varepsilon > 0 : P\left(\underbrace{\bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} [|X_m - X| \geq \varepsilon]}_{=: A_n(\varepsilon)}\right) = 0 \end{aligned}$$

Nun ist $A_n \downarrow A$, d.h. wir können dies auch schreiben als

$$\begin{aligned} & \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(A_n(\varepsilon)) = 0 \\ \Leftrightarrow & \forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P\left(\sup_{m \geq n} |X_m - X| \geq \varepsilon\right) = 0. \end{aligned}$$

Weiterhin gilt

$$P[|X_n - X| \geq \varepsilon] \leq P[\sup_{m \geq n} |X_m - X| \geq \varepsilon] \leq \sum_{m \geq n} P[|X_m - X| \geq \varepsilon].$$

a) Falls $X_n \xrightarrow{c} X$, dann gilt $\sum_{m \geq n} P[|X_m - X| \geq \varepsilon] \rightarrow 0$ und damit auch

$$X_n \rightarrow X \text{ f.s.}$$

b) Falls $X_n \rightarrow X$ f.s., so gilt $P[\sup_{m \geq n} |X_m - X| \geq \varepsilon] \rightarrow 0$ und damit auch

$$P[|X_n - X| \geq \varepsilon] \rightarrow 0.$$

c) Unter Verwendung der Markov-Ungleichung erhalten wir

$$P[|X_n - X| \geq \varepsilon] \leq \frac{1}{\varepsilon^r} \mathbf{E}|X_n - X|^r$$

also $X_n \xrightarrow{P} X$, wenn $X_n \rightarrow X$ im r -ten Mittel.

d) Sei $g : \mathbb{R} \rightarrow \mathbb{R}$ beschränkt und stetig und $X_n \xrightarrow{P} X$. Es existiert also zu jeder Teilfolge $(X_{n'})$ eine konvergente Teilfolge $(X_{n''})$, die gegen X konvergiert. Somit gilt $g(X_{n''}) \rightarrow g(X)$ f.s.

Mit dem Satz von der dominierten Konvergenz erhalten wir somit

$$\mathbf{E}g(X_{n''}) \rightarrow \mathbf{E}g(X).$$

Nun verwenden wir den Satz aus der Analysis, dass eine Folge genau dann konvergiert, wenn jede Teilfolge eine konvergente Teilfolge besitzt, und alle diese Teilfolgen denselben Grenzwert haben und erhalten somit, $\mathbf{E}g(X_n) \rightarrow \mathbf{E}g(X)$, d.h. $X_n \xrightarrow{D} X$. «

8-B Schwache und starke Gesetze der großen Zahlen

- 8.2 **Definition** Eine Folge (X_n) von integrierbaren reellen Zufallsvariablen auf einem W -Raum (Ω, \mathcal{A}, P) genügt dem *schwachen* bzw. *starken Gesetz der großen Zahlen*, wenn

$$\frac{1}{n} \sum_{k=1}^n (X_k - \mathbf{E}X_k) \rightarrow 0, \quad n \rightarrow \infty \text{ nach Wahrscheinlichkeit bzw. } P\text{-f.s.} \quad \times$$

- 8.2 **Bemerkung.** Genügt eine Folge von Zufallsvariablen dem starken Gesetz der großen Zahlen, dann genügt sie auch dem schwachen Gesetz der großen Zahlen. Die Umkehrung gilt im Allgemeinen nicht. \rightarrow

- 8.2 **Kriterium von Kolmogorov für das starke Gesetz der großen Zahlen** Eine unabhängige Folge (X_n) quadratisch integrierbarer reeller Zufallsvariablen mit

$$\sum_{n=1}^{\infty} n^{-2} \mathbf{V}(X_n) < \infty$$

genügt dem starken Gesetz der großen Zahlen. \times

- » Wir wollen den Satz zunächst nur unter der stärkeren Bedingung

$$\sup_n \mathbf{V}(X_n) =: C < \infty$$

(jedoch unter der schwächeren Voraussetzung der paarweisen Unkorreliertheit der (X_n) anstatt der Unabhängigkeit). Einen vollständigen Beweis werden wir mit Hilfe der Martingalthorie geben können.

Wir zeigen nun

$$\forall \varepsilon > 0 : P \left[\left| \frac{1}{n} \sum_{j=1}^n X_j - \mathbf{E}X_j \right| \geq \varepsilon \right] \leq \frac{C}{n\varepsilon^2} \quad (1)$$

und

$$\frac{1}{n} \sum_{j=1}^n (X_j - \mathbf{E}X_j) \rightarrow 0$$

im quadratischen Mittel und P -f.s.. Setze $Y_j := X_j - \mathbf{E}X_j$, $Z_n := \frac{1}{n} \sum_{j=1}^n Y_j$. Mit Hilfe der Markov-Ungleichung erhalten wir

$$\varepsilon^2 P[|Z_n| \geq \varepsilon] \leq \mathbf{V}(Z_n) = \frac{1}{n^2} \mathbf{V}\left(\sum_{j=1}^n Y_j\right) \stackrel{\text{Bienayme}}{=} \frac{1}{n^2} \sum_{j=1}^n \mathbf{V}(Y_j) \leq \frac{C}{n}.$$

Somit folgt (1) und damit

$$\frac{1}{n} \sum_{j=1}^n (X_j - \mathbf{E}X_j) \rightarrow 0, \quad \text{im quadratischen Mittel.}$$

Es verbleibt die P -f.s. konvergenz zu zeigen. Betrachte dazu

$$\sum_{n=1}^{\infty} P[|Z_{n^2}| \geq \varepsilon] \leq \frac{C}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

also konvergiert Z_{n^2} vollständig und insbesondere auch P -f.s. gegen 0 (mit Satz 8.1). Sei nun $n \in \mathbb{N}$, setze

$$m(n) := \max \{m \in \mathbb{N} : m^2 \leq n\}$$

dann folgt $m(n)^2 \leq n < (m(n) + 1)^2$.

$$|Z_n| \leq \left| \frac{1}{n} \sum_{j=1}^{m(n)^2} Y_j \right| + \underbrace{\left| \frac{1}{n} \sum_{j=m(n)^2+1}^n Y_j \right|}_{=: R_n}$$

Nun ist $\left| \frac{1}{n} \sum_{j=1}^{m(n)^2} Y_j \right| \leq \left| \frac{1}{m(n)^2} \sum_{j=1}^{m(n)^2} Y_j \right| \rightarrow 0$ P -f.s.. Es genügt also zu zeigen, dass $R_n \rightarrow 0$ P -f.s.. Sei also $\varepsilon > 0$, dann folgt mit der Markov-Ungleichung und dem Satz von Bienayme

$$\begin{aligned} \varepsilon^2 P[|R_n| \geq \varepsilon] &\leq \mathbf{V}(R_n) \leq \frac{1}{n^2} \sum_{j=m(n)^2+1}^n \mathbf{V}(Y_j) \leq \frac{C}{n^2} (n - m(n)^2) \\ &\leq \frac{C}{n^2} ((m(n) + 1)^2 - 1 - m(n)^2) \\ &= \frac{C}{n^2} (2m(n)) \leq 2C \frac{\sqrt{n}}{n^2} = 2C n^{-3/2}. \end{aligned}$$

Da $\sum_{n \geq 0} n^{-3/2} < \infty$, folgt (R_n) konvergiert vollständig gegen 0. «

8.3 **Kolmogorovsches starkes Gesetz der großen Zahlen** Für eine unabhängige Folge (X_n) identisch verteilter integrierbarer reeller Zufallsvariablen gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mathbf{E}X_1, \quad n \rightarrow \infty, \text{ f.s. } \times$$

» Um die quadratische Integrierbarkeit der Zufallsvariablen zu gewährleisten, führen wir gestutzte Zufallsvariablen ein durch

$$X'_i := X_i \mathbf{1}_{\{|X_i| \leq i\}}$$

Insbesondere folgt aus der Unabhängigkeit der X_n auch die Unabhängigkeit der X'_n . Setze weiterhin

$$S'_n = \sum_{i=1}^n X'_i, \quad S_n := \sum_{i=1}^n X_i, \\ k_n := \lfloor \vartheta^n \rfloor$$

für beliebiges aber fest gewähltes $\vartheta > 1$.

$$\begin{aligned} \varepsilon^2 \sum_{n=1}^{\infty} P \left[\frac{1}{k_n} \left| S'_{k_n} - \mathbf{E}S'_{k_n} \right| \right] &\leq \sum_{n=1}^{\infty} \frac{1}{k_n^2} \mathbf{V}(S'_{k_n}) = \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{k=1}^{k_n} \mathbf{V}(X'_k) \\ &\leq \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{k=1}^{k_n} \mathbf{E}((X'_k)^2) \leq \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{k=1}^{k_n} \mathbf{E}(X_k^2 \mathbf{1}_{\{|X_k| \leq k\}}) \\ &\leq \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{k=1}^{k_n} \mathbf{E}(X_1^2 \mathbf{1}_{\{|X_1| \leq k_n\}}), \end{aligned}$$

denn da alle Zufallsvariablen dieselbe Verteilung besitzen, ist es für den Erwartungswert unerheblich, welche Zufallsvariable speziell ausgewählt wird. Aufgrund der Linearität von \mathbf{E} und dem Satz der monotonen Konvergenz folgt,

$$\leq \mathbf{E} \left(X_1^2 \sum_{n=1}^{\infty} \frac{1}{k_n} \mathbf{1}_{\{|X_1| \leq k_n\}} \right).$$

Setzen wir $n_0 := \min \{n \in \mathbb{N} : |X_1| \leq k_n\}$, so gilt

$$= \mathbf{E} \left(X_1^2 \sum_{n=n_0}^{\infty} \frac{1}{k_n} \mathbf{1}_{\{|X_1| \leq k_n\}} \right).$$

Außerdem $|X_1| \leq k_{n_0} = \lfloor \vartheta^{n_0} \rfloor \leq \vartheta^{n_0}$, also

$$\sum_{n=n_0}^{\infty} \frac{1}{k_n} \mathbf{1}_{\{|X_1| \leq k_n\}} \leq \sum_{n=n_0}^{\infty} \frac{1}{\lfloor \vartheta^n \rfloor} \leq \frac{1}{\vartheta^{n_0}} \sum_{n=0}^{\infty} \frac{2}{\vartheta^n} = \frac{2}{\vartheta^{n_0}} \frac{1}{1 - \frac{1}{\vartheta}} \leq \frac{2}{|X_1|} \frac{\vartheta}{\vartheta - 1},$$

wobei $\frac{1}{\lfloor \vartheta^{n_0} \rfloor} \leq \frac{2}{\vartheta^{n_0}}$, denn es gilt allgemein für positive reelle Zahlen $\lfloor p \rfloor > \frac{p}{2}$. Also erhalten wir für den Erwartungswert,

$$\mathbf{E} \left(X_1^2 \sum_{n=n_0}^{\infty} \frac{1}{k_n} \mathbf{1}_{\{|X_1| \leq k_n\}} \right) \leq \mathbf{E} \left(X_1^2 \frac{2}{|X_1|} \frac{\vartheta}{\vartheta - 1} \right) = \frac{2\vartheta}{\vartheta - 1} \mathbf{E} |X_1| < \infty,$$

da $\mathbf{E} |X_n| < \infty$ nach Voraussetzung.

Mit Satz 8.1 folgt

$$\frac{1}{k_n} (S'_{k_n} - \mathbf{E} S'_{k_n}) \rightarrow 0, \quad P\text{-f.s. für } n \rightarrow \infty$$

wobei

$$\frac{1}{k_n} \mathbf{E} S'_{k_n} = \frac{1}{k_n} \sum_{k=1}^{k_n} \mathbf{E}(X_1 \mathbf{1}_{\{|X_1| \leq k\}}).$$

Mit dem Satz von der dominierten Konvergenz folgt $\mathbf{E}(X_1 \mathbf{1}_{\{|X_1| \leq k\}}) \rightarrow \mathbf{E} X_1$ und damit,

$$\frac{1}{k_n} \sum_{k=1}^{k_n} \mathbf{E}(X_1 \mathbf{1}_{\{|X_1| \leq k\}}) \rightarrow \mathbf{E} X$$

nach dem Satz von Césaro-Stolz bzw. dem Cauchyschen Grenzwertsatz.

Außerdem konvergiert $\frac{S_{k_n}}{k_n} \rightarrow \mathbf{E} X_1$ P -f.s., denn

$$\begin{aligned} \sum_{n=1}^{\infty} P[X_n \neq X'_n] &= \sum_{n=1}^{\infty} P[|X_n| > n] \leq \sum_{n=1}^{\infty} \int_{n-1}^n P[|X_n| > t] dt \\ &\stackrel{\text{dom.konv}}{=} \int_0^{\infty} P[|X_n| > t] dt = \mathbf{E} X_1. \end{aligned}$$

Mit Lemma von Borel-Cantelli folgt dass X_n und X'_n für alle bis auf endlich viele n übereinstimmen, damit

$$\frac{S_{k_n} - S'_{k_n}}{k_n} \rightarrow 0, \quad P\text{-f.s. für } n \rightarrow \infty.$$

Zusammenfassend erhalten wir also

$$\frac{S_{k_n}}{k_n} = \underbrace{\frac{S_{k_n} - S'_{k_n}}{k_n}}_{-0 \text{ f.s.}} + \underbrace{\frac{S'_{k_n} - ES'_{k_n}}{k_n}}_{-0 \text{ f.s.}} + \underbrace{\frac{ES'_{k_n}}{k_n}}_{-EX_1 \text{ f.s.}} \rightarrow EX_1 \text{ f.s.}$$

Sei zunächst $X_i \geq 0$ für alle $i \in \mathbb{N}$. Dann gilt für alle i mit $k_n \leq i \leq k_{n+1}$,

$$\frac{k_n}{k_{n+1}} \frac{S_{k_n}}{k_n} \leq \frac{k_n S_i}{k_{n+1} k_n} \leq \frac{S_i}{i} \leq \frac{S_{k_{n+1}}}{i} \leq \frac{S_{k_{n+1}}}{k_{n+1}} \frac{k_{n+1}}{k_n}.$$

per definitionem ist

$$\frac{k_n}{k_{n+1}} = \frac{\lfloor \vartheta^n \rfloor}{\lfloor \vartheta^{n+1} \rfloor} \geq \frac{\vartheta^n - 1}{\vartheta^{n+1}} \rightarrow \frac{1}{\vartheta}$$

analog erhält man $\frac{k_{n+1}}{k_n} \rightarrow \vartheta$. Somit folgt

$$\frac{1}{\vartheta} EX_1 \leq \liminf_n \frac{S_n}{n} \leq \limsup_n \frac{S_n}{n} \leq \vartheta EX_1.$$

Da $\vartheta > 1$ beliebig und obige Gleichung für alle ϑ gilt, folgt durch den Übergang $\vartheta \downarrow 1$,

$$EX_1 = \lim_{n \rightarrow \infty} \frac{S_n}{n}, \quad P\text{-f.s. } n \rightarrow \infty.$$

Sei jetzt X_i reellwertig, mit $X_i = X_i^+ - X_i^-$, dann ist

$$\frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n X_i^+ - \frac{1}{n} \sum_{i=1}^n X_i^- \rightarrow EX_1^+ - EX_1^- = EX_1, \quad P\text{-f.s. für } n \rightarrow \infty. \quad \ll$$

8.3 *Bemerkungen.* A. In Satz 8.3 darf die Integrierbarkeitsvoraussetzung nicht weggelassen werden.

B. Die Voraussetzung der Unabhängigkeit in Satz 8.4 kann zur Voraussetzung der paarweisen Unabhängigkeit abgeschwächt werden (Etemadi).

C. In Satz 8.2 kann bei $\sum n^{-2} \mathbf{V}(X_n) (\log n)^2 < \infty$ die Unabhängigkeitsvoraussetzung zur Voraussetzung der paarweisen Unkorreliertheit (s.u.) abgeschwächt werden (Rademacher, Menchov). \rightarrow

8.4 **Satz von Tschebyschev** Eine Folge (X_n) quadratisch integrierbarer paarweise unkorrelierter, d.h. es gilt

$$\forall j \neq k : \mathbf{E}(X_j - \mathbf{E}X_j)(X_k - \mathbf{E}X_k) = 0,$$

reeller Zufallsvariablen mit

$$n^{-2} \sum_{k=1}^n \mathbf{V}(X_k) \rightarrow 0, \quad n \rightarrow \infty,$$

genügt dem schwachen Gesetz der großen Zahlen. \times

8.4 **Bemerkung.** Für eine unabhängige Folge (X_n) identisch verteilter reeller Zufallsvariablen mit $P[X_1 = 1] = p$, $P[X_1 = 0] = 1 - p$ (festes $p \in [0, 1]$) gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow p, \quad n \rightarrow \infty \quad P\text{-f.s. und nach Wahrscheinlichkeit.}$$

Borelsches Gesetz der großen Zahlen bzw. Bernoullisches schwaches Gesetz der großen Zahlen. \rightarrow

Bemerkung 8.4 stellt ein theoretisches Gegenstück zu der Erfahrungstatsache dar, dass im Allgemeinen die relative Häufigkeit eines Ereignisses bei großer Zahl der unter gleichen Bedingungen unabhängig durchgeführten Zufallsexperimente näherungsweise konstant ist.

Wir betrachten nun zwei Beispiele zu Satz 8.3 und Bemerkung 8.4.

BSP 1 Es seien X_1, X_2, \dots unabhängige identisch verteilte Zufallsvariablen Z_n mit existierenden (endlichen) $\mathbf{E}X_1 =: a$ und $\mathbf{V}(X_1) =: \sigma^2$. Aufgrund der beobachteten Realisierungen x_1, \dots, x_n von X_1, \dots, X_n wird a geschätzt durch

$$\bar{x}_n := \frac{1}{n} \sum_{i=1}^n x_i$$

das sogenannte **empirische Mittel** und σ^2 durch

$$s_n^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

die sogenannte **empirische Varianz**. Für

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

gilt $E\bar{X}_n = a$, $ES_n^2 = \sigma^2$ (sogenannte Erwartungstreue der Schätzung) und ferner für $n \rightarrow \infty$,

$$\bar{X}_n \rightarrow a \text{ f.s.}, \quad S_n^2 \rightarrow \sigma^2 \text{ f.s.},$$

sogenannte Konsistenz der Schätzfolgen. ■

» Wir zeigen $S_n^2 \rightarrow \sigma^2$ *P*-f.s.. Setzen wir $\mu = EX_1$, so gilt

$$\begin{aligned} S_n^2 &:= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu + \mu - \bar{X}_n)^2 \right) \\ &= \frac{n}{n-1} \left(\underbrace{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2}_{-\sigma^2 \text{ P-f.s.}} + \underbrace{\frac{2}{n} \sum_{i=1}^n (X_i - \mu) (\mu - \bar{X}_n)}_{-0 \text{ P-f.s.}} + \underbrace{\frac{1}{n} \sum_{i=1}^n (\mu - \bar{X}_n)^2}_{(\mu - \bar{X}_n)^2 - 0 \text{ f.s.}} \right), \\ &\rightarrow \sigma^2 \text{ f.s.} \quad \blacksquare \end{aligned}$$

BSP 2 Für eine Menge von n radioaktiven Atomen mit unabhängigen $\exp(\lambda)$ -verteilten Lebensdauern (mit Verteilungsfunktion F) gebe - bei festem $T > 0$ - die Zufallsvariable Z_n die zufällige Anzahl der Atome an, die von jetzt (Zeitpunkt 0) an bis zum Zeitpunkt T zerfallen. Die Wahrscheinlichkeit, dass ein gewisses Atom im Zeitintervall $[0, T]$ zerfällt, ist $p = F(T) = 1 - e^{-\lambda T}$. Nach Bemerkung 8.4 konvergiert mit Wahrscheinlichkeit Eins $Z_n/n \rightarrow p$ für $n \rightarrow \infty$. Wählt man T so, dass $F(T) = \frac{1}{2}$, d.h. $T = (\ln 2)/\lambda$ (T sog. Median von $\exp(\lambda)$), dann gilt

$$\text{f.s.} \quad \frac{Z_n}{n} \rightarrow \frac{1}{2}, \quad n \rightarrow \infty.$$

Dieses T wird als Halbwertszeit des radioaktiven Elements bezeichnet (nach T Zeiteinheiten ist i.A. bei großem n ungefähr die Hälfte der Atome zerfallen). ■

BSP 3 Aus der Analysis ist der Satz von Weierstraß bekannt, dass jede stetige Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$ auf einem kompakten Intervall gleichmäßiger Limes von Polynomen ist.

Wir wollen diesen Satz nun unter Verwendung des Kolmogorovschen starken Gesetz der großen Zahlen beweisen. Dazu verwenden wir Bernstein-Polynome, um stetige Funktionen über einem kompakten Intervall zu approximieren.

Sei $f : [0, 1] \rightarrow \mathbb{R}$ stetig. Die **Bernsteinpolynome n -ten Grades** sind definiert durch

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

Wir versehen also $[0, 1]$ mit einem Gitter mit Gitterabstand $\frac{1}{n}$, werten f an $\frac{k}{n}$ aus und multiplizieren mit speziellen Gewichten.

Seien U_1, \dots, U_n unabhängige auf $[0, 1]$ identisch gleichverteilte Zufallsvariablen. Wähle $p \in [0, 1]$ fest und setze

$$X_i := \mathbf{1}_{[0,p]}(U_i) = \mathbf{1}_{U_i \in [0,p]}.$$

Die X_i sind dann unabhängige und $b(1, p)$ -verteilte Zufallsvariablen.

$$\begin{aligned} \mathbf{E} \left(f \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[0,p]}(U_i) \right) \right) &= \mathbf{E} \left(f \left(\underbrace{\frac{1}{n} \sum_{i=1}^n X_i}_{\in \{0, 1/n, 2/n, \dots, 1\}} \right) \right) \\ &= \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^k = f_n(p) \end{aligned}$$

Sei $\varepsilon > 0$ beliebig. Da f auf dem kompakten Intervall $[0, 1]$ gleichmäßig stetig, existiert ein $\delta > 0$, so dass

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon, \quad \text{für } x, y \in [0, 1].$$

Somit gilt

$$\begin{aligned} |f_n(p) - f(p)| &= \left| \mathbf{E} \left(f \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[0,p]}(U_i) \right) \right) - f(p) \right| \\ &= \left| \mathbf{E} \left(f \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[0,p]}(U_i) \right) - f(p) \right) \right| \leq \mathbf{E} \left| f \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[0,p]}(U_i) \right) - f(p) \right| \end{aligned}$$

Falls nun $\left| \frac{1}{n} \sum_{i=0}^n \mathbf{1}_{[0,p]}(U_i) - p \right| < \delta$, ist der Erwartungswert $< \varepsilon$, andernfalls schät-

zen wir grob durch die Dreiecksungleichung ab,

$$\begin{aligned}
 &\leq \mathbf{E} \left| \varepsilon + 2 \|f\|_\infty \cdot \mathbf{1} \left[\left| \frac{1}{n} \sum_{i=0}^n \mathbf{1}_{[0,p]}(U_i) - p \right| \geq \delta \right] \right| \\
 &= \varepsilon + 2 \|f\|_\infty P \left[\left| \frac{1}{n} \sum_{i=0}^n \mathbf{1}_{[0,p]}(U_i) - p \right| \geq \delta \right] \leq \varepsilon + 2 \|f\|_\infty \frac{\mathbf{V}(\mathbf{1}_{[0,p]}(U_1))}{n\delta^2} \\
 &= \varepsilon + 2 \|f\|_\infty \frac{p(1-p)}{n\delta^2} \leq \varepsilon + 2 \|f\|_\infty \frac{1}{n\delta^2}
 \end{aligned}$$

Somit $f_n \rightarrow f$ gleichmäßig auf $[0, 1]$.

Bernstein-Polynome haben sehr angenehme Eigenschaften. Ist beispielsweise f monoton, so ist auch f_n monoton. Ist f konvex, so ist auch f_n konvex. ■

9 Zentrale Grenzwertsätze

9-A Schwache Konvergenz von Wahrscheinlichkeitsmaßen in \mathbb{R}

Im Folgenden sei stets $n \in \mathbb{N}$ und “ \rightarrow ” bezeichne die Konvergenz für $n \rightarrow \infty$.

- 9.1 **Definition** 1.) Seien Q_n, Q W -Maße auf der σ -Algebra \mathcal{B} der Borelschen Mengen in \mathbb{R} . Die Folge (Q_n) heißt gegen Q *schwach konvergent (weakly convergent)* — Schreibweise $Q_n \rightarrow Q$ schwach —, wenn für jede beschränkte stetige Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ gilt,

$$\int_{\mathbb{R}} g \, dQ_n \rightarrow \int_{\mathbb{R}} g \, dQ, \quad n \rightarrow \infty.$$

- 2.) Seien X_n, X reelle Zufallsvariablen (nicht notwendig auf demselben W -Raum definiert). Die Folge (X_n) heißt gegen X *nach Verteilung konvergent (convergent in distribution, convergent in law)* — Schreibweise $X_n \xrightarrow{D} X$ ($n \rightarrow \infty$) —, wenn $P_{X_n} \rightarrow P_X$ schwach, d.h. für jede beschränkte stetige Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ gilt

$$\int_{\mathbb{R}} g \, dP_{X_n} \rightarrow \int_{\mathbb{R}} g \, dP_X, \quad n \rightarrow \infty$$

oder (äquivalent nach dem Transformationssatz für Integrale)

$$\mathbf{E}g(X_n) \rightarrow \mathbf{E}g(X), \quad n \rightarrow \infty. \quad \times$$

- 9.1 **Bemerkungen.** A. In Definition 9.1a ist das Grenz- W -Maß Q eindeutig bestimmt.

- B. In Definition 9.1b können X_n, X durch reelle Zufallsvariablen X'_n, X' mit $P_{X'_n} = P_{X_n}, P_{X'} = P_X$ ersetzt werden. \rightarrow

Bevor wir die Eindeutigkeit des Grenz-W-Maßes beweisen, betrachten wir zunächst den Poissonschen Grenzwertsatz

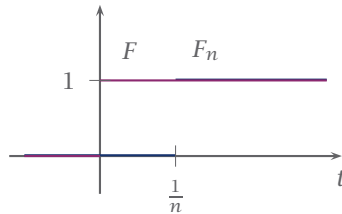
$$\forall k \in \mathbb{N}_0 : b(n, p_n; k) \rightarrow \pi(\lambda; k), \quad n \rightarrow \infty,$$

falls $np_n \rightarrow \lambda$. Dies ist (hier) äquivalent zu

$$\forall x \in \mathbb{R} : F_n(x) \rightarrow F(x),$$

wobei F_n Verteilungsfunktion einer $b(n, p)$ -verteilten und F Verteilungsfunktion einer $\pi(\lambda)$ -verteilten Zufallsvariable ist.

Seien X_n, X reelle Zufallsvariablen mit $X_n = \frac{1}{n}$ f.s., $X = 0$ f.s., so gilt $X_n \rightarrow X$ f.s. für $n \rightarrow \infty$.



9.1 Verteilungsfunktionen von X und X_n .

Für die Verteilungsfunktionen folgt

$$\forall x \in \mathbb{R} \setminus \{0\} : F_n(x) \rightarrow F(x),$$

denn 0 ist Unstetigkeitsstelle von F . Wir sehen an diesem Beispiel, dass die Konvergenz der Verteilungsfunktion in Unstetigkeitsstellen im Allgemeinen *nicht* fordern können. Wir können jedoch die Konvergenz nach Verteilung durch die Konvergenz der F_n in Stetigkeitspunkten charakterisieren.

Klassische Definition der Verteilungskonvergenz Seien X_n, X reelle Zufallsvariablen bzw. Q_n, Q W-Maße mit Verteilungsfunktion F_n bzw. F . Dann ist

$$X_n \xrightarrow{D} X,$$

bzw.

$$Q_n \rightarrow Q \text{ schwach},$$

wenn

$$\forall \text{ Stetigkeitspunkte } x \text{ von } F : F_n(x) \rightarrow F(x), \quad n \rightarrow \infty. \quad \times$$

Diese Definition ist jedoch nicht so allgemein wie die in 9.1, da sie sich nicht auf unendlichdimensionale Räume übertragen lässt.

» *Beweis der Eindeutigkeit des Grenz-W-Maßes.* Falls $Q_n \rightarrow Q^*$ schwach und $Q_n \rightarrow Q^{**}$ schwach, so gilt nach Satz 9.2 (noch zu zeigen),

$$F_n(x) \rightarrow F^*(x), \quad F_n(x) \rightarrow F^{**}(x)$$

für jede Stetigkeitsstelle x von F^* bzw. F^{**} . Aufgrund der rechtsseitigen Stetigkeit einer Verteilungsfunktion folgt,

$$\forall x \in \mathbb{R} : F^*(x) = F^{**}(x)$$

und damit auch $Q^* = Q^{**}$. «

9.1 **Satz** Seien X_n, X reelle Zufallsvariablen auf (Ω, \mathcal{A}, P) . Dann gilt

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X. \quad \times$$

» Sei $g : \mathbb{R} \rightarrow \mathbb{R}$ stetig und beschränkt. Angenommen $E(g(X_n) - g(X))$ konvergiert nicht gegen Null, es gibt also eine Teilfolge (n_k) , so dass

$$|E(g(X_{n_k}) - g(X))| > \varepsilon, \quad \forall k \in \mathbb{N}.$$

Aber da auch $X_{n_k} \xrightarrow{P} X$, besitzt X_{n_k} eine Teilfolge, die P -f.s. konvergiert. Da g stetig, folgt somit auch

$$g(X_{n_{k_l}}) \rightarrow g(X) \text{ } P\text{-f.s.}$$

Da außerdem g beschränkt, können wir den Satz von der dominierten Konvergenz anwenden und erhalten somit

$$Eg(X_{n_{k_l}}) \rightarrow Eg(X),$$

im Widerspruch zur Annahme, dass $Eg(X_n)$ nicht gegen $Eg(X)$ konvergiert. «

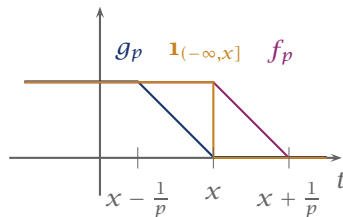
9.2 **Satz** Seien Q_n, Q W -Maße auf \mathcal{B} bzw. reelle Zufallsvariablen X_n, X mit Verteilungsfunktion F_n, F .

$Q_n \rightarrow Q$ schwach bzw. $X_n \xrightarrow{D} X$ gilt genau dann, wenn $F_n(x) \rightarrow F(x)$ für alle Stetigkeitspunkte x von F . \times

» “ \Rightarrow ”: Es gelte $X_n \xrightarrow{D} X$. Sei $D := \{x \in \mathbb{R} : F(x-) = F(x)\}$ die Menge der Stetigkeitspunkte von F . D ist dicht in \mathbb{R} , denn F ist monoton und daher ist D^c höchstens abzählbar. Sei $x \in D$. Für jedes $p \in \mathbb{N}$ setze

$$f_p(y) := \begin{cases} 1, & y \leq x, \\ p(x - y), & x \leq y \leq x + \frac{1}{p}, \\ 0, & x + \frac{1}{p} \leq y, \end{cases}$$

$$g_p(y) := \begin{cases} 1, & y \leq x - \frac{1}{p}, \\ p(x - y), & x - \frac{1}{p} \leq y \leq x, \\ 0, & x \leq y. \end{cases}$$



9.2 **Verteilungsfunktionen von X und X_n .**

$f_p, g_p : \mathbb{R} \rightarrow [0, 1]$ sind stetige, beschränkte Approximationen von $\mathbf{1}_{(-\infty, x]}(y)$, wobei

$$f_p(y) \downarrow \mathbf{1}_{(-\infty, x]}(y), \quad g_p(y) \uparrow \mathbf{1}_{(-\infty, x]}(y), \quad p \rightarrow \infty.$$

Da X_n nach Verteilung konvergiert, gilt

$$\forall p \in \mathbb{N} : \lim_{n \rightarrow \infty} \mathbf{E}f_p(X_n) = \mathbf{E}f_p(X),$$

$$\forall p \in \mathbb{N} : \lim_{n \rightarrow \infty} \mathbf{E}g_p(X_n) = \mathbf{E}g_p(X).$$

Nach Definition der Verteilungsfunktion ist $F_n(x) = \mathbf{E}\mathbf{1}_{(-\infty, x]}(X)$, also gilt auch

$$\mathbf{E}g_p(X_n) \leq F_n(x) \leq \mathbf{E}f_p(X_n).$$

Mit dem Lemma von Fatou folgt außerdem

$$\begin{aligned} \forall p \in \mathbb{N} : \mathbf{E}g_p(X) &\leq \liminf_{n \rightarrow \infty} \mathbf{E}g_p(X_n) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \\ &\leq \limsup_{n \rightarrow \infty} \mathbf{E}f_p(X_n) \leq \mathbf{E}f_p(X), \end{aligned} \quad (*)$$

da f_p beschränkt. Weiterhin folgt mit dem Satz von der monotonen Konvergenz,

$$\begin{aligned} \lim_{p \rightarrow \infty} \mathbf{E}f_p(X) &= \mathbf{E}\mathbf{1}_{(-\infty, x]}(X) = P[X \leq x] = F(x), \\ \lim_{p \rightarrow \infty} \mathbf{E}g_p(X) &= \mathbf{E}\mathbf{1}_{(-\infty, x)}(X) = P[X < x] = F(x-). \end{aligned}$$

Mit (*) folgt $F(x-0) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x)$. Für $x \in D$ gilt

$$F(x-) = F(x),$$

also auch

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

“ \Leftarrow ”: Diese Richtung ist etwas mühsamer. Hier approximiert man umgekehrt zur Hinrichtung eine stetige, beschränkte Funktion g durch Stufenfunktionen. Details findet man in [Jacod J, Protter P. Probability Essentials]. \llcorner

9.3 **Satz** Seien X_n, X reelle Zufallsvariablen auf (Ω, \mathcal{A}, P) . Ist X P -f.s. konstant, so gilt:

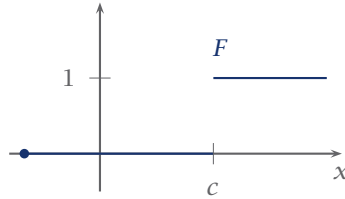
$$X_n \xrightarrow{D} X \Leftrightarrow X_n \xrightarrow{P} X. \quad \times$$

» “ \Leftarrow ”: Klar nach Satz 9.1.

“ \Rightarrow ”: Sei $X = c$ P -f.s. mit Verteilungsfunktion F . c ist die einzige Unstetigkeitsstelle von F .

Da $X_n \xrightarrow{D} X$ nach Voraussetzung folgt mit Satz 9.1,

$$\forall \mathbb{R} \setminus \{c\} : F_n(x) \rightarrow F(x).$$



9.3 Verteilungsfunktionen von X .

Sei $\varepsilon > 0$ beliebig aber fest, so gilt

$$\begin{aligned} P[|X_n - X| > \varepsilon] &= 1 - P[|X_n - X| \leq \varepsilon] \\ P[|X_n - X| \leq \varepsilon] &= P[c - \varepsilon \leq X_n \leq c + \varepsilon] \geq P[c - \varepsilon < X_n \leq c + \varepsilon] \\ &= \underbrace{F_n(c + \varepsilon)}_{\rightarrow F(c + \varepsilon)} - \underbrace{F_n(c - \varepsilon)}_{\rightarrow F(c - \varepsilon)} \rightarrow 1. \end{aligned}$$

Somit konvergiert $P[c - \varepsilon < X_n \leq c + \varepsilon] \rightarrow 1$, also $P[|X_n - X| > \varepsilon] \rightarrow 0$. «

9.4 **Satz** Seien X_n, Y_n, X reelle Zufallsvariablen auf (Ω, \mathcal{A}, P) . Gilt

$$X_n \xrightarrow{D} X, \quad |X_n - Y_n| \xrightarrow{P} 0,$$

so folgt

$$Y_n \xrightarrow{D} X. \quad \times$$

» Sei $g: \mathbb{R} \rightarrow \mathbb{R}$ stetig und beschränkt. Ohne Einschränkung können wir annehmen, dass g gleichmäßig stetig ist (siehe Satz 9.9). Außerdem gilt ohne Einschränkung

$$|X_n - Y_n| \rightarrow 0 \text{ P-f.s.}$$

ansonsten gehen wir zu einer Teilfolge (n_k) über. Betrachte

$$\mathbf{E}g(Y_n) = \mathbf{E}g(X_n) + \mathbf{E}[g(Y_n) - g(X_n)].$$

da g gleichmäßig stetig und $|X_n - Y_n| \rightarrow 0$ P-f.s. gilt $g(Y_n) - g(X_n) \rightarrow 0$ P-f.s.. $|g(Y_n) - g(X_n)| \leq c$, da g beschränkt, also können wir den Satz von der dominierten Konvergenz anwenden und erhalten $\mathbf{E}[g(Y_n) - g(X_n)] \rightarrow 0$.

Somit gilt

$$\mathbf{E}g(Y_n) = \underbrace{\mathbf{E}g(X_n)}_{\rightarrow \mathbf{E}g(X) \text{ P-f.s.}} + \underbrace{\mathbf{E}[g(Y_n) - g(X_n)]}_{\rightarrow 0 \text{ P-f.s.}} \rightarrow \mathbf{E}g(X) \text{ P-f.s.} \quad \ll$$

9.5 **Spezialfall des Darstellungssatzes von Skorokhod** Seien X_n, X reelle Zufallsvariablen mit $X_n \xrightarrow{D} X$. Dann existiert ein W -Raum $(\Omega^*, \mathcal{A}^*, P^*)$ — wobei man $\Omega^* = [0, 1]$, $\mathcal{A}^* = [0, 1] \cap \mathcal{B}$, $P^* = L$ -B-Maß auf \mathcal{A}^* wählen kann — und reelle Zufallsvariablen X_n^*, X^* auf $(\Omega^*, \mathcal{A}^*, P^*)$ derart, dass

$$\forall n \in \mathbb{N} : P_{X_n} = P_{X_n^*}^*, P_X = P_{X^*}^*, X_n^* \rightarrow X^* \text{ P}^*\text{-f.s.} \quad \times$$

» *Beweisidee.* Wähle als Ansatz für X_n^*, X^* die verallgemeinerte Inverse von F_n und F ,

$$X_n^*(\omega) := \inf \{t \in \mathbb{R} : F_n(t) \geq \omega\},$$

$$X^*(\omega) := \inf \{t \in \mathbb{R} : F(t) \geq \omega\}.$$

Nach Annahme konvergiert $X_n \xrightarrow{D} X$ und daher $F_n(x) \rightarrow F(x)$ in den Stetigkeitspunkten von F . Zu zeigen ist nun

$$X_n^*(\omega) \rightarrow X^*(\omega)$$

in den Stetigkeitspunkten ω von X^* , also λ -f.ü..

Formaler Beweis. Wähle $\Omega^* = (0, 1)$, $\mathcal{A}^* = (0, 1) \cap \mathcal{B}$ und $P^* = \lambda|_{\mathcal{A}^*}$. Seien F_n und F die Verteilungsfunktionen von X_n und X . Nun setzen wir

$$X_n^*(\omega) := \inf \{t \in \mathbb{R} : F_n(t) \geq \omega\},$$

$$X^*(\omega) := \inf \{t \in \mathbb{R} : F(t) \geq \omega\},$$

d.h. X_n^* und X^* sind die verallgemeinerten Inversen von F_n bzw. F . Somit sind X_n^* und X^* monotone Funktionen und damit insbesondere messbar. Wähle für die Verteilung von X^* ,

$$P^*[X^* \leq x] := F(x) = P[X \leq x],$$

so gilt $P_{X_n^*}^* = P_X$, entsprechend für X_n und X_n^* . Somit ist auch klar, dass X^* höchstens abzählbar viele Unstetigkeitsstellen hat.

Es genügt nun zu zeigen $X_n^* \rightarrow X^*$ für alle Stetigkeitspunkte ω von X^* . Sei ω Stetigkeitspunkt von X^* und $\varepsilon > 0$. Es existieren dann Stetigkeitspunkte x_1 und x_2 von F mit

$$x_1 < X^*(\omega) < x_2 \text{ und } x_2 - x_1 < \varepsilon, \quad (*)$$

Hierbei gilt auch $F(x_1) < \omega < F(x_2)$, da ω Stetigkeitspunkt von X^* , sowie $F_n(x_1) \rightarrow F(x_1)$ und $F_n(x_2) \rightarrow F(x_2)$, da x_1, x_2 Stetigkeitspunkte von F und $X_n \xrightarrow{D} X$. Also

$$\exists n_0 \in \mathbb{N} \forall n \geq n_0 : F_n(x_1) < \omega < F_n(x_2)$$

und damit auch

$$\forall n \geq n_0 : x_1 \leq X_n^*(\omega) \leq x_2, \quad (**)$$

wobei lediglich benötigt wird, dass ω Stetigkeitspunkt von X^* , nicht unbedingt von X_n^* .

Aus (*) und (**) folgt nun $|X_n^*(\omega) - X^*(\omega)| < \varepsilon$. «

9.6 **Satz von der stetigen Abbildung** Seien X_n, X reelle Zufallsvariablen mit $X_n \xrightarrow{D} X$. Sei die Abbildung

$$h : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$$

P_X -f.ü. stetig. Dann gilt $h(X_n) \xrightarrow{D} h(X)$. \times

» Sei $D := \{x \in \mathbb{R} : h \text{ unstetig in } x\}$. Den Beweis $D \in \mathcal{B}$ überspringen wir. Nach Voraussetzung ist $P_X(D) = 0$.

Nach Satz 9.5 existieren ein W-Raum $(\Omega^*, \mathcal{A}^*, P^*)$ und Zufallsvariablen $X_n^*, X^* : (\Omega^*, \mathcal{A}^*, P^*) \rightarrow (\mathbb{R}, \mathcal{B})$ mit $X_n^* \rightarrow X^*$ P^* -f.ü. und

$$P_{X_n^*}^* = P_{X_n}, \quad P_{X^*}^* = P_X.$$

Da h stetig, gilt

$$h(X_n^*(\omega)) \rightarrow h(X^*(\omega))$$

für $\omega \in D^c$, aber

$$P^*(\{\omega \in \Omega^* : X^*(\omega) \in D\}) = P^*[X^* \in D] = P_{X^*}^*(D) = P_X(D) = 0,$$

also gilt $h(X_n^*) \rightarrow h(X^*)$ P^* f.s. Somit folgt nach Satz 9.1 $h(X_n^*) \xrightarrow{D} h(X^*)$, also

$$P_{h(X_n^*)}^* \rightarrow P_{h(X^*)}^* \quad \text{schwach.}$$

Nun ist $P_{h(X_n^*)}^* = (P_{X_n^*}^*)_h = (P_{X_n})_h = P_{h(X_n)}$, analog $P_{h(X^*)}^* = P_{h(X^*)}$. Somit

$$P_{h(X_n)} \rightarrow P_{h(X)} \quad \text{schwach.} \quad \ll$$

Wie wir gesehen haben, vertauscht der Limes Operator der Verteilungskonvergenz im Allgemeinen nicht mit der Addition oder der Multiplikation. Für spezielle Zufallsvariablen erhalten wir jedoch Vertauschbarkeit.

9.7 **Satz von Slutsky** Seien X_n, Y_n und X reelle Zufallsvariablen auf (Ω, \mathcal{A}, P) und $c \in \mathbb{R}$.

$$\left. \begin{array}{l} X_n \xrightarrow{D} X, \\ Y_n \xrightarrow{P} c, \end{array} \right\} \Rightarrow \begin{cases} X_n + Y_n \xrightarrow{D} X + c, \\ Y_n X_n \xrightarrow{D} cX. \quad \times \end{cases}$$

» Sei $c \in \mathbb{R}$ und $Z_n := X_n + (Y_n - c)$, so gilt

$$|X_n - Z_n| = |Y_n - c| = 0 \text{ f.s.,}$$

insbesondere $|X_n - Z_n| \xrightarrow{P} 0$. Somit ist 9.4 anwendbar und daher,

$$X_n + (Y_n - c) = Z_n \xrightarrow{D} X.$$

Nun wenden wir den Satz von der stetigen Abbildung an mit

$$h : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto x + c,$$

so gilt

$$X_n + Y_n = h(X_n + (Y_n - c)) \xrightarrow{D} h(X) = X + c. \quad \ll$$

9.8 **Satz** Für reellwertige Zufallsvariablen X_n, X mit Dichten f_n bzw. f gilt:

$$f_n \rightarrow f \text{ } \lambda\text{-f.}\ddot{u}. \Rightarrow X_n \xrightarrow{D} X. \quad \times$$

» Sei $g \in C_b(\mathbb{R})$. Wähle $c \in \mathbb{R}$, so dass $|g| \leq c$, so folgt

$$g + c \geq 0.$$

Da $dP_X = f \, d\lambda$ erhalten wir,

$$\int_{\mathbb{R}} g \, dP_X = -c + \int_{\mathbb{R}} (c + g) \, dP_X = -c + \int_{\mathbb{R}} (c + g) f \, d\lambda.$$

Nun ist $f = \lim_{n \rightarrow \infty} f_n = \liminf_{n \rightarrow \infty} f_n$ λ -f.ü. nach Voraussetzung. Da f, f_n Dichten und daher positiv, können wir das Lemma von Fatou anwenden und erhalten,

$$\begin{aligned} \int_{\mathbb{R}} g \, dP_X &\leq -c + \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} (c + g) f_n \, d\lambda = \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} g f_n \, d\lambda \\ &\leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}} g f_n \, d\lambda = - \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} -g f_n \, d\lambda \\ &= c - \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} \underbrace{(c - g)}_{\geq 0} f_n \, d\lambda \leq c - \int_{\mathbb{R}} \liminf_{n \rightarrow \infty} (c - g) f_n \, d\lambda \\ &= \int_{\mathbb{R}} g f \, d\lambda. \end{aligned}$$

Es gilt also

$$\int_{\mathbb{R}} g \, dP_X \leq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} g f_n \, d\lambda \leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}} g f_n \, d\lambda \leq \int_{\mathbb{R}} g \, dP_X.$$

Somit existiert $\lim_{n \rightarrow \infty} \int_{\mathbb{R}} g f_n \, d\lambda$ und es gilt

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} g \, dP_{X_n} = \int_{\mathbb{R}} g \, dP_X. \quad \llcorner$$

Um $X_n \xrightarrow{D} X$ nachzuweisen, müssen wir für jedes stetige beschränkte g nachweisen

$$\mathbf{E}g(X_n) \rightarrow \mathbf{E}g(X).$$

Der nächste Satz zeigt, dass es genügt, sich auf eine kleinere Menge von Funktionen zurückzuziehen.

Definition Eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ heißt *Lipschitz-stetig*, falls eine Konstante $L \in \mathbb{R}$ existiert, so dass

$$|g(x) - g(y)| \leq L |x - y|, \quad \forall x, y \in \mathbb{R}. \quad \times$$

Jede lipschitzstetige Funktion ist gleichmäßig stetig und jede gleichmäßig stetige Funktion ist stetig. Die Umkehrungen gelten im Allgemeinen *nicht*.

Die gleichmäßig stetigen Funktionen auf \mathbb{R} bilden eine echte Teilmenge der stetigen Funktionen.

9.9 **Satz** Für reellwertige Zufallsvariablen X_n, X gilt:

$$X_n \xrightarrow{D} X \Leftrightarrow \mathbf{E}f(X_n) \rightarrow \mathbf{E}f(X)$$

für alle beschränkten gleichmäßig stetigen Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$. \times

» “ \Rightarrow ”: Klar, denn die Aussage gilt nach Voraussetzung für stetige beschränkte und somit insbesondere für gleichmäßig stetige beschränkte Funktionen.

“ \Leftarrow ”: Sei $f \in C_b(\mathbb{R})$. Zu zeigen ist $\mathbf{E}f(X_n) \rightarrow \mathbf{E}f(X)$, da f lediglich stetig und beschränkt. Setzen wir

$$\alpha := \sup_{x \in \mathbb{R}} |f(x)| = \|f\|_\infty.$$

Wir zeigen nun, dass für jedes $i \in \mathbb{N}$ eine Lipschitz-stetige Funktion $g_i : \mathbb{R} \rightarrow \mathbb{R}$ existiert mit

$$-\alpha \leq g_i \leq g_{i+1} \leq \alpha, \quad \forall x \in \mathbb{R} : g_i(x) \rightarrow f(x), \quad i \rightarrow \infty. \quad (*)$$

Haben wir die Existenz der g_i gezeigt, so gilt für $i \in \mathbb{N}$

$$\liminf_n \mathbf{E}f(X_n) \geq \liminf_{n \rightarrow \infty} \mathbf{E}g_i(X_n) = \lim_{n \rightarrow \infty} \mathbf{E}g_i(X_n) = \mathbf{E}g_i(X), \quad (**)$$

denn die g_i sind gleichmäßig stetig und daher konvergiert der Erwartungswert nach Voraussetzung.

Da die g_i durch α beschränkt sind, ist $g_i(X) + \alpha \geq 0$ und es gilt

$$\lim_{i \rightarrow \infty} \mathbf{E}(g_i(X) + \alpha) \stackrel{\text{mon.konv}}{=} \mathbf{E}(f(X) + \alpha).$$

Somit gilt auch $\lim_{i \rightarrow \infty} \mathbf{E}(g_i(X)) = \mathbf{E}f(X)$. Zusammen mit (**) erhalten wir also

$$\liminf_{n \rightarrow \infty} \mathbf{E}f(X_n) \geq \mathbf{E}f(X).$$

Analog erhalten wir mit f ersetzt durch $-f$,

$$\limsup_{n \rightarrow \infty} \mathbf{E}f(X_n) \leq \mathbf{E}f(X).$$

Insgesamt gilt also $\mathbf{E}f(X_n) \rightarrow \mathbf{E}f(X)$ und die Behauptung ist gezeigt.

Wir müssen also noch die Behauptung (*) nachweisen. Es genügt, eine Folge $(h_k)_{k \in \mathbb{N}}$ Lipschitz-stetiger Funktionen zu finden mit

$$h_k \geq -\alpha \text{ und } \forall x \in \mathbb{R} : \sup_{k \in \mathbb{N}} h_k(x) = f(x).$$

Denn dann leistet g_i mit $g_i(x) = \max\{h_1(x), \dots, h_i(x)\}$ das Gewünschte, denn das Maximum über endlich viele Lipschitz-stetige Funktionen ist wieder Lipschitz-stetig.

Wir können ohne Einschränkung davon ausgehen, dass $f \geq 0$, ansonsten ersetzen wir f durch $\tilde{f} = f + \alpha \geq 0$. Wähle $A \in \mathcal{B}$ und setze

$$d_A(x) = \inf\{|x - y| : y \in A\}.$$

d_A ist der Hausdorffabstand des Punktes x von der Menge A . Sei $r \geq 0$ rational, $m \in \mathbb{N}$ und

$$h_{m,r}(x) = \min\left\{r, (m \cdot d_{\{t:f(t) \leq r\}}(x))\right\},$$

so sind die $h_{m,r}$ Lipschitz, denn

$$|h_{m,r}(x) - h_{m,r}(y)| \leq m \left| d_{\{t:f(t) \leq r\}}(x) - d_{\{t:f(t) \leq r\}}(y) \right| \leq m |x - y|,$$

außerdem sind die $h_{m,r}$ beschränkt, denn $|h_{m,r}| \leq r$, und $|h_{m,r}|(x) = 0$ für x mit $f(x) \leq r$. Insbesondere

$$0 \leq h_{m,r}(x) \leq f(x), \quad \forall x \in \mathbb{R}.$$

Wähle $x \in \mathbb{R}$, $\varepsilon > 0$ beliebig aber fest. Wähle außerdem $0 \leq r \in \mathbb{Q}$ so, dass

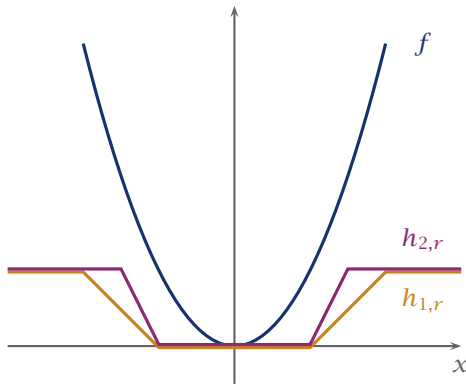
$$f(x) - \varepsilon < r < f(x).$$

Es gilt $f(y) > r$ für alle y aus einer hinreichend kleinen Umgebung von x , da f stetig. Somit folgt

$$d_{\{t:f(t) \leq r\}}(x) > 0,$$

also ist auch

$$h_{m,r}(x) = r \begin{cases} < f(x), \\ > f(x) - \varepsilon, \end{cases}$$



9.4 $h_{m,r}(x)$ für $f(x) = x^2$, $r = 1$ und $m = 1, 2$.

für m hinreichend groß.

Die Menge $\{h_{m,r} : m \in \mathbb{N}, 0 \leq r \in \mathbb{Q}\}$ ist abzählbar. Sei

$$\{h_k : k \in \mathbb{N}\}$$

eine Abzählung dieser Menge. Nach Konstruktion gilt nun

$$\sup_{k \in \mathbb{N}} h_k(x) \begin{cases} \leq f(x), \\ \geq f(x) - \varepsilon, \end{cases}$$

also $\sup_{k \in \mathbb{N}} h_k(x) = f(x)$, da $\varepsilon > 0$ beliebig. «

9.10 **Satz von Lévy-Cramér; Stetigkeitssatz** Seien Q_n, Q W -Maße auf \mathcal{B} mit charakteristischen Funktionen $\varphi_n, \varphi : \mathbb{R} \rightarrow \mathbb{C}$. Dann gilt:

$$Q_n \rightarrow Q \text{ schwach} \Leftrightarrow \forall u \in \mathbb{R} : \varphi_n(u) \rightarrow \varphi(u). \quad \times$$

» “ \Rightarrow ”: Sei φ_n die charakteristische Funktion von Q_n , also

$$\varphi_n(u) = \int_{\mathbb{R}} e^{iux} dQ_n(x) = \int_{\mathbb{R}} \cos(ux) dQ_n(x) + i \int_{\mathbb{R}} \sin(ux) dQ_n(x).$$

sin und cos sind beschränkte Funktionen also,

$$\begin{aligned} \varphi_n(u) &\rightarrow \int_{\mathbb{R}} \cos(ux) dQ(x) + i \int_{\mathbb{R}} \sin(ux) dQ(x) = \int_{\mathbb{R}} e^{iux} dQ(x) \\ &= \varphi(u). \end{aligned}$$

“ \Leftarrow ”: Seien Y_n, Y_0, X unabhängige Zufallsvariablen auf (Ω, \mathcal{A}, P) mit $Y_n \sim Q_n, Y_0 \sim Q, X \sim N(0, 1)$. Dann gilt für alle $\alpha > 0$,

$$P_{Y_n + \alpha X} = P_{Y_n} * P_{\alpha X}.$$

Nach Übungsaufgabe 40 besitzt $P_{Y_n + \alpha X}$ die Dichte

$$g_{n,\alpha}(x) = \frac{1}{2\pi} \int e^{-ux} \varphi_n(u) e^{-\frac{(\alpha u)^2}{2}} du.$$

Der Integrand besitzt eine integrierbare Majorante, denn

$$\left| e^{-ux} \varphi_n(u) e^{-\frac{(\alpha u)^2}{2}} \right| \leq e^{-ux} e^{-\frac{(\alpha u)^2}{2}},$$

also können wir den Satz von der dominierten Konvergenz anwenden und erhalten,

$$\forall x \in \mathbb{R} : g_{n,\alpha}(x) \rightarrow \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ux} \varphi(u) e^{-\frac{(\alpha x)^2}{2}} du.$$

Unter Verwendung des Satzes 9.8 erhalten wir

$$Y_n + \alpha X \xrightarrow{D} Y_0 + \alpha X. \quad (*)$$

Zum Nachweis von $Y_n \xrightarrow{D} Y_0$ genügt es nach Satz 9.9 zu zeigen, dass für beliebige beschränkte gleichmäßig stetige Funktionen gilt $f : \mathbb{R} \rightarrow \mathbb{R}$

$$\int f(Y_n) dP \rightarrow \int f(Y_0) dP.$$

Seien also f wie vorausgesetzt, $\varepsilon > 0$ und $\delta > 0$, so dass

$$|f(y+x) - f(y)| < \frac{\varepsilon}{6}, \quad \forall x, y \in \mathbb{R} \text{ mit } |x| < \delta.$$

Wähle außerdem $\alpha > 0$ mit

$$P[|\alpha X| \geq \delta] \leq \frac{\varepsilon}{12 \|f\|_{\infty}}.$$

Nach (*) existiert ein n_0 so, dass

$$\forall n \geq n_0 : \left| \int_{\Omega} f(Y_n + \alpha X) - f(Y_0 + \alpha X) dP \right| < \frac{\varepsilon}{3}.$$

Dann gilt für $n \geq n_0$,

$$\begin{aligned} \left| \int_{\Omega} f(Y_n) - f(Y_0) \, dP \right| &\leq \underbrace{\int_{\Omega} |f(Y_n) - f(Y_n + \alpha X)| \, dP}_{(1)} \\ &+ \underbrace{\left| \int_{\Omega} f(Y_n + \alpha X) - f(Y_0 + \alpha X) \, dP \right|}_{(2)} \\ &+ \underbrace{\int_{\Omega} |f(Y_0 + \alpha X) - f(Y_0)| \, dP}_{(3)}. \end{aligned}$$

(1): Wir nutzen die gleichmäßige Stetigkeit von f aus,

$$\begin{aligned} \int_{\Omega} |f(Y_n) - f(Y_n + \alpha X)| \, dP &= \int_{|\alpha X| \geq \delta} |f(Y_n) - f(Y_n + \alpha X)| \, dP \\ &+ \int_{|\alpha X| < \delta} |f(Y_n) - f(Y_n + \alpha X)| \, dP \\ &< 2 \|f\|_{\infty} P[|\alpha X| \geq \delta] + \frac{\varepsilon}{6} P[|\alpha X| < \delta] \\ &\leq 2 \|f\|_{\infty} \frac{\varepsilon}{12 \|f\|_{\infty}} + \frac{\varepsilon}{6} \cdot 1 = \frac{\varepsilon}{3}. \end{aligned}$$

(2) $< \frac{\varepsilon}{3}$ für $n \geq n_0$.

(3) $< \frac{\varepsilon}{6} P[|\alpha X| < \delta] + 2 \|f\|_{\infty} P[|\alpha X| \geq \delta] \leq \frac{\varepsilon}{3}$.

Somit $E f(Y_n) \rightarrow E f(Y_0)$, d.h. $Y_n \xrightarrow{D} Y_0$. «

9.2 *Bemerkung.* Die obigen Definitionen und Sätze lassen sich auf W-Maße auf \mathcal{B}_k bzw. k -dim. Zufallsvektoren übertragen. \rightarrow

9-B Zentrale Grenzwertsätze

Aussagen über die Approximation von Verteilungen, insbesondere der Verteilungen von Summen von Zufallsvariablen, durch die Normalverteilung im Sinne der schwachen Konvergenz werden als zentrale Grenzwertsätze bezeichnet.

9.11 **Zentraler Grenzwertsatz von Lindeberg-Lévy im 1-dim. Fall** Sei $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Folge identisch verteilter quadratisch integrierbarer reeller Zufallsvariablen mit $E X_1 = a$, $V(X_1) = \sigma^2$ mit $\sigma > 0$. Dann

$$\frac{1}{\sqrt{n}\sigma} \sum_{k=1}^n (X_k - a) \xrightarrow{D} N(0, 1)\text{-verteilte reelle Zufallsvariable. } \times$$

» Ohne Einschränkung können wir $a = 0$ und $\sigma = 1$ annehmen (andernfalls ersetzen wir X_k durch $\frac{X_k - a}{\sigma}$). X_1 habe die charakteristische Funktion φ , dann besitzt

$$\sum_{k=1}^n X_k$$

nach Satz 6.4 die charakteristische Funktion φ^n . Somit hat

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k$$

die charakteristische Funktion $\varphi^n\left(\frac{\cdot}{\sqrt{n}}\right)$. Nach Satz 9.10 genügt es zu zeigen,

$$\forall x \in \mathbb{R} : \varphi^n\left(\frac{u}{\sqrt{n}}\right) \rightarrow e^{-\frac{u^2}{2}}.$$

$e^{-\frac{u^2}{2}}$ ist nach Bemerkung 6.4 die charakteristische Funktion einer $N(0, 1)$ -Verteilung.

Nach Satz 6.6 besitzt φ wegen $\text{EX}_1^2 < \infty$ eine stetige 2. Ableitung und daher nach dem Satz von Taylor eine Darstellung

$$\begin{aligned} \varphi(u) &= \varphi(0) + u\varphi'(0) + \frac{u^2}{2}\varphi''(0) + \rho(u) \\ &= 1 + iu\text{EX}_1 - \frac{u^2}{2}\text{EX}_1^2 + \rho(u), \end{aligned}$$

mit einem Restterm $\rho(u)$ der Ordnung $o(u^2)$, d.h.

$$\lim_{u \rightarrow 0} \frac{\rho(u)}{u^2} = 0.$$

Nach Voraussetzung ist $\text{EX}_1 = a = 0$ und $\text{EX}_1(X_1 - 1) = \text{EX}_1^2 = \text{VX}_1 = 1$ und daher

$$\varphi(u) = 1 - \frac{u^2}{2} + \rho(u).$$

Für $u \in \mathbb{R}$ beliebig aber fest gilt

$$\varphi^2\left(\frac{u}{\sqrt{n}}\right) = \left(1 - \frac{u^2}{2n} + \rho\left(\frac{u}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{\frac{u^2}{2} + n\rho\left(\frac{u}{\sqrt{n}}\right)}{n}\right)^n$$

Nach Voraussetzung $n\rho\left(\frac{u}{\sqrt{n}}\right) \rightarrow 0$ für $n \rightarrow \infty$, also

$$\left(1 - \frac{\frac{u^2}{2} + n\rho\left(\frac{u}{\sqrt{n}}\right)}{n}\right)^n \rightarrow e^{-\frac{u^2}{2}}. \quad \ll$$

Als unmittelbares Korollar erhalten wir.

9.2 **Zentraler Grenzwertsatz von de Moivre und Laplace** Für eine unabhängige Folge $(X_n)_{n \in \mathbb{N}}$ identisch verteilter reeller Zufallsvariablen auf (Ω, \mathcal{A}, P) mit

$$P[X_1 = 1] = p, \quad P[X_1 = 0] = 1 - p =: q, \quad (0 < p < 1)$$

gilt für $n \rightarrow \infty$

$$\forall \alpha < \beta \in \mathbb{R} : P \left[\alpha \underset{(\equiv)}{<} \frac{\sum_{k=1}^n X_k - np}{\sqrt{npq}} \underset{(\equiv)}{<} \beta \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-t^2/2} dt. \quad \times$$

BSP 1 Volksabstimmung zu Vorschlägen A und B. Eine resolute Minderheit von 3.000 Personen stimmt für A. Weitere 1.000.000 Personen stimmen zufällig ab.

Wie groß ist die Wahrscheinlichkeit, dass A angenommen wird? Bezeichne die "gleichgültigen" Wähler mit den Nummern $k = 1, 2, 3, \dots, 1.000.000$,

$$X_k = \begin{cases} 1, & \text{Wähler wählt A,} \\ 0, & \text{Wähler wählt B.} \end{cases}$$

Somit sind X_1, X_2, \dots unabhängig und $b(1, 1/2)$ -verteilt. Seien $n = 1.000.000$, $r = 3.000$, so wird der Vorschlag A angenommen, wenn

$$\sum_{k=1}^n X_k + r > n - \sum_{k=1}^n X_k,$$

d.h. genau dann, wenn

$$\sum_{k=1}^n X_k > \frac{n-r}{2} = 498.500.$$

Die Wahrscheinlichkeit dafür ist gegeben durch,

$$P \left[\sum_{k=1}^n X_k > 498.500 \right] = P \left[\frac{\sum_{k=1}^n X_k - \frac{n}{2}}{\sqrt{n \frac{1}{4}}} > \frac{498.500 - 500.000}{500} = -3 \right].$$

Korollar 9.2 besagt, dass $\frac{\sum_{k=1}^n X_k - \frac{n}{2}}{\sqrt{n \frac{1}{4}}}$ annähernd $N(0, 1)$ -verteilt ist, d.h.

$$P \left[\frac{\sum_{k=1}^n X_k - \frac{n}{2}}{\sqrt{n \frac{1}{4}}} > -3 \right] = 1 - P \left[\frac{\sum_{k=1}^n X_k - \frac{n}{2}}{\sqrt{n \frac{1}{4}}} \leq -3 \right] \\ \approx 1 - \Phi(-3) = 0.9986.$$

Obwohl lediglich 3.000 Personen sicher für A stimmen, wird der Vorschlag mit einer Wahrscheinlichkeit von 99.86% angenommen.

Dies ist auch der Grund dafür, dass Vorschläge, die den Großteil der Abstimmenden nicht interessieren, bereits von einer kleinen Gruppe von Entschlossenen durchgesetzt werden können. ■

Normal- oder Poisson-Approximation?

Seien X_n unabhängige $b(1, p)$ -verteilte Zufallsvariablen. Nach dem Grenzwertsatz von de Moivre-Laplace gilt

$$\underbrace{\sum_{i=1}^n \frac{X_i - p}{\sqrt{np(1-p)}}}_{\stackrel{!}{=} N(0,1)\text{-vert. ZV}} \approx N(0, 1)\text{-verteilt für "große" } n. \quad (1)$$

Andererseits gilt nach dem Grenzwertsatz von Poisson

$$\underbrace{\sum_{j=1}^n X_j}_{\stackrel{!}{=} \pi(\lambda)} \approx \pi(np_n), \text{ falls } n \text{ "groß" und } p_n \text{ "klein",} \quad (2)$$

und $np_n \rightarrow \lambda \in (0, \infty)$ (siehe Übungsaufgabe 45).

Im Fall (1) sind die Summanden dem Betrag nach klein für großes n , während im Fall (2) nur die Wahrscheinlichkeit, $p_n = P[X_n = 1]$, klein ist, dass die Summanden *nicht* klein sind.

Als *Faustregel* gilt

(a) Normal-Approximation ist "gut", falls $np(1-p) \geq 9$.

(b) Poisson-Approximation ist "gut", falls $n \geq 50$ und $p \leq 0.05$.

Im Allgemeinen können Poisson- und Normal-Approximation (für kleine n) nicht gleichzeitig angewendet werden.

- 9.3 *Bemerkungen.* A. Seien X_n unabhängige, identisch verteilte reelle Zufallsvariablen mit endlichen Varianzen und $\mu = EX_1$, so gilt nach dem starken Gesetz der großen Zahlen,

$$\frac{1}{n} S_n := \frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mu, \quad P\text{-f.s. und } L^2.$$

Sei $\varepsilon > 0$. Wie groß ist nun n_0 zu wählen, so dass

$$\left| \frac{1}{n} S_n - \mu \right| < \varepsilon, \quad n \geq n_0?$$

Mit Standardmethoden der Analysis lässt sich die Fragestellung umformulieren zu

$$\exists \alpha > 0 \exists c \neq 0 : \lim_{n \rightarrow \infty} n^\alpha \left| \frac{S_n}{n} - \mu \right| = c, \quad P\text{-f.s.}?$$

Ein solches α existiert nicht, da nach dem zentralen Grenzwertsatz

$$n^{\frac{1}{2}} \left(\frac{S_n}{n} - \mu \right) \stackrel{D}{\rightarrow} N(0, \mathbf{V}(X_1))\text{-verteilte Zufallsvariable.}$$

Für $\alpha < \frac{1}{2}$ liegt nach dem Satz von Slutsky Verteilungs-Konvergenz gegen Null vor, also auch nach Wahrscheinlichkeit.

- B. In der Praxis ist die Verteilungsfunktion F der Zufallsvariablen X meist unbekannt. Seien dazu X_1, \dots, X_n unabhängige Zufallsvariablen mit Verteilungsfunktion F . Schätze F_n durch

$$\hat{F}_n(x) := \hat{F}_n(x, \omega) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[X_i(\omega) \leq x]}, \quad x \in \mathbb{R},$$

wobei $\hat{F}_n(x)$ die relative Anzahl derjenigen X_1, \dots, X_n bezeichnet mit $X_i \leq x$. \hat{F}_n heißt **empirische Verteilungsfunktion** zu X_1, \dots, X_n . Nach dem starken Gesetz der großen Zahlen von Kolmogorov gilt P -f.s.

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{F}_n(x) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbf{1}_{[X_i \leq x]}}_{=: Y_i(x)} = \mathbf{E}Y_1(x) = \mathbf{E}\mathbf{1}_{[X_1 \leq x]} = P[X_1 \leq x] \\ &= F(x), \end{aligned}$$

also $\hat{F}_n(x) \rightarrow F(x)$ P -f.s. und L^2 . Eine Verschärfung dieser Aussage liefert der

Satz von Glivenko-Cantelli Seien X_n, X unabhängige und identisch verteilte Zufallsvariablen. Die empirische Verteilungsfunktion \hat{F}_n konvergiert P -f.s. gleichmäßig gegen F , d.h.

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} \forall x \in \mathbb{R}, n \geq n_0 : \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| < \varepsilon \text{ f.s.} \quad \times$$

Dieser Satz wird in der mathematischen Statistik bewiesen und heißt auch Hauptsatz der Statistik.

Nach dem zentralen Grenzwertsatz gilt

$$\begin{aligned}\sqrt{n}(\hat{F}_n(x) - F(x)) &= \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[X_i \leq x]} - \mathbf{E} \mathbf{1}_{[X_1 \leq x]} \right) \\ &= \frac{\sum_{i=1}^n (\mathbf{1}_{[X_i \leq x]} - \mathbf{E} \mathbf{1}_{[X_1 \leq x]})}{\sqrt{n}} \stackrel{D}{\rightarrow} N(0, \sigma^2(x))\end{aligned}$$

wobei $\sigma^2(x) = \mathbf{V}(\mathbf{1}_{[X_1 \leq x]}) = F(x)(1 - F(x))$ und $\mathbf{1}_{[X_1 \leq x]}$ eine $b(1, F(x))$ verteilte Zufallsvariable. Somit ist $\hat{F}_n(x) - F(x)$ approximativ $N\left(0, \frac{F(x)(1-F(x))}{n}\right)$ -verteilt.

Da $\sigma(x) \leq 1/4$ für alle $x \in \mathbb{R}$, gilt für ein vorgegebenes $\varepsilon > 0$ und eine $N(0, 1)$ -verteilte Zufallsvariable Z ,

$$\begin{aligned}P[|\hat{F}_n(x) - F(x)| \leq \varepsilon] &\approx P\left[|Z| \leq \varepsilon \sqrt{\frac{n}{F(x)(1-F(x))}}\right] \\ &\geq P[|Z| \leq 2\sqrt{n}\varepsilon] \\ &= 2\Phi(2\varepsilon\sqrt{n}) - 1\end{aligned}$$

ZAHLENBEISPIEL $\varepsilon = 0.1$, $n = 100$, dann ist $P[|\hat{F}_n(x) - F(x)| \leq \varepsilon] \gtrsim 0,955$ für jedes x und jede Verteilungsfunktion F . ■

- c. **Satz von Berry-Esseen** Seien X_n unabhängige identisch verteilte Zufallsvariablen mit $\mathbf{E}X_1 = \mu$, $\mathbf{V}(X_1) = \sigma^2$ und $\mathbf{E}|X_1|^3 < \infty$. Dann gilt

$$\sup_{x \in \mathbb{R}} \left| P \left[\sqrt{n} \frac{\sum_{i=1}^n (X_i - \mu)}{n\sigma} \leq x \right] - \Phi(x) \right| \leq c \frac{\mathbf{E}|X_1|^3}{\sigma^3 \sqrt{n}},$$

mit $c < 0.7975$, wobei

$$P \left[\sqrt{n} \frac{\sum_{i=1}^n (X_i - \mu)}{n\sigma} \leq x \right] \rightarrow \Phi(x) = \int_{-\infty}^x \frac{1}{2\pi} e^{-\frac{t^2}{2}} dt.$$

nach dem zentralen Grenzwertsatz. ✕ ∞

9.12 **Zentraler Grenzwertsatz von Lindeberg** Die Folge $(X_n)_{n \in \mathbb{N}}$ quadratisch integrierbarer reeller Zufallsvariablen mit $\mathbf{E}X_1^2 > 0$, $\mathbf{E}X_n = 0$, sei unabhängig und erfülle — mit $s_n^2 := \sum_{i=1}^n \mathbf{E}X_i^2$, $s_n = \sqrt{s_n^2}$ — die **klassische Lindeberg-Bedingung**

$$\forall \varepsilon > 0 : \frac{1}{s_n^2} \sum_{i=1}^n \mathbf{E}(X_i^2 \mathbf{1}_{[|X_i| > \varepsilon s_n]}) \rightarrow 0. \quad (\text{LB})$$

Dann

$$\frac{1}{s_n} \sum_{i=1}^n X_i \xrightarrow{\mathcal{D}} N(0, 1)\text{-verteilte Zufallsvariable. } \times$$

Die Lindeberg-Bedingung stellt sicher, dass der Einfluss aller Zufallsvariablen ungefähr “gleich groß ist”. Salop kann man sagen, dass ein Zentraler Grenzwertsatz immer existiert, wenn man eine Größe betrachtet, die aus sehr vielen aber kleinen und nahezu unabhängigen Einflüssen besteht. In diesem Fall kann man stets vermuten, dass die Summe der Einflüsse normalverteilt ist.

Als Beispiel sei der Kurs einer Aktien genannt, die sich im Streubesitz befindet. Durch unabhängige Kauf- und Verkaufsaktionen haben die Aktienbesitzer nur einen geringen Einfluss auf den Kurs, in ihrer Gesamtheit führt dies aber zu normalverteilten Tages-Renditen. Befindet sich die Aktie dagegen im Besitz einiger weniger Großaktionäre, sind die Aktienkurse nicht mehr (log)normalverteilt.

9.4 **Bemerkungen.** A. In Satz 9.12 dient die Folge (s_n) zur Normierung. Die Lindeberg-Bedingung (LB) schränkt den Einfluss der einzelnen Zufallsvariablen ein.

B. In Satz 9.12 — mit am Erwartungswert zentrierten Zufallsvariablen — impliziert die klassische Lindeberg-Bedingung (LB) die **klassische Feller-Bedingung**

$$\max_{i=1, \dots, n} \left(\frac{1}{s_n^2} \mathbf{E}X_i^2 \right) \rightarrow 0$$

und wird ihrerseits durch die **klassische Ljapunov-Bedingung**

$$\exists \delta > 0 : \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbf{E}|X_i|^{2+\delta} \rightarrow 0$$

impliziert. Bei nicht am Erwartungswert zentrierten Zufallsvariablen ist jeweils X_i durch $X_i - \mathbf{E}X_i$, auch in der Definition von s_n , zu ersetzen.

C. Satz 9.12 \Rightarrow Satz 9.11. \rightarrow

Satz 9.12 folgt aus

- 9.13 **Zentraler Grenzwertsatz von Lindeberg für Dreiecksschemata von ZVn** Für jedes $n \in \mathbb{N}$ seien $X_{n,1}, \dots, X_{n,m_n}$ unabhängige quadratisch integrierbare reelle Zufallsvariablen mit $m_n \rightarrow \infty$ für $n \rightarrow \infty$. Ferner seien

$$\mathbf{E}X_{n,i} = 0, \quad \sum_{i=1}^{m_n} \mathbf{E}X_{n,i}^2 = 1$$

und die Lindeberg-Bedingung

$$\forall \varepsilon > 0: \sum_{i=1}^{m_n} \mathbf{E} \left(X_{n,i}^2 \mathbf{1}_{[|X_{n,i}| > \varepsilon]} \right) \rightarrow 0.$$

erfüllt. Dann

$$\sum_{i=1}^{m_n} X_{n,i} \xrightarrow{\mathcal{D}} N(0, 1)\text{-verteilte Zufallsvariable.} \quad \times$$

Bevor wir den Satz beweisen, betrachten wir folgendes Beispiel.

BSP 2 Seien X_n unabhängige, identisch verteilte, quadratisch integrierbare reelle Zufallsvariablen mit

$$\mathbf{E}X_1 = 0, \quad \mathbf{E}X_1^2 = 1, \quad S_n := \sum_{i=1}^n X_i.$$

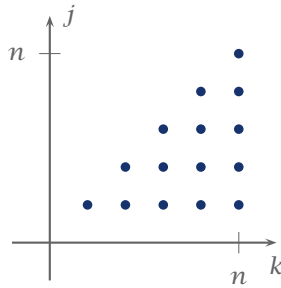
Wir zeigen

$$n^{-\frac{3}{2}} \sum_{k=1}^n S_k \rightarrow N\left(0, \frac{1}{3}\right)\text{-verteilte Zufallsvariable.}$$

Der zentrale Grenzwertsatz von Lindeberg 9.12 lässt sich so nicht anwenden, da beispielsweise die Unabhängigkeit in Bezug auf die S_k verletzt ist. Es gilt jedoch,

$$\sqrt{3}n^{-\frac{3}{2}} \sum_{k=1}^n S_k = \sqrt{3}n^{-\frac{3}{2}} \sum_{k=1}^n \sum_{j=1}^k X_j = \sqrt{3}n^{-\frac{3}{2}} \sum_{j=1}^n \underbrace{\sum_{k=j}^n X_j}_{(n-j+1)X_j}.$$

Die Vertauschbarkeit der Summen macht man sich schnell an der Skizze klar.



9.5 Zur Vertauschbarkeit der Summen.

Setzen wir nun $X_{n,j} = \sqrt{3}n^{-\frac{3}{2}}(n-j+1)X_j$, so sind die Voraussetzung von Satz 9.13 erfüllt, denn

$$\begin{aligned} \mathbf{E}X_{n,j} &= 0, \\ \sum_{j=1}^{m_n} \mathbf{E}X_{n,j}^2 &= \sum_{j=1}^n \mathbf{E}(\sqrt{3}n^{-\frac{3}{2}}(n-j+1)X_j)^2 = 3 \frac{1}{n} \sum_{j=1}^n \left(1 - \frac{j-1}{n}\right)^2 \\ &\rightarrow 3 \int_0^1 (1-t)^2 dt = 1. \end{aligned}$$

Außerdem ist (LB) erfüllt, da für $\varepsilon > 0$,

$$\begin{aligned} 3 \sum_{j=1}^n \mathbf{E}n^{-3}(n-j+1)^2 X_j^2 \mathbf{1}_{[\sqrt{3}n^{-\frac{3}{2}}(n-j+1)X_j > \varepsilon]} \\ \leq \frac{3}{n} \sum_{j=1}^n \underbrace{\mathbf{E}X_1^2 \mathbf{1}_{[\sqrt{3}n^{-\frac{1}{2}}|X_1| > \varepsilon]}}_{\rightarrow 0 \text{ in } \Omega} \rightarrow 0, \end{aligned}$$

denn $X_1^2 \mathbf{1}_{[\sqrt{3}n^{-\frac{1}{2}}|X_1| > \varepsilon]} \leq X_1^2$ und $\mathbf{E}X_1^2 < \infty$, also können wir den Satz von der dominierten Konvergenz anwenden. Somit konvergieren die Erwartungswerte gegen Null und nach dem Satz von Stolz-Cesàro konvergiert daher auch das arithmetische Mittel gegen Null.

Somit gilt

$$\sqrt{3}n^{-\frac{3}{2}} \sum_{k=1}^n S_k = \sum_{j=1}^n X_{n,j} \xrightarrow{D} N(0, 1) \quad \blacksquare$$

Zur Beweisvorbereitung benötigen wir noch einige Ergebnisse.

Lemma Seien $z_i, \eta_i \in \mathbb{C}$ mit $|z_i|, |\eta_i| < 1$, so gilt,

$$\left| \prod_{i=1}^n z_i - \prod_{i=1}^n \eta_i \right| \leq \sum_{i=1}^n |z_i - \eta_i|. \quad \times$$

» Der Beweis erfolgt durch Induktion. Der Induktionsanfang mit $n = 1$ ist klar. „ $n = 2$ “:

$$\begin{aligned} z_1 z_2 - \eta_1 \eta_2 &= z_1(z_2 - \eta_2) + \eta_2(z_1 - \eta_1), \\ \Rightarrow |z_1 z_2 - \eta_1 \eta_2| &\leq |z_2 - \eta_2| + |z_1 - \eta_1|. \end{aligned}$$

Der Induktionsschritt erfolgt analog. «

Aus der Analysis kennen wir die

Taylor-Formel mit Integralrestglied Sei $I \subset \mathbb{R}$ ein Intervall und $f : I \rightarrow \mathbb{C}$ mindestens $(n + 1)$ -mal stetig differenzierbar, so gilt für $a \in I$

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t) dt. \quad \times$$

Als unmittelbare Konsequenz erhalten wir

Hilfsformel 1 $\left| e^{ix} - \left(1 + ix - \frac{x^2}{2}\right) \right| \leq \min \{ |x^2|, |x^3| \}. \quad \times$

» Nach dem Taylor-Formel gilt für $a = 0$ und $m \in \mathbb{N}$,

$$e^{ix} = \sum_{k=0}^m \frac{1}{k!} (ix)^k + \frac{1}{m!} \int_0^x i^m (x-s)^m e^{is} ds.$$

$$m = 1 : \left| e^{ix} - (1 + ix) \right| = \left| \int_0^x i e^{is} (x-s) ds \right| \leq \frac{|x|^2}{2},$$

$$m = 2 : \left| e^{ix} - \left(1 + ix - \frac{x^2}{2}\right) \right| \leq \frac{|x|^3}{6} \leq |x|^3.$$

und folglich

$$\left| e^{ix} - \left(1 + ix - \frac{x^2}{2}\right) \right| \leq \left| e^{ix} - (1 + ix) \right| + \frac{|x|^2}{2} \leq |x|^2.$$

Zusammenfassend also

$$\left| e^{ix} - \left(1 + ix - \frac{x^2}{2}\right) \right| \leq \min \{ |x^2|, |x^3| \}. \quad \ll$$

Hilfsformel 2 Für $x \geq 0$ gilt $|e^{-x} - (1 - x)| \leq \frac{1}{2}x^2$. \times

$$\gg |e^{-x} - (1 - x)| \leq \int_0^x |e^{-s}(x - s)| ds \leq \int_0^x |x - s| ds = \frac{1}{2}x^2. \ll$$

» *Beweis von Satz 9.13* Wir zeigen zunächst, dass das Maximum der Varianzen von $X_{n,1}, \dots, X_{n,n_m}$ gegen Null konvergiert. Sei also $\varepsilon > 0$

$$\begin{aligned} \max_{i \in \{1, \dots, m_n\}} \mathbf{V}(X_{n,i}) &= \max_{i \in \{1, \dots, m_n\}} \int_{\Omega} X_{n,i}^2 \mathbf{1}_{[|X_{n,i}| \leq \varepsilon]} + X_{n,i}^2 \mathbf{1}_{[|X_{n,i}| > \varepsilon]} dP \\ &\leq \varepsilon^2 + \underbrace{\int_{\Omega} \sum_{i=1}^{m_n} X_{n,i}^2 \mathbf{1}_{[|X_{n,i}| > \varepsilon]} dP}_{-0 \text{ nach Vor.}} \end{aligned}$$

Somit konvergiert das Maximum gegen Null, da ε beliebig.

Seien nun $S_n = \sum_{i=1}^n X_i$ und $C_{n,i} := \mathbf{V}X_{n,i}$, so gilt nach Annahme $\sum_{i=1}^n C_{n,i} = 1$. Nach Satz 6.7 gilt für $t \in \mathbb{R}$,

$$\left| \varphi_{S_n}(t) - e^{-\frac{t^2}{2}} \right| = \left| \prod_{i=1}^n \varphi_{X_{n,i}}(t) - \prod_{i=1}^{m_n} e^{-C_{n,i} \frac{t^2}{2}} \right| \leq \sum_{i=1}^n \left| \varphi_{X_{n,i}}(t) - e^{-C_{n,i} \frac{t^2}{2}} \right|, \quad (*)$$

nach dem obigen Lemma. Anwendung von Hilfsformel 1 und $\mathbf{E}X_{n,i} = 0$ ergibt,

$$\begin{aligned} \left| \varphi_{n,i}(t) - \left(1 - C_{n,i} \frac{t^2}{2}\right) \right| &= \left| \int_{\Omega} e^{itX_{n,i}} - \left(1 - itX_{n,i} - X_{n,i}^2 \frac{t^2}{2}\right) dP \right| \\ &\stackrel{\text{HF1}}{\leq} \int_{\Omega} \min \left\{ |tX_{n,i}|^2, |tX_{n,i}|^3 \right\} dP \\ &\leq \underbrace{\max \{1, |t|^3\}}_{\alpha} \int_{\Omega} X_{n,i}^2 \min \{1, |X_{n,i}|\} dP, \end{aligned}$$

wobei

$$\begin{aligned} \int X_{n,i}^2 \min \{1, |X_{n,i}|\} dP &\leq \int_{[|X_{n,i}| > \varepsilon]} X_{n,i}^2 dP + \int_{[|X_{n,i}| \leq \varepsilon]} X_{n,i}^2 \underbrace{|X_{n,i}|}_{\leq \varepsilon} dP \\ &\leq \mathbf{E} \left(X_{n,i}^2 \mathbf{1}_{[|X_{n,i}| > \varepsilon]} \right) + \varepsilon C_{n,i}. \end{aligned}$$

Durch Summation erhalten wir,

$$\sum_{i=1}^{m_n} \left| \varphi_{n,i}(t) - \left(1 - C_{n,i} \frac{t^2}{2}\right) \right| \leq \alpha \left(\underbrace{\sum_{i=1}^{m_n} \mathbf{E} \left(X_{n,i}^2 \mathbf{1}_{[|X_{n,i}| > \varepsilon]} \right)}_{-0, \quad n \rightarrow \infty} + \varepsilon \underbrace{\sum_{i=1}^{m_n} C_{n,i}}_{=1} \right) \quad (1)$$

Also konvergiert der gesamte Ausdruck gegen Null für $n \rightarrow \infty$, da $\varepsilon > 0$ beliebig.

Da $C_{n,i}$ Varianz, ist $C_{n,i} \geq 0$. Wir können somit Hilfsformel 2 anwenden und erhalten,

$$\sum_{i=1}^{m_n} \left| e^{-C_{n,i} \frac{t^2}{2}} - \left(1 - C_{n,i} \frac{t^2}{2} \right) \right| \stackrel{\text{HF2}}{\leq} \frac{t^2}{4} \sum_{i=1}^{m_n} C_{n,i}^2 \leq \frac{t^2}{4} \underbrace{\max_{1 \leq i \leq m_n} C_{n,i}}_{\rightarrow 0, n \rightarrow \infty} \underbrace{\sum_{j=1}^{m_n} C_{n,i}}_{=1}. \quad (2)$$

Wenden wir die Dreiecksungleichung sowie (1) und (2) auf (*) an, so folgt

$$\forall t \in \mathbb{R} : \varphi_{S_n}(t) \rightarrow e^{-\frac{t^2}{2}}. \quad \ll$$

■ Multivariate zentrale Grenzwertsätze

Ist X ein d -dimensionaler integrierbarer Zufallsvektor, d.h. $\mathbf{E} \|X\| < \infty$, so heißt

$$\mathbf{E}X = (\mathbf{E}X_1, \dots, \mathbf{E}X_d)^\top$$

Erwartungsvektor von X .

Ist X ein d -dimensionaler quadratisch integrierbarer Zufallsvektor, d.h. $\mathbf{E} \|X\|^2 < \infty$, so heißt

$$\mathbf{Cov}(X) := (\mathbf{Cov}(X_i, X_j))_{i,j \in \{1, \dots, d\}}$$

Kovarianzmatrix von X , wobei die einzelnen Einträge die **Kovarianzen**

$$\mathbf{Cov}(X_i, X_j) := \mathbf{E}(X_i - \mathbf{E}X_i)(X_j - \mathbf{E}X_j)$$

der reellwertigen Zufallsvariablen X_i und X_j darstellen.

Auf der Hauptdiagonalen der Kovarianzmatrix stehen die Varianzen der X_i . Insbesondere ist im eindimensionalen Fall gerade $\mathbf{V}X = \mathbf{Cov}(X)$. Die Nebendiagonalelemente können als Maß für die stochastische Abhängigkeit der X_i, X_j interpretiert werden. Sind alle Komponenten unabhängig, so ist die Kovarianzmatrix eine Diagonalmatrix.

- 9.2 **Definition** Ein d -dimensionaler Zufallsvektor $X = (X_1, \dots, X_d)^\top$ heißt **multivariat normalverteilt** (oder auch **d -dimensional normalverteilt**), falls für jedes $u \in \mathbb{R}^d$ die Zufallsvariable $\langle u, X \rangle = u^\top X = \sum_{i=1}^d u_i X_i$ eindimensional normalverteilt ist, wobei eine 1-dimensionale Normalverteilung mit Varianz 0 als eine Dirac-Verteilung δ_a im Punkt a interpretiert wird. \times

9.3 **Definition** Seien X_n, X d -dimensionale Zufallsvektoren so konvergiert $X_n \rightarrow X$ nach Verteilung für $n \rightarrow \infty$, wenn

$$\forall f \in C_b(\mathbb{R}^n) : \mathbf{E}f(X_n) \rightarrow \mathbf{E}f(X).$$

Schreibe $X_n \xrightarrow{D} X$. \times

9.14 **Satz (Cramér-Wold-Device)** Für d -dimensionale Zufallsvektoren X_n und X gilt $X_n \xrightarrow{D} X$ genau dann, wenn $\langle u, X_n \rangle \xrightarrow{D} \langle u, X \rangle$ für alle $u \in \mathbb{R}^d$. \times

» Der Beweis wird in den Übungen behandelt. «

9.15 **Multivariater zentraler Grenzwertsatz** Seien $X_{n,1}, \dots, X_{n,m_n}$ unabhängige quadratisch integrierbare d -dimensionale Zufallsvariablen mit $m_n \rightarrow \infty$ für $n \rightarrow \infty$. Des Weiteren gelte

$$(a) \quad \forall n \in \mathbb{N} \quad \forall i \in \{1, \dots, m_n\} : \mathbf{E}X_{n,i} = 0,$$

$$(b) \quad \forall n \in \mathbb{N} : \sum_{i=1}^{m_n} \mathbf{Cov}(X_{n,i}) = C \in \mathbb{R}^{d \times d},$$

$$(c) \quad \forall \varepsilon > 0 : \sum_{i=1}^{m_n} \mathbf{E} \left(\|X_{n,i}\|_{\mathbf{1}_{\|X_{n,i}\| > \varepsilon}}^2 \right) \rightarrow 0.$$

Dann

$$\sum_{i=1}^{m_n} X_{n,i} \xrightarrow{D} N(0, C)\text{-verteilten Zufallsvektor. } \times$$

» $S_n := \sum_{i=1}^{m_n} X_{n,i}$. Wegen Satz 9.14 genügt es zu zeigen,

$$\forall u \in \mathbb{R}^d : \langle u, S_n \rangle \xrightarrow{D} N(0, \langle u, Cu \rangle).$$

1. Fall $\langle u, Cu \rangle = 0$. Dann ist

$$\mathbf{V}(\langle u, S_n \rangle) = 0,$$

also $\langle u, S_n \rangle = 0$ P-f.s.

2. Fall $\langle u, Cu \rangle > 0$. Wende Satz 9.13 auf

$$\tilde{X}_{n,i} = \frac{1}{\sqrt{\langle u, Cu \rangle}} \langle u, X_{n,i} \rangle$$

an, so erhalten wir für den Erwartungswert

$$\mathbf{E}\tilde{X}_{n,i} = \frac{1}{\sqrt{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle}} \langle \mathbf{u}, \mathbf{E}X_{n,i} \rangle = 0,$$

also

$$\sum_{i=1}^{m_n} \mathbf{V}(\tilde{X}_{n,i}) = \frac{1}{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle} \underbrace{\sum_{i=1}^{m_n} \mathbf{V}(\langle \mathbf{u}, X_{n,i} \rangle)}_{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle} = 1.$$

Zum Nachweis der Lindebergbedingung betrachte

$$\begin{aligned} \sum_{i=1}^{m_n} \mathbf{E} \left(\tilde{X}_{n,i}^2 \mathbf{1}_{[|\tilde{X}_{n,i}| > \varepsilon]} \right) &= \frac{1}{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle} \sum_{i=1}^{m_n} \mathbf{E} \underbrace{\langle \mathbf{u}, X_{n,i} \rangle^2}_{\leq \|\mathbf{u}\|^2 \|X_{n,i}\|^2} \mathbf{1}_{\left[\frac{1}{\sqrt{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle}} |\langle \mathbf{u}, X_{n,i} \rangle| > \varepsilon \right]} \\ &\leq \frac{1}{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle} \sum_{i=1}^{m_n} \mathbf{E} \left(\|\mathbf{u}\|^2 \|X_{n,i}\|^2 \right) \mathbf{1}_{\left[\frac{\|\mathbf{u}\|}{\sqrt{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle}} \|X_{n,i}\| > \varepsilon \right]} \\ &\leq \frac{\|\mathbf{u}\|^2}{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle} \sum_{i=1}^{m_n} \mathbf{E} \|X_{n,i}\|^2 \mathbf{1}_{\left[\|X_{n,i}\| > \frac{\sqrt{\langle \mathbf{u}, \mathbf{C}\mathbf{u} \rangle}}{\|\mathbf{u}\|} \varepsilon \right]} \rightarrow 0, \end{aligned}$$

nach Voraussetzung für $n \rightarrow \infty$. Damit sind alle Voraussetzungen von Satz 9.13 erfüllt. «

9.3 **Korollar** Ist $(X_n)_{n \in \mathbb{N}}$ eine unabhängige Folge identisch verteilter quadratisch integrierbarer Zufallsvektoren, so gilt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{m_n} (X_i - \mathbf{E}X_i) \xrightarrow{\mathcal{D}} N(0, \mathbf{Cov}(X_1))\text{-verteilte ZV.} \quad \times$$

10 Bedingte Erwartungen

Bedingte Erwartungen stellen den mathematischen Rahmen zur Untersuchung der Fragestellung, welchen Mittelwert eine Zufallsvariable Y annimmt unter der Voraussetzung, dass eine andere Zufallsvariable X den Wert x annimmt,

$$\mathbf{E}(Y \mid X = c).$$

Als Beispiel sei die Suche des mittleren Körpergewichts einer gegebenen Bevölkerungsschicht, unter Voraussetzung einer gewissen Körpergröße, genannt.

Erinnern wir uns zurück an die Definition des Erwartungswerts, so können wir diesen als Verallgemeinerung des Begriffes der Wahrscheinlichkeit für das Eintreten eines Ereignisses interpretieren. Allgemeiner gilt

$$P(A \mid X = x) = \mathbf{E}(\mathbf{1}_A \mid X = x).$$

Als Beispiel sei das Ereignis A , die Straße ist glatt, unter der Voraussetzung, dass die Lufttemperatur X den Wert x annimmt.

Unsere bisherige Definition von $P(A \mid X = x)$ schließt den Fall $P[X = x] = 0$ nicht mit ein. Wir werden dieses Manko durch die Definition einer allgemeinen "bedingten Erwartung" beheben.

10-A Grundlagen

Wie immer sei (Ω, \mathcal{A}, P) ein W-Raum.

Erinnern wir uns zunächst an die Definition der Messbarkeit einer Abbildung. f ist \mathcal{C} - \mathcal{B} -messbar $\Leftrightarrow f^{-1}(\mathcal{B}) \subseteq \mathcal{C}$.

Eine reellwertige Funktion ist eine Zufallsvariable, falls sie \mathcal{A} – \mathcal{B} messbar ist. Bedingte Erwartungen sind spezielle Zufallsvariablen, bei denen die Messbarkeitsforderung noch verschärft wird.

Zum Beweis des folgenden Satzes benötigen wir den Satz von Radon-Nikodym (siehe Anhang), der ein Verhältnis zwischen zwei Maßen μ und ν herstellt.

10.1 **Satz** Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ integrierbare Zufallsvariable und $\mathcal{C} \subset \mathcal{A}$ σ -Algebra. Dann existiert eine Zufallsvariable

$$Z : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$$

mit folgenden Eigenschaften:

(a) Z ist integrierbar und \mathcal{C} - \mathcal{B} -messbar,

(b) $\forall C \in \mathcal{C} : \int_C X \, dP = \int_C Z \, dP$. \times

» Ohne Einschränkung sei $X \geq 0$, ansonsten zerlegen wir X in X_+ und X_- . Sei $\varphi : \mathcal{C} \rightarrow \mathbb{R}$ definiert durch

$$\varphi(C) := \int_C X \, dP, \quad C \in \mathcal{C},$$

so ist φ aufgrund der Integrierbarkeit von X wohldefiniert und ein Maß. Außerdem ist φ ein $P|_{\mathcal{C}}$ -stetiges endliches Maß auf \mathcal{C} .

Somit sind alle Voraussetzungen des Satzes von Radon-Nikodym erfüllt und es folgt die Existenz einer bis auf die Äquivalenz “= $P|_{\mathcal{C}}$ -f.ü.” eindeutige, \mathcal{C} – \mathcal{B} -messbare Funktion

$$Z : \Omega \rightarrow \mathbb{R}_+.$$

Weiterhin ist $Z P|_{\mathcal{C}}$ -integrierbar, wobei

$$\forall C \in \mathcal{C} : \varphi(C) = \int_C Z \, dP|_{\mathcal{C}} = \int_C Z \, dP. \quad \ll$$

Z ist auch \mathcal{A} – \mathcal{B} -messbar und bezüglich \mathcal{P} integrierbar. \ll

Die Aussage des Satzes mag zunächst unscheinbar sein. Es ist jedoch zu beachten, dass die Zufallsvariable Z tatsächlich so konstruiert werden kann, dass sie \mathcal{C} - \mathcal{B} -messbar ist. Da man \mathcal{C} als echte Teilmenge von \mathcal{A} wählen kann, ist dies überhaupt nicht offensichtlich

Z ist eindeutig bis auf die Äquivalenz " $P|_{\mathcal{C}}$ -f.ü."

10.1 Definition Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \overline{\mathcal{B}})$ integrierbare Zufallsvariable und $\mathcal{C} \subset \mathcal{A}$ σ -Algebra. Die Äquivalenzklasse (im oberen Sinne) der Zufallsvariablen $Z : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ mit (a) und (b) — oder auch ein Repräsentant dieser Äquivalenzklasse — heißt *bedingte Erwartung von X bei gegebenem \mathcal{C}* . Man bezeichnet Z mit $\mathbf{E}(X | \mathcal{C})$.

Häufig wird ein Repräsentant dieser Äquivalenzklasse als eine Version von $\mathbf{E}(X | \mathcal{C})$ bezeichnet. \times

Die bedingte Erwartung $\mathbf{E}(X | \mathcal{C})$ stellt im Allgemeinen also *keine* reelle Zahl sondern eine Zufallsvariable dar! $\mathbf{E}(X | \mathcal{C})$ kann man als eine "Vergrößerung" von X betrachten, da \mathcal{C} gröber als \mathcal{A} und Z daher weniger Möglichkeiten der Variation als X besitzt. (Siehe Abbildung 11.2)

Bemerkungen. A. Zur Eindeutigkeit von Satz 10.1 betrachte zwei Zufallsvariablen Z_1 und Z_2 , so gilt

$$\forall C \in \mathcal{C} : \int_C (Z_1 - Z_2) dP = 0 \Rightarrow Z_1 = Z_2 \text{ } P|_{\mathcal{C}}\text{-f.ü.}$$

B. $\mathbf{E}(X | \mathcal{C}) = \frac{d\varphi}{dP|_{\mathcal{C}}}$, wobei $\varphi(C) = \int_C X dP$ für $C \in \mathcal{C}$. \rightarrow

BSP 1 a.) Sei $\mathcal{C} = \mathcal{A}$, so gilt $\mathbf{E}(X | \mathcal{C}) = X$ f.s.

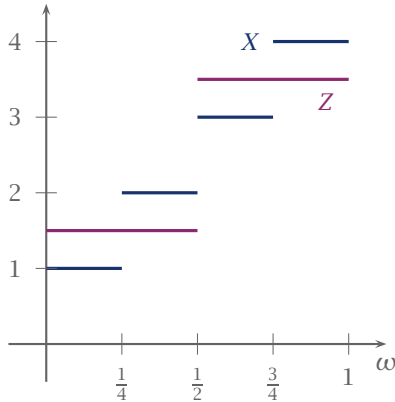
b.) Sei $\mathcal{C} = \{\emptyset, \Omega\}$ so gilt $\mathbf{E}(X | \mathcal{C}) = \mathbf{E}X$, denn

$$\underbrace{\int_{\emptyset} X dP}_{=0} = \underbrace{\int_{\emptyset} \mathbf{E}X dP}_{=0}, \quad \int_{\Omega} X dP = \int_{\Omega} \mathbf{E}X dP.$$

c.) Sei $\mathcal{C} = \{\emptyset, B, B^c, \Omega\}$ für festes B mit $0 < P(B) < 1$. So gilt

$$(\mathbf{E}(X | \mathcal{C}))(\omega) = \begin{cases} \frac{1}{P(B)} \int_B X dP =: \mathbf{E}(X | B), & \omega \in B \\ \frac{1}{P(B^c)} \int_{B^c} X dP, & \omega \in B^c. \end{cases}$$

$\mathbf{E}(X | B)$ heißt *bedingter Erwartungswert von X unter der Hypothese B* .



$$10.1 \quad \Omega = [0, 1], \mathcal{A} = \sigma\left(\left\{[0, \frac{1}{4}), [\frac{1}{4}, \frac{1}{2}), [\frac{1}{2}, \frac{3}{4}), [\frac{3}{4}, 1]\right\}\right), \mathcal{C} = \sigma\left(\left\{[0, \frac{1}{2}), [\frac{1}{2}, 1]\right\}\right) \subseteq \mathcal{A}.$$

» Scharfes Hinsehen liefert, dass die rechte Seite nach obiger Definition \mathcal{C} - \mathcal{B} -messbar ist. Weiterhin gilt

$$\begin{aligned} \int_{\emptyset} \text{"rechte Seite"} dP &= 0 = \int_{\emptyset} X dP, \\ \int_B \text{"rechte Seite"} dP &= \frac{\int_B 1 dP}{P(B)} \int_B X dP = \int_B X dP, \\ \int_{B^c} \text{"rechte Seite"} dP &= \frac{\int_{B^c} 1 dP}{P(B^c)} \int_{B^c} X dP = \int_{B^c} X dP, \\ \int_{\Omega} \text{"rechte Seite"} dP &= \int_B \text{"rechte Seite"} dP + \int_{B^c} \text{"rechte Seite"} dP \\ &= \int_B X dP + \int_{B^c} X dP = \int_{\Omega} X dP. \quad \ll \blacksquare \end{aligned}$$

10.2 **Satz** Seien X, X_i integrierbar, $\mathcal{C} \subset \mathcal{A}$ σ -Algebra und $c, \alpha_{1,2} \in \mathbb{R}$.

- $\forall C \in \mathcal{C} : \int_C \mathbf{E}(X | \mathcal{C}) dP = \int_C X dP.$
- $X = c$ P-f.s. $\Rightarrow \mathbf{E}(X | \mathcal{C}) = c$ f.s.
- $X \geq 0$ P-f.s. $\Rightarrow \mathbf{E}(X | \mathcal{C}) \geq 0$ f.s.
- $\mathbf{E}(\alpha_1 X_1 + \alpha_2 X_2 | \mathcal{C}) = \alpha_1 \mathbf{E}(X_1 | \mathcal{C}) + \alpha_2 \mathbf{E}(X_2 | \mathcal{C})$ f.s.
- $X_1 \leq X_2$ P-f.s. $\Rightarrow \mathbf{E}(X_1 | \mathcal{C}) \leq \mathbf{E}(X_2 | \mathcal{C})$ f.s.

f) X \mathcal{C} - \mathcal{B} -messbar $\Rightarrow X = \mathbf{E}(X \mid \mathcal{C})$ f.s.

g) X integrierbar, Y \mathcal{C} - \mathcal{B} -messbar, XY integrierbar

$$\Rightarrow \mathbf{E}(XY \mid \mathcal{C}) = Y\mathbf{E}(X \mid \mathcal{C}) \text{ f.s.}$$

g') X, X' integrierbar, $X\mathbf{E}(X' \mid \mathcal{C})$ integrierbar

$$\Rightarrow \mathbf{E}(X\mathbf{E}(X' \mid \mathcal{C}) \mid \mathcal{C}) = \mathbf{E}(X \mid \mathcal{C})\mathbf{E}(X' \mid \mathcal{C}) \text{ f.s.}$$

h) σ -Algebra $\mathcal{C}_{1,2}$ mit $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{A}$, X integrierbar

$$\mathbf{E}(\mathbf{E}(X \mid \mathcal{C}_1) \mid \mathcal{C}_2) = \mathbf{E}(X \mid \mathcal{C}_1) \text{ f.s.}$$

$$\mathbf{E}(\mathbf{E}(X \mid \mathcal{C}_2) \mid \mathcal{C}_1) = \mathbf{E}(X \mid \mathcal{C}_1) \text{ f.s.}$$

Hier bedeutet f.s., Rest \mathcal{C}_2 P-f.s. bzw. Rest \mathcal{C}_1 P-f.s. \times

- » a) Folgt sofort aus der Definition der bedingten Erwartung.
- b) Klar, denn Glättung einer Konstanten ergibt die Konstante selbst.
- c) Für eine $\mathcal{C} - \mathcal{B}$ -messbare Funktion Z mit $\forall C \in \mathcal{C} : \int_C Z dP \geq 0$ folgt $Z \geq 0$ $P \mid_{\mathcal{C}}$ -f.s.
- d) Folgt direkt aus der Linearität des Intergrals.
- e) Folgt direkt aus der Monotonie des Integrals, denn für $\mathcal{C} - \mathcal{B}$ -messbare Zufallsvariablen $Z_{1,2}$ mit

$$\forall C \in \mathcal{C} : \int_C Z_1 dP \leq \int_C Z_2 dP$$

gilt auch $Z_1 \leq Z_2$ $P \mid_{\mathcal{C}}$ -f.s.

f) Klar.

g) Es folgt sofort, dass XY und $Y\mathbf{E}(X \mid \mathcal{C})$ $\mathcal{C} - \mathcal{B}$ -messbar. Es verbleibt zu zeigen, dass

$$\forall C \in \mathcal{C} : \int_C Y\mathbf{E}(X \mid \mathcal{C}) dP = \int_C XY dP.$$

Ohne Einschränkung ist $X \geq 0$, ansonsten gehen wir zu X_+ und X_- über. Sei $C \in \mathcal{C}$ so gilt

$$\int_C Y \mathbf{E}(X | \mathcal{C}) dP = \int_C Y \mathbf{E}(X | \mathcal{C}) dP |_{\mathcal{C}},$$

wobei $\mathbf{E}(X | \mathcal{C}) = \frac{d\varphi}{dP} |_{\mathcal{C}}$ mit $\varphi(C) = \int_C X dP$ für $C \in \mathcal{C}$, so dass nach dem Zusatz zum Satz von Radon-Nikodym,

$$\int_C Y \mathbf{E}(X | \mathcal{C}) dP |_{\mathcal{C}} = \int_C Y d\varphi = \int_C Y d\varphi^*,$$

mit $\varphi^*(A) = \int_A X dP$ für $A \in \mathcal{A}$ (also $\varphi = \varphi^* |_{\mathcal{C}}$). Somit ist φ^* ein endliches Maß auf \mathcal{A} . Erneute Anwendung des Zusatzes zum Satz von Radon-Nikodym ergibt,

$$\int_C Y d\varphi^* = \int_C Y \frac{d\varphi^*}{dP} dP.$$

Nun ist $\frac{d\varphi^*}{dP} = X$ P -f.s. und daher

$$\int_C Y \frac{d\varphi^*}{dP} dP = \int_C YX dP.$$

g') Folgt sofort aus g), wenn wir $Y = \mathbf{E}(X | \mathcal{C})$ setzen.

h) (α) Aus g) folgt direkt, dass

$$\mathbf{E}(\mathbf{E}(X | \mathcal{C}_1) | \mathcal{C}_2) = \mathbf{E}(X | \mathcal{C}_1) \underbrace{\mathbf{E}(1 | \mathcal{C}_2)}_1,$$

da $\mathbf{E}(X | \mathcal{C}_1)$ $\mathcal{C}_1 - \mathcal{B}$ -messbar.

(β) Die zweite Gleichung ist plausibel, da die Vergrößerung von X über \mathcal{C}_2 zur noch kleineren σ -Algebra \mathcal{C}_1 dasselbe liefert, wie die unmittelbare Vergrößerung über \mathcal{C}_1 .

Seien also $Z := \mathbf{E}(X | \mathcal{C}_1)$ und $Y := \mathbf{E}(X | \mathcal{C}_2)$. Es ist zu zeigen, dass $Z = \mathbf{E}(Y | \mathcal{C}_1)$. Zunächst ist nach Definition Z auch $\mathcal{C}_1 - \mathcal{B}$ -messbar. Weiterhin sei $C \in \mathcal{C}_1 \subseteq \mathcal{C}_2$, so gilt per definitionem,

$$\begin{aligned} \int_C Z dP &= \int_C X dP, & \int_C Y dP &= \int_C X dP, \\ \Rightarrow \int_C Y dP &= \int_C Z dP. \quad \ll \end{aligned}$$

10.2 **Definition** Sei $\mathcal{C} \subset \mathcal{A}$ σ -Algebra und $A \in \mathcal{A}$.

$$P(A | \mathcal{C}) := \mathbf{E}(\mathbf{1}_A | \mathcal{C})$$

heißt *bedingte Wahrscheinlichkeit von A bei gegebenem \mathcal{C}* . \times

Setzen wir $\mathcal{C} = \{\emptyset, \Omega\}$, so ist $\mathbf{E}(\mathbf{1}_A | \mathcal{C}) = \mathbf{E}\mathbf{1}_A = P(A)$ (vgl. BSP 1). Nach Definition 10.2 gilt dann

$$\mathbf{E}(\mathbf{1}_A | \mathcal{C}) = P(A | \mathcal{C}).$$

10.1 **Bemerkung zu Definition 10.2.**

$$\forall C \in \mathcal{C} : \int_C P(A | \mathcal{C}) dP = P(A \cap C). \quad \rightarrow$$

» Sei $C \in \mathcal{C}$, so gilt

$$\int_C P(A | \mathcal{C}) dP = \int_C \mathbf{E}(\mathbf{1}_A | \mathcal{C}) dP = \int_C \mathbf{1}_A dP = \int_{\Omega} \mathbf{1}_A \mathbf{1}_C dP = P(A \cap C). \quad \leftarrow$$

BSP 2 Sei $\mathcal{C} = \{\emptyset, B, B^c, \Omega\}$ mit $0 < P(B) < 1$. Sei $A \in \mathcal{A}$, so gilt

$$(P(A | \mathcal{C}))(\omega) = \begin{cases} \frac{P(A \cap B)}{P(B)} =: P(A | B), & \omega \in B \\ \frac{P(A \cap B^c)}{P(B^c)} =: P(A | B^c), & \omega \in B^c, \end{cases}$$

denn

$$\begin{aligned} (P(A | \mathcal{C}))(\omega) &\stackrel{\text{vgl. BSP 1}}{=} \mathbf{E}(\mathbf{1}_A | \mathcal{C})(\omega) \\ &= \begin{cases} \frac{1}{P(B)} \int_B \mathbf{1}_A dP = \frac{P(A \cap B)}{P(B)} = P(A | B), & \omega \in B, \\ \frac{1}{P(B^c)} \int_{B^c} \mathbf{1}_A dP = \frac{P(A \cap B^c)}{P(B^c)} = P(A | B^c), & \omega \notin B. \quad \blacksquare \end{cases} \end{aligned}$$

10.3 **Definition** 1.) Seien $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$, $Y : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ Zufallsvariablen und X integrierbar.

$$\mathbf{E}(X | Y) := \mathbf{E}(X | Y^{-1}(\mathcal{A}'))$$

heißt *bedingte Erwartung von X bei gegebenem Y*. $Y^{-1}(\mathcal{A})$ bezeichnet hier die kleinste σ -Algebra in Ω , bzgl. der Y messbar ist.

2.) Seien $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$, $Y_i : (\Omega, \mathcal{A}, P) \rightarrow (\Omega'_i, \mathcal{A}'_i)$ für $i \in I$ Zufallsvariablen und X integrierbar.

$\mathcal{C} \subset \mathcal{A}$ sei die kleinste σ -Algebra in Ω , bzgl. der alle Y_i messbar sind, d.h. $\mathcal{C} = \mathcal{F} \left(\bigcup_{i \in I} Y_i^{-1}(\mathcal{A}_i) \right)$. So heißt

$$\mathbf{E}(X \mid (Y_i)_{i \in I}) := \mathbf{E}(X \mid \mathcal{C})$$

bedingte Erwartung von X bei gegebenen Y_i , $i \in I$.

3.) Sei $A \in \mathcal{A}$ und $Y : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ Zufallsvariable.

$$P(A \mid Y) := \mathbf{E}(\mathbf{1}_A \mid Y)$$

heißt *bedingte Wahrscheinlichkeit von A bei gegebenem Y* . \times

10.2 *Bemerkungen.* Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ eine integrierbare Zufallsvariable.

A. Sei \mathcal{C} σ -Algebra in \mathcal{A} . Dann gilt

$$(X^{-1}(\overline{\mathcal{B}}), \mathcal{C}) \text{ unabhängig} \Rightarrow \mathbf{E}(X \mid \mathcal{C}) = \mathbf{E}X \text{ f.s.}$$

B. Sei $Y : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ eine Zufallsvariable. Dann gilt

$$(X, Y) \text{ unabhängig} \Rightarrow \mathbf{E}(X \mid Y) = \mathbf{E}X \text{ f.s.} \quad \rightarrow$$

» A. Zu zeigen ist, dass

$$\forall C \in \mathcal{C} : \int_C \mathbf{E}X \, dP = \int_C X \, dP.$$

Schreiben wir

$$\int_C \mathbf{E}X \, dP = (\mathbf{E}X) \int_{\Omega} \mathbf{1}_C \, dP = (\mathbf{E}X)(P(C)),$$

sowie

$$\int_C X \, dP = \int_{\Omega} \mathbf{1}_C X \, dP = \mathbf{E}(\mathbf{1}_C X) = \mathbf{E}(\mathbf{1}_C) \mathbf{E}X,$$

mit Hilfe der Unabhängigkeit, so ist die Gleichheit offensichtlich.

B. Folgt direkt mit a). «

10.3 **Satz** Seien $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$, $Y : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ Zufallsvariablen. Dann existiert eine Abbildung

$$g : (\Omega', \mathcal{A}') \rightarrow (\mathbb{R}, \mathcal{B}),$$

mit $E(X | Y) = g \circ Y$. g ist die sog. **Faktorisierung der bedingten Erwartung**. g ist eindeutig bis auf die Äquivalenz “= P_Y -f.ü.”. \rightarrow

» Sei $Z := E(X | Y)$, dann

$$Z : (\Omega, \mathcal{F}(Y)) \rightarrow (\mathbb{R}, \mathcal{B}).$$

a) **Existenz 1. Schritt.** Rückführung auf Indikatorfunktionen. Sei $Z = \mathbf{1}_A$ für ein $A \in \mathcal{F}(Y) = Y^{-1}(\mathcal{A}')$, also existiert ein $A' \in \mathcal{A}'$, so dass $Y^{-1}(A') = A$.

$$\mathbf{1}_{A'}(Y(\omega)) = \begin{cases} 1, & Y(\omega) \in A', \text{ d.h. } \omega \in A, \\ 0, & \text{sonst} \end{cases} = \mathbf{1}_A(\omega).$$

Also ist $Z = g \circ Y$ mit $g = \mathbf{1}_{A'}$.

2. **Schritt.** Z sei nichtnegativ und einfach. Dann folgt die Existenz von g mit $Z = g \circ Y$ direkt aus dem 1. Schritt.

3. **Schritt.** Z sei nichtnegativ und $\mathcal{C} - \mathcal{B}$ -messbar. Dann existiert einer Folge von einfachen Zufallsvariablen Z_n mit $Z_n \uparrow Z$. Nach dem 2. Schritt existieren $g_n : (\Omega', \mathcal{A}') \rightarrow (\mathbb{R}, \mathcal{B}')$ mit $Z_n = g_n \circ Y$. Setzen wir

$$g^* := \sup_{n \in \mathbb{N}} g_n,$$

so ist g^* messbar und setzen wir weiterhin

$$g = g^* \mathbf{1}_{[g^* < \infty]},$$

so ist g reellwertig. Da $Z = \sup_{n \in \mathbb{N}} Z_n$, gilt auch P -f.s.

$$Z = g \circ Y.$$

4. **Schritt.** Sei Z messbar, so besitzt Z eine Darstellung $Z = Z_+ - Z_-$, mit Z_+, Z_- positiv und messbar. Nach dem 3. Schritt existieren Abbildungen g_+, g_- mit $Z_+ = g_+ \circ Y$ und $Z_- = g_- \circ Y$. $g := g_+ - g_-$ ist die gesuchte Abbildung.

b) *Eindeutigkeit.* Es gelte $\mathbf{E}(X | Y) = g_1 \circ Y$ f.s. = $g_2 \circ Y$ f.s. und somit

$$\begin{aligned} \forall C \in \mathcal{F}(Y) : \int_C g_1 \circ Y \, dP &= \int_C X \, dP = \int_C g_2 \circ Y \, dP. \\ \Rightarrow \forall C' \in \mathcal{A}' : \int_{Y^{-1}(C')} g_1 \circ Y \, dP &= \int_{Y^{-1}(C')} g_2 \circ Y \, dP. \end{aligned}$$

Nach dem Transformationssatz gilt somit

$$\forall C' \in \mathcal{A}' : \int_{C'} g_1 \, dP_Y = \int_{C'} g_2 \, dP_Y,$$

d.h. gerade

$$\forall C' \in \mathcal{A}' : \int_{C'} (g_1 - g_2) \, dP_Y = 0$$

und daher $g_1 = g_2$ P_Y -f.ü. «

10.4 **Definition** Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ bzw. $A \in \mathcal{A}$ eine integrierbare Zufallsvariable und $Y : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ Zufallsvariable. Sei g bzw. g_A eine — bis auf Äquivalenz “= P_Y -f.ü.” eindeutig bestimmte — Faktorisierung von $\mathbf{E}(X|Y)$ bzw. von $P(A | Y)$.

$$\mathbf{E}(X | Y = y) := g(y)$$

heißt *bedingte Erwartung von X unter der Hypothese $Y = y$.*

$$P(A | Y = y) := g_A(y)$$

heißt *bedingte Wahrscheinlichkeit von A unter der Hypothese $Y = y$*

$$\mathbf{E}(X | Y = \cdot) = g,$$

$$P(A | Y = \cdot) = g_A. \quad \times$$

10.4 **Satz** Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ bzw. $A \in \mathcal{A}$ integrierbare Zufallsvariable und $Y : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ Zufallsvariable. Dann gelten

$$\begin{aligned} a) \quad \forall A' \in \mathcal{A}' : \int_{A'} \mathbf{E}(X | Y = y) \, dP_Y(y) &= \int_{Y^{-1}(A')} X \, dP, \\ \text{insbesondere } \int_{\Omega'} \mathbf{E}(X | Y = y) \, dP_Y(y) &= \mathbf{E}X. \end{aligned}$$

$$\begin{aligned} b) \quad \forall A' \in \mathcal{A}' : \int_{A'} P(A | Y = y) \, dP_Y(y) &= P(Y^{-1}(A') \cap A), \\ \text{insbesondere } \int_{\Omega'} P(A | Y = y) \, dP_Y(y) &= P(A). \quad \times \end{aligned}$$

In vielen Anwendungen ist es tatsächlich einfacher, die bedingte Erwartung $\int_{\Omega'} \mathbf{E}(X | Y = y) dP_Y(y)$ anstelle des Erwartungswerts $\mathbf{E}Y$ zu berechnen.

- » a) Sei g eine Faktorisierung von $\mathbf{E}(X | Y)$ (deren Existenz sichert Satz 10.3) und $A' \in \mathcal{A}'$. So gilt

$$\begin{aligned} \int_{Y^{-1}(A')} X dP &= \int_{Y^{-1}(A')} \mathbf{E}(X | Y) dP \stackrel{\text{Satz 10.3}}{=} \int_{Y^{-1}(A')} g \circ Y dP \\ &= \int_{A'} g(y) dP_Y(y) \stackrel{\text{Def 10.4}}{=} \int_{A'} \mathbf{E}(X | Y = y) dP_Y(y). \end{aligned}$$

- b) Folgt direkt aus dem Vorangegangenen mit $X = \mathbf{1}_A$. «

BSP 3 Seien X bzw. A sowie Y wie zuvor. Sei $y \in \Omega'$ wobei $\{y\} \in \mathcal{A}'$ und $P[Y = y] = P_Y(\{y\}) > 0$.

$$\text{a.) } \underbrace{\mathbf{E}(X | Y = y)}_{\text{s. Definition 10.4}} = \underbrace{\mathbf{E}(X | [Y = y])}_{\text{s. BSP 1}}.$$

» Sei also $y \in \Omega'$ fest mit $\{y\} \in \mathcal{A}'$ und $P_Y(\{y\}) > 0$. Nach Satz 10.4 ist $\mathbf{E}(X | Y = y)$ eindeutig, denn für zwei Faktorisierungen g_1 und g_2 von $\mathbf{E}(X | Y)$ stimmen g_1 und g_2 bis auf eine Menge vom P_Y -Maß Null überein, aber $P_Y(\{y\}) > 0$, also $g_1(y) = g_2(y)$.

Sei also g eine Faktorisierung von $\mathbf{E}(X | Y)$, so ist $\mathbf{E}(X | Y) = g(y)$ unabhängig von der Wahl von g . Nach Satz 10.4 gilt

$$\begin{aligned} \int_{\{y\}} g(y') dP_Y(y') &= \int_{[Y=y]} X dP, \\ \Rightarrow g(y) &= \frac{1}{P[Y = y]} \int_{[Y=y]} X dP. \quad \ll \end{aligned}$$

$$\text{b.) } \underbrace{P(A | Y = y)}_{\text{s. Definition 10.4}} = \underbrace{P(A | [Y = y])}_{\text{s. BSP 2}}. \quad \blacksquare$$

10.5 **Satz** Sei $X : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ integrierbare Zufallsvariable und $Y : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ Zufallsvariable.

a) $X = c$ f.s. $\Rightarrow \mathbf{E}(X | Y = \cdot) = c$ P_Y -f.ü..

b) $X \geq 0$ f.s. $\Rightarrow \mathbf{E}(X | Y = \cdot) \geq 0$ P_Y -f.ü..

c) $\mathbf{E}(\alpha X_1 + \beta X_2 | Y = \cdot) = \alpha \mathbf{E}(X_1 | Y = \cdot) + \beta \mathbf{E}(X_2 | Y = \cdot)$ P_Y -f.ü..

$$d) X_1 \leq X_2 \text{ f.s.} \Rightarrow \mathbf{E}(X_1 | Y = \cdot) \leq \mathbf{E}(X_2 | Y = \cdot) \text{ } P_Y\text{-f.ü.} \quad \times$$

» Diese Eigenschaften ergeben sich sofort aus Satz 10.2 und Satz 10.3. «

10.6 **Satz** Seien $\mathcal{C} \subset \mathcal{A}$ eine Sub- σ -Algebra und X, X_n integrierbare Zufallsvariablen. Dann gilt:

a) Ist $0 \leq X_n \uparrow X$ f.s., so folgt

$$\mathbf{E}(X_n | \mathcal{C}) \rightarrow \mathbf{E}(X | \mathcal{C}) \text{ f.s.}$$

(Satz von der monotonen Konvergenz für bedingte Erwartungen).

b) Ist $X_n \rightarrow X$ f.s., $|X_n| \leq Y$ f.s. und Y eine integrierbare Zufallsvariable, so folgt

$$\mathbf{E}(X_n | \mathcal{C}) \rightarrow \mathbf{E}(X | \mathcal{C}) \text{ f.s.}$$

(Satz von der dominierten Konvergenz für bedingte Erwartungen). \times

» a) Sei $0 \leq X_n \uparrow X$ f.s., so folgt mit dem klassischen Satz von der monotonen Konvergenz, dass

$$\mathbf{E}|X_n - X| = \mathbf{E}X_n - \mathbf{E}X \rightarrow 0.$$

Außerdem folgt aus der Dreiecksungleichung oder aus Satz 10.7,

$$\mathbf{E}(\mathbf{E}(|X_n - X| | \mathcal{C})) \geq \mathbf{E}(|\mathbf{E}(X_n - X) | \mathcal{C}|) = \mathbf{E}(|\mathbf{E}(X_n | \mathcal{C}) - \mathbf{E}(X | \mathcal{C})|),$$

wobei die linke Seite gegen Null konvergiert, also $\mathbf{E}(X_n | \mathcal{C}) \xrightarrow{L^1} \mathbf{E}(X | \mathcal{C})$. Nun existiert eine Teilfolge, so dass $\mathbf{E}(X_{n_k} | \mathcal{C}) \rightarrow \mathbf{E}(X | \mathcal{C})$ f.s. und da $X_{n_k} \uparrow X$ f.s. ist auch die Konvergenz von $(\mathbf{E}(X_{n_k} | \mathcal{C}))$ monoton und da $(\mathbf{E}(X_n | \mathcal{C}))$ monoton folgt

$$\mathbf{E}(X_n | \mathcal{C}) \uparrow \mathbf{E}(X | \mathcal{C}).$$

b) Sei $X_n \rightarrow X$ f.s. Setzen wir $Z_n := \sup_{k \geq n} |X_k - X|$, so ist klar, dass $Z_n \downarrow 0$ P-f.s. und es gilt

$$|\mathbf{E}(X_n | \mathcal{C}) - \mathbf{E}(X | \mathcal{C})| \leq \mathbf{E}(|X_n - X| | \mathcal{C}) \leq \mathbf{E}(Z_n | \mathcal{C}),$$

aufgrund der Monotonie der bedingten Erwartung. Es genügt nun zu zeigen, dass $\mathbf{E}(Z_n | \mathcal{C}) \downarrow 0$ P -f.s.

Da $Z_n \downarrow 0$ P -f.s., konvergiert auch $\mathbf{E}(Z_n | \mathcal{C})$ P -f.s. punktweise gegen eine nicht-negative Zufallsvariable. Sei also $U := \lim_{n \rightarrow \infty} \mathbf{E}(Z_n | \mathcal{C})$. Nun gilt

$$0 \leq Z_n \leq \sup_{n \in \mathbb{N}} |X_n| + |X| \leq 2Y.$$

Wir können schreiben

$$\mathbf{E}(U) = \mathbf{E}(\mathbf{E}(U | \mathcal{C})) \leq \mathbf{E}(\mathbf{E}(Z_n | \mathcal{C})) = \mathbf{E}(Z_n) \rightarrow 0,$$

nach dem klassischen Satz von der dominierten Konvergenz, denn die Z_n konvergieren f.s. und werden durch eine integrierbare Zufallsvariable majorisiert. Somit ist $U = 0$ f.s. «

10.7 **Jensen'sche Ungleichung** Sei $\mathcal{C} \subset \mathcal{A}$ eine Sub- σ -Algebra, $I \subset \mathbb{R}$ ein Intervall, $f : I \rightarrow \mathbb{R}$ eine konvexe Funktion und $X : \Omega \rightarrow I$ eine integrierbare Zufallsvariable. Dann ist $\mathbf{E}(X | \mathcal{C}) \in I$ fast sicher. Ist $f(X)$ integrierbar, so gilt

$$f(\mathbf{E}(X | \mathcal{C})) \leq \mathbf{E}(f(X) | \mathcal{C}) \text{ f.s. } \times$$

» Falls $a \leq X$ bzw. $X \leq b$, so ist $a \leq \mathbf{E}(X | \mathcal{C})$ bzw. $\mathbf{E}(X | \mathcal{C}) \leq b$ und daher $\mathbf{E}(X | \mathcal{C}) \in I$.

Eine konvexe Funktion $f : I \rightarrow \mathbb{R}$ besitzt eine Darstellung als

$$f(x) = \sup_{\nu \in V} \nu(x), \quad V := \{\nu : I \rightarrow \mathbb{R} : \nu(t) = a + bt \leq f(t), \quad t \in I\}.$$

Somit gilt aufgrund der Linearität

$$f(\mathbf{E}(X | \mathcal{C})) = \sup_{\nu \in V} \nu(\mathbf{E}(X | \mathcal{C})) = \sup_{\nu \in V} (\mathbf{E}(\nu(X) | \mathcal{C})).$$

Nun ist $\sup_{\nu \in V} \nu(X) \leq f(X)$ und somit,

$$\sup_{\nu \in V} (\mathbf{E}(\nu(X) | \mathcal{C})) \leq \mathbf{E}(f(X) | \mathcal{C}). \quad \ll$$

11 Martingale

Ziel dieses Kapitels ist es, Kriterien für ein starkes Gesetz der großen Zahlen auch für abhängige Zufallsvariablen zu finden. Dazu untersuchen wir zunächst spezielle Folgen von Zufallsvariablen, die Martingale, für die sehr angenehme Konvergenzsätze existieren, auch wenn die Zufallsvariablen der Folge abhängig sind.

Martingale spielen eine große Rolle in der Spieltheorie, in der Finanzmathematik und in der stochastischen Analysis.

Für alles Weitere sei (Ω, \mathcal{A}, P) ein W-Raum, $n \in \mathbb{N}$ und sofern nicht anders angegeben, mit “ \rightarrow ” die Konvergenz für $n \rightarrow \infty$ bezeichnet.

11.1 **Definition** Eine Folge (X_n) integrierbarer Zufallsvariablen $X_n : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ heißt bei gegebener monoton wachsender Folge (\mathcal{A}_n) von σ -Algebren $\mathcal{A}_n \subset \mathcal{A}$ mit \mathcal{A}_n - $\overline{\mathcal{B}}$ -Messbarkeit von X_n

a) ein *Martingal* bzgl. (\mathcal{A}_n) , wenn

$$\mathbf{E}(X_{n+1} \mid \mathcal{A}_n) = X_n \text{ f.s.}, \quad n \geq 1,$$

b) ein *Submartingal* bzgl. (\mathcal{A}_n) , wenn

$$\mathbf{E}(X_{n+1} \mid \mathcal{A}_n) \geq X_n \text{ f.s.}, \quad n \geq 1,$$

c) ein *Supermartingal* bzgl. (\mathcal{A}_n) , wenn $(-X_n)$ ein Submartingal bzgl. (\mathcal{A}_n) ist. \times

Eine Folge von Zufallsvariablen ist nicht per se ein Martingal sondern immer nur in Bezug auf eine Folge von σ -Algebren. Ein wichtiger Spezialfall ist $\mathcal{A}_n = \mathcal{F}(X_1, \dots, X_n)$ und sofern nicht anders angegeben, gehen wir immer von dieser Wahl von \mathcal{A}_n aus.

11.1 **Bemerkung.** Ein Martingal (X_n) bezüglich (\mathcal{A}_n) ist stets auch ein Martingal bezüglich $(\mathcal{F}(X_1, \dots, X_n))$. Denn nach Voraussetzung ist jedes X_n \mathcal{A}_n - $\overline{\mathcal{B}}$ -messbar, so

dass $X_n^{-1}(\overline{\mathcal{B}}) = \mathcal{F}(X_n) \subset \mathcal{A}_n$. Aus der Monotonie von \mathcal{A}_n folgt dass $\mathcal{F}(X_1, \dots, X_n) \subset \mathcal{A}_n$ und somit

$$\begin{aligned} \mathbf{E}(X_{n+1} \mid X_1, \dots, X_n) &= \mathbf{E}(\underbrace{\mathbf{E}(X_{n+1} \mid \mathcal{A}_n)}_{\geq X_n} \mid \mathcal{F}(X_1, \dots, X_n)) \\ &\geq \mathbf{E}(X_n \mid \mathcal{F}(X_1, \dots, X_n)) = X_n. \end{aligned}$$

Entsprechend für Sub-, Supermartingal. \rightarrow

11.1 **Satz** Sei (V_n) eine Folge von Zufallsvariablen $V_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$. Die Partialsummenfolge $S_n = \sum_{j=1}^n V_j$ ist genau dann ein Martingal bzw. Submartingal bzgl. $(\mathcal{F}(V_1, V_1 + V_2, \dots, V_1 + \dots + V_n)) = (\mathcal{F}(V_1, \dots, V_n))$, wenn für jedes n gilt

a) V_n ist integrierbar, und

b) $\mathbf{E}(V_{n+1} \mid V_1, \dots, V_n) = 0$ bzw. ≥ 0 f.s. \times

» Die Integrierbarkeit sichert die Existenz der bedingten Erwartungen. Weiterhin gilt für jedes $n \geq 0$

$$\mathbf{E}\left(\sum_{j=1}^{n+1} V_j \mid \mathcal{F}_n\right) = \mathbf{E}(V_{n+1} \mid \mathcal{F}_n) + \sum_{j=1}^n V_j,$$

nach Satz 10.2, denn V_j ist \mathcal{F}_n -messbar für $j \leq n$. Somit ist die Partialsummenfolge genau dann ein Martingal, wenn $\mathbf{E}(V_{n+1} \mid \mathcal{F}_n) = 0$, bzw. ein Submartingal, wenn ≥ 0 . \ll

11.2 **Definition** Ein Spiel mit zufälligen Gewinnständen X_1, X_2, \dots nach dem 1., 2., ... Schritt heißt *fair*, wenn $\mathbf{E}X_1 = 0$ und (X_n) ein Martingal ist, d.h. für jedes n gilt $\mathbf{E}X_n = 0$ und

$$\mathbf{E}(X_{n+1} \mid X_1 = x_1, \dots, X_n = x_n) = x_n \text{ für } P_{(X_1, \dots, X_n)}\text{-f.a. } (x_1, \dots, x_n). \quad \times$$

11.2 **Satz** Seien die $V_n : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ quadratisch integrierbare Zufallsvariablen und die Partialsummenfolge $S_n = \sum_{j=1}^n V_j$ ein Martingal. Dann sind die V_n paarweise unkorreliert, d.h.

$$\mathbf{E}(V_i V_j) = 0, \quad i \neq j. \quad \times$$

» Da V_i und V_j quadratisch integrierbar sind, folgt mit der Cauchy-Schwartz-Ungleichung, dass

$$(\mathbf{E}(V_j V_i))^2 \leq (\mathbf{E}V_j^2)(\mathbf{E}V_i^2) < \infty.$$

Also ist $V_j V_i$ integrierbar. Falls $i < j$, so ist V_i \mathcal{F}_j -messbar und wir erhalten

$$\begin{aligned} \mathbf{E}(V_j V_i) &= \mathbf{E}(\mathbf{E}(V_j V_i \mid V_i)) \stackrel{10.2}{=} \mathbf{E}(V_i \mathbf{E}(V_j \mid V_i)) \\ &= \mathbf{E}(\underbrace{\mathbf{E}(V_j \mid V_1, \dots, V_{j-1})}_{=0 \text{ f.s. nach 11.1}} \mid V_i). \quad \ll \end{aligned}$$

BSP 1 Ein Beispiel für ein Martingal ist die Partialsummenfolge $(\sum_{i=1}^n V_j)_n$ zu einer unabhängigen Folge (V_n) von integrierbaren reellen Zufallsvariablen mit Erwartungswerten 0. ■

» Nach Satz 11.1 genügt es zu zeigen, dass

$$\mathbf{E}(V_{n+1} \mid V_1, \dots, V_n) = 0 \text{ f.s.}$$

da $(\mathcal{F}(V_{n+1}), \mathcal{F}(V_1, \dots, V_n))$ ein unabhängiges Paar von σ -Algebren ist, gilt nach Bemerkung 10.2,

$$\mathbf{E}(V_{n+1} \mid V_1, \dots, V_n) = \mathbf{E}V_{n+1} = 0. \quad \ll$$

11.3 **(Sub-/Super-)Martingalkonvergenztheorem von Doob** Sei (X_n) ein Sub-, Super- oder Martingal mit $\limsup_{n \rightarrow \infty} \mathbf{E}|X_n| < \infty$. Dann existiert eine integrierbare reelle Zufallsvariable X , so dass $X_n \rightarrow X$ P-f.s. ✕

Zur Beweisvorbereitung benötigen wir noch Definition 11.3 und 11.4, sowie Satz 11.4 und 11.5.

11.3 **Definition** Sei (\mathcal{A}_n) eine monoton wachsende Folge von σ -Algebren $\mathcal{A}_n \subset \mathcal{A}$. Eine Zufallsvariable

$$T : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{N}}, \mathcal{P}(\overline{\mathbb{N}})), \quad \overline{\mathbb{N}} := \mathbb{N} \cup \{\infty\}$$

heißt **Stopzeit** bzgl. (\mathcal{A}_n) , wenn

$$\forall k \in \mathbb{N} : [T = k] \in \mathcal{A}_k.$$

Hierbei heißt T **Stopzeit im engeren Sinne**, falls $P[T < \infty] = 1$ ("Kein Vorgriff auf die Zukunft").

Wichtiger Spezialfall: $\mathcal{A}_n = \mathcal{F}(X_1, \dots, X_n)$ mit Zufallsvariablen X_n . ✕

Man kann (X_n) als Folge der Gewinnstände in einem Spiel interpretieren. Ein Spieler ohne prophetische Gaben bricht das Spiel im zufälligen Zeitpunkt T aufgrund des bisherigen Spielverlaufs, d.h. aufgrund der Informationen, die bis zu diesem Zeitpunkt zur Verfügung stehen, ab.

BSP 2 $T(\omega) = \inf \{n \in \mathbb{N} : X_n(\omega) \in B\}$, $\omega \in \Omega$ — festes messbares B .

Zum Beispiel habe sich ein Aktienhändler einen festen Minimalwert vorgegeben, bei dessen Unterschreitung er seine Aktien verkaufen will. B stellt dann das Ereignis dar, dass eine Aktie den Minimalwert unterschreitet, und X_n den Aktienkurs seiner Aktien. Seine Bedingung lautet damit "Verkaufe die Aktien für das kleinste n , so dass $X_n \in B$ ". ■

11.4 **Definition** Sei (X_n) eine Folge von Zufallsvariablen $X_n : (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ und (T_n) eine Folge von Stoppzeiten bzgl. (X_n) [d.h. bzgl. $(\mathcal{F}(X_1, \dots, X_n))$] mit $T_1 \leq T_2 \leq \dots < \infty$.

So wird eine neue Folge $(X_{T_n})_n$ von Zufallsvariablen definiert durch

$$(X_{T_n})(\omega) := X_{T_n(\omega)}(\omega), \quad \omega \in \Omega.$$

Der Übergang von (X_n) zu (X_{T_n}) heißt *optional sampling* [frei gewählte Stichprobenbildung]. ✕

Man kann optional sampling z.B. als ein Testen des Spielverlaufs zu den Zeitpunkten $T_n(\omega)$ interpretieren.

Anschaulich greift man aus einer vorgegebenen Menge von Zufallsvariablen eine zufällige Teilmenge heraus. Dabei ist es durchaus möglich, dass eine Zufallsvariable mehrfach auftritt.

Vorsicht: Die (X_{T_n}) stellen keine Teilfolge von (X_n) dar!

11.4 **Optional Sampling Theorem** Sei (X_n) ein Submartingal, $M \in \mathbb{N}$ fest und (T_n) eine Folge von Stoppzeiten bzgl. (X_n) mit $T_1 \leq T_2 \leq \dots \leq M$.

Die durch optional sampling erhaltene Folge $(X_{T_n})_{n \in \mathbb{N}}$ ist ebenfalls ein Submartingal. — Entsprechend für Martingal statt Submartingal. ✕

Die Martingaleigenschaft ist invariant unter "Stoppen". Interpretieren wir (X_n) z.B. als Folge von Gewinnständen in einem Spiel, so besagt der Satz, dass sich die Fairness eines Spielverlaufs nicht ändert.

» Wir führen den Beweis für Submartingale. Martingale werden analog behandelt.

Für alles weitere sei $n \in \mathbb{N}$ fest und $C \in \mathcal{F}(X_{T_1}, \dots, X_{T_n})$. Setzen wir $T = T_{n+1}$ und $S = T_n$, so ist $S \leq T \leq M$ und wir haben zu zeigen, dass

$$\int_C X_T \, dP \geq \int_C X_S \, dP.$$

Da $C = \sum_{j=1}^M C \cap [S = j]$ genügt es obige Ungleichung auf $D_j := C \cap [S = j]$ zu zeigen. Es sei noch bemerkt, dass $D_j \in \mathcal{A}_j = \mathcal{F}(X_1, \dots, X_j)$.

Wir zeigen nun per Induktion, dass für jedes $m \geq j$

$$\int_{D_j \cap [T \geq m]} X_T \, dP \geq \int_{D_j \cap [T \geq m]} X_m \, dP.$$

Somit folgt da $j \leq T$ auf D_j ,

$$\int_{D_j} X_T \, dP = \int_{D_j \cap [T \geq j]} X_T \, dP \geq \int_{D_j} X_j \, dP = \int_{D_j} X_S \, dP.$$

Für den Induktionsanfang bemerken wir, dass $[T \geq M] = [T = M]$, also folgt die Behauptung unmittelbar. Für $m \geq j$ ist $D_j \cap [T \geq m+1] = D_j \cap [T \leq m]^c \in \mathcal{A}_m$, also folgt mit der Submartingaleigenschaft von X

$$\int_{D_j \cap [T \geq m+1]} X_{m+1} \, dP \geq \int_{D_j \cap [T \geq m+1]} X_m \, dP.$$

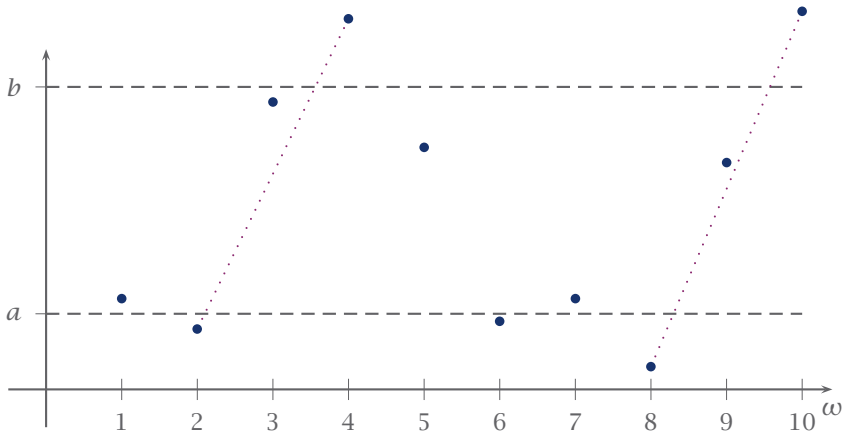
Zusammen mit der Induktionsvoraussetzung gilt also

$$\begin{aligned} \int_{D_j \cap [T \geq m]} X_T \, dP &= \int_{D_j \cap [T \geq m+1]} X_T \, dP + \int_{D_j \cap [T=m]} X_T \, dP \\ &\geq \int_{D_j \cap [T \geq m+1]} X_{m+1} \, dP + \int_{D_j \cap [T=m]} X_T \, dP \\ &\geq \int_{D_j \cap [T \geq m]} X_m \, dP. \end{aligned}$$

Damit ist die Induktion komplett und es gilt $\mathbf{E}(X_{T_{n+1}} \mid X_{T_1}, \dots, X_{T_n}) \geq X_{T_n}$, was zu zeigen war. «

11.5 Upcrossing inequality von Doob Sei (X_1, \dots, X_n) ein bei $n \geq 1$ abbrechendes Submartingal und $a < b$ reelle Zahlen. Die Zufallsvariable $U[a, b]$ gebe die Anzahl der aufsteigenden Überquerungen des Intervalls $[a, b]$ durch X_1, \dots, X_n an (d.h. die Anzahl der Übergänge der abbrechenden Folge von einem Wert $\leq a$ zu einem Wert $\geq b$). Dann gilt

$$(b - a)\mathbf{E}U[a, b] \leq \mathbf{E}(X_n - a)^+ - \mathbf{E}(X_1 - a)^+. \quad \times$$



11.1 Zur Upcrossing Inequality. $X_n(\omega)$ mit Überquerungen, $U_{10}[a, b] = 2$.

» Da (X_1, \dots, X_n) ein Submartingal ist, ist auch $(X_1 - a, \dots, X_n - a)$ ein Submartingal und ebenso $((X_1 - a)^+, \dots, (X_n - a)^+)$, aufgrund der Monotonie der bedingten Erwartung. Somit gibt $U[a, b]$ die Anzahl der aufsteigenden Überschreitungen des Intervalls $[0, b - a]$ durch $((X_1 - a)^+, \dots, (X_n - a)^+)$ an. Deshalb können wir ohne Einschränkung $a = 0$ und $X_i \geq 0$ annehmen.

Zu zeigen ist nun $bEU[0, b] \leq EX_n - EX_1$. Wir definieren uns Zufallsvariablen T_1, \dots, T_{n+1} durch folgende Vorschrift:

$$\begin{aligned} T_1(\omega) &= 1, \\ T_{2j}(\omega) &= \begin{cases} \min \{i : T_{2j-1} \leq i \leq n, X_i(\omega) = 0\}, & \text{falls existent,} \\ n, & \text{sonst,} \end{cases} \\ T_{2j+1}(\omega) &= \begin{cases} \min \{i : T_{2j}(\omega) \leq i \leq n, X_i(\omega) \geq b\}, & \text{falls existent,} \\ n, & \text{sonst,} \end{cases} \\ T_{n+1}(\omega) &= n. \end{aligned}$$

Die geraden Zeiten T_{2j} geben die Zeitpunkte der Unterschreitungen von $[0, b - a]$ an, und die ungeraden Zeiten T_{2j+1} geben die Zeitpunkte der Überschreitungen an. So definiert sind T_1, \dots, T_{n+1} sind Stoppzeiten mit $T_1 \leq \dots \leq T_{n+1} = n$.

Nach Satz 11.4 überträgt sich die Submartingaleigenschaft von X auf $(X_{T_1}, \dots, X_{T_n})$

und ein Teleskopsummenargument liefert

$$X_n - X_1 = \sum_{k=1}^n (X_{T_{k+1}} - X_{T_k}) = \underbrace{\sum (X_{T_{2j+1}} - X_{T_{2j}})}_{\geq b, \text{ da Aufsteigung}} + \underbrace{\sum (X_{T_{2j}} - X_{T_{2j-1}})}_{E... \geq 0, \text{ da Submartingal}}$$

$\geq bU[0, b]$

Also $\mathbf{E}X_n - \mathbf{E}X_1 \geq b\mathbf{E}U[0, b]$. «

Nun können wir einen Beweis für Satz 11.3 geben.

» *Beweis von 11.3.* Es gilt $\liminf_n X_n \leq \limsup_n X_n$ und beide Zufallsvariablen sind messbar. Angenommen $[\liminf_n X_n \neq \limsup_n X_n]$ habe positives Maß, so gilt aufgrund der Stetigkeit von P , dass für ein $\varepsilon > 0$

$$P[\limsup_n X_n - \liminf_n X_n > \varepsilon] > 0.$$

Also existieren reelle Zahlen $a < b$, so dass auch

$$P[\liminf_n X_n < a < b < \limsup_n X_n] > 0. \quad (\star)$$

Gebe nun $U_n[a, b]$ die Anzahl der aufsteigenden Überquerungen des Intervalls $[a, b]$ durch X_1, \dots, X_n an. Nach Voraussetzung ist $\sup_{n \geq 1} \mathbf{E}|X_n| \leq M < \infty$, also gilt nach Doobs Upcrossing Inequality 11.5

$$(b - a)\mathbf{E}U_n[a, b] \leq \mathbf{E}(X_n - a)^+ \leq M + |a| < \infty, \quad n \geq 1.$$

Nach (\star) überqueren die X_n auf einer Menge mit positivem Maß das Intervall $[a, b]$ jedoch unendlich oft, d.h. $\mathbf{E}U_n[a, b] \uparrow \infty$ im Widerspruch zur Upcrossing Inequality. Somit ist die Annahme falsch und es gilt

$$P[\liminf_n X_n = \limsup_n X_n] = 1.$$

Setzen wir also $X = \lim_{n \rightarrow \infty} X_n$ wann immer der Limes erklärt ist, so gilt $X_n \rightarrow X$ f.s. «

11.1 **Korollar** Ist (U_n) eine Folge integrierbarer nichtnegativ-reeller Zufallsvariablen auf (Ω, \mathcal{A}, P) und (\mathcal{A}_n) eine monoton wachsende Folge von Sub- σ -Algebren von \mathcal{A} mit \mathcal{A}_n - \mathcal{B}_+ -Messbarkeit von U_n und gilt weiterhin

$$\mathbf{E}(U_{n+1} \mid \mathcal{A}_n) \leq (1 + \alpha_n)U_n + \beta_n,$$

wobei $\alpha_n, \beta_n \in \mathbb{R}_+$ mit $\sum \alpha_n < \infty, \sum \beta_n < \infty$, dann konvergiert (U_n) f.s. — (U_n) ist "fast ein nichtnegatives Supermartingal".

Auch $(\mathbf{E}U_n)$ konvergiert. \times

Zusatz zum Martingalkonvergenztheorem Ist das Martingal (X_n) bezüglich der Folge (\mathcal{A}_n) gleichgradig integrierbar, d.h.

$$\sup_{n \geq 1} \mathbf{E}(|X_n| \cdot \mathbf{1}_{[|X_n| > c]}) \rightarrow 0, \quad c \rightarrow \infty,$$

so gilt zusätzlich

$$X_n \xrightarrow{L^1} X, \quad X_n = \mathbf{E}(X \mid \mathcal{A}_n). \quad \times$$

» Sei $\varepsilon > 0$ fest. Dann existiert ein $c > 0$, so dass

$$\sup_{n \geq 1} \mathbf{E}(|X_n| \cdot \mathbf{1}_{[|X_n| > c]}) < \varepsilon,$$

da (X_n) gleichgradig integrierbar. Nun gilt

$$\mathbf{E}|X_n| = \underbrace{\mathbf{E}(|X_n| \mid \mathbf{1}_{[|X_n| > c]})}_{< \varepsilon} + \underbrace{\mathbf{E}(|X_n| \mid \mathbf{1}_{[|X_n| \leq c]})}_{\leq c} < c + \varepsilon.$$

Also ist (X_n) L^1 -beschränkt. Mit Satz 11.3 folgt somit $\lim_{n \rightarrow \infty} X_n = X$ existiert f.s. und $X \in L^1$.

Wir zeigen nun, dass $X_n \rightarrow X$ in L^1 , d.h. $\mathbf{E}|X_n - X| \rightarrow 0$. Setze dazu

$$f_c(x) := \begin{cases} c, & x > c, \\ x, & -c \leq x \leq c, \\ -c, & x < -c, \end{cases}$$

so ist f lipschitz stetig und wegen der gleichgradigen Integrierbarkeit der X_n existiert ein $c > 0$, so dass

$$\mathbf{E}|f_c(X_n) - X_n| < \frac{\varepsilon}{3}, \quad n \geq 1 \quad (1)$$

$$\mathbf{E}|f_c(X) - X| < \frac{\varepsilon}{3}, \quad (2)$$

Da $\lim_{n \rightarrow \infty} X_n = X$ f.s. folgt

$$f_c(X_n) \rightarrow f_c(X) \text{ f.s.}$$

und $|f_c(X_n)| \leq c$. Mit dem Satz von der dominierten Konvergenz folgt

$$\mathbf{E}|f_c(X_n) - f_c(X)| \rightarrow 0 \text{ f.s.}$$

Zusammenfassend ergibt sich,

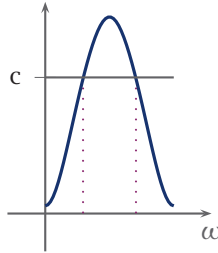
$$\mathbf{E} |X_n - X| \leq \mathbf{E} |X_n - f_c(X_n)| + \mathbf{E} |f_c(X_n) - f_c(X)| + \mathbf{E} |f_c(X) - X| < \varepsilon$$

also $X_n \rightarrow X$ in L^1 .

Es verbleibt zu zeigen, dass $X_n = \mathbf{E}(X | \mathcal{A}_n)$ f.s.. Sei $C \in \mathcal{A}_m$ und $n \geq m$. Da (X_n) ein Martingal ist, gilt $\mathbf{E}(X_n \mathbf{1}_C) = \mathbf{E}(X_m \mathbf{1}_C)$. Betrachte

$$|\mathbf{E}(X_n \mathbf{1}_C) - \mathbf{E}(X \mathbf{1}_C)| \leq \mathbf{E}(|X_n - X| \mathbf{1}_C) \leq \mathbf{E}(|X_n - X|) \rightarrow 0,$$

so folgt $\mathbf{E}(X_m \mathbf{1}_C) = \mathbf{E}(X \mathbf{1}_C)$. «



11.2 Zur gleichgradigen Integrierbarkeit.

11.6 **Satz** Sei (V_n) eine Folge von quadratisch integrierbaren reellen Zufallsvariablen mit $\sum_{n=1}^{\infty} \mathbf{V}(V_n) < \infty$. Dann ist

$$\sum_{n=1}^{\infty} (V_n - \mathbf{E}(V_n | V_1, \dots, V_{n-1})) \quad \text{f.s. konvergent.}$$

Sind die V_n zusätzlich unabhängig, dann ist $\sum_{n=1}^{\infty} (V_n - \mathbf{E}V_n)$ f.s. konvergent. ✕

» Ohne Einschränkung sei $\mathbf{E}V_n = 0$ ansonsten ersetze V_n durch $V_n - \mathbf{E}V_n$. Setze

$$W_n = V_n - \mathbf{E}(V_n | V_1, \dots, V_{n-1}),$$

so ist W_n integrierbar und $\mathcal{F}_n = \mathcal{F}(V_1, \dots, V_n)$ -messbar. Wir haben zu zeigen, dass $S_n = \sum_{j=1}^n W_j$ f.s. konvergiert. Zunächst ist

$$\begin{aligned} \mathbf{E}(W_{n+1} | V_1, \dots, V_n) \\ = \mathbf{E}(V_{n+1} | V_1, \dots, V_n) - \mathbf{E}(\mathbf{E}(V_{n+1} | V_1, \dots, V_n) | V_1, \dots, V_n) = 0, \end{aligned}$$

also ist S_n ein Martingal.

Im nächsten Schritt zeigen wir, dass $\sup_{n \geq 1} \mathbf{E}|S_n| < \infty$ ist. Betrachte dazu

$$S_n^2 = (W_n + S_{n-1})^2 = W_n^2 + 2W_n S_{n-1} + S_{n-1}^2.$$

Nach Satz 11.2 sind W_n und S_{n-1} unkorreliert, so dass

$$\mathbf{E}S_n^2 = \mathbf{E}W_n^2 + \mathbf{E}S_{n-1}^2 = \dots = \sum_{j=1}^n \mathbf{E}W_j^2.$$

Weiterhin gilt

$$\begin{aligned} \mathbf{E}W_n^2 &= \mathbf{E}(V_n - \mathbf{E}(V_n | V_1, \dots, V_{n-1}))^2 \\ &\leq 2\mathbf{E}(V_n^2 + \mathbf{E}(V_n | V_1, \dots, V_{n-1})^2) \\ &\leq 2\mathbf{E}(V_n^2 + \mathbf{E}(V_n^2 | V_1, \dots, V_{n-1})) \\ &= 4\mathbf{E}V_n^2 = 4\mathbf{V}V_n^2. \end{aligned}$$

Somit folgt $\mathbf{E}S_n^2 = \sum_{j=1}^n W_j^2 \leq 4 \sum_{j=1}^n \mathbf{V}V_j^2$. Also ist $(\mathbf{E}|S_n|)^2 \leq \mathbf{E}S_n^2$ beschränkt.

Mit dem Martingalkonvergenzatz 11.3 folgt nun die Behauptung, dass S_n f.s. konvergiert.

Sind die V_n außerdem unabhängig, so können wir Bemerkung 10.2 anwenden und erhalten $\mathbf{E}(V_n | V_1, \dots, V_{n-1}) = \mathbf{E}V_n$ f.s. «

11.7 **Satz** Sei (V_n) eine Folge von quadratisch integrierbaren reellen Zufallsvariablen mit $\sum_{n=1}^{\infty} n^{-2} \mathbf{V}(V_n) < \infty$. Dann gilt

$$\frac{1}{n} \sum_{j=1}^n (V_j - \mathbf{E}(V_j | V_1, \dots, V_{j-1})) \rightarrow 0 \quad \text{f.s.}$$

Falls zusätzlich (V_n) unabhängig, dann $\frac{1}{n} \sum_{j=1}^n (V_j - \mathbf{E}V_j) \rightarrow 0$ f.s.

(Kriterium von Kolmogorov zum starken Gesetz der großen Zahlen) ✕

» Wenden wir Satz 11.6 auf $n^{-1} V_n$ an, erhalten wir

$$\sum_{n=1}^{\infty} \frac{V_n - \mathbf{E}(V_n | V_1, \dots, V_{n-1})}{n} \text{ konvergiert f.s..}$$

Daraus folgt mit dem Lemma von Kronecker (s.u.),

$$\frac{1}{n} \sum_{i=1}^n (V_i - \mathbf{E}(V_i | V_1, \dots, V_{i-1})) \rightarrow 0 \text{ f.s..}$$

Falls die V_n unabhängig sind, können wir Bemerkung 10.2 anwenden und erhalten so den Zusatz. «

11.1 **Lemma von Kronecker** Sei (c_n) eine Folge reeller Zahlen.

$$\sum_{n=1}^{\infty} \frac{c_n}{n} \text{ konvergiert} \Rightarrow \frac{1}{n} \sum_{j=1}^n c_j \rightarrow 0. \quad \times$$

» Sei $s_n = \sum_{j=1}^n j^{-1} c_j$, so ist (s_n) nach Voraussetzung konvergent, sei s der Grenzwert. Offensichtlich ist $s_n - s_{n-1} = n^{-1} c_n$ mit $s_0 = 0$ und daher

$$\sum_{j=1}^n c_j = \sum_{j=1}^n j(s_j - s_{j-1}) = n s_n - \sum_{j=1}^n \underbrace{(j - (j-1))}_1 s_{j-1}.$$

Somit gilt

$$\frac{1}{n} \sum_{j=1}^n c_j = s_n - \underbrace{\frac{1}{n} \sum_{j=1}^n s_{j-1}}_{\rightarrow s} \rightarrow 0. \quad \ll$$

11.2 **Bemerkung.** Aus Satz 11.6 bzw. 11.7 ergibt sich unmittelbar für eine Folge (V_n) quadratisch integrierbarer reeller Zufallsvariablen eine hinreichende Bedingung für die f.s.-Konvergenz der Reihe $\sum_{j=1}^n V_n$ bzw. für $\frac{1}{n} \sum_{j=1}^n V_j \rightarrow 0$ f.s. \times

11.8 **Kriterium von Kolmogorov für das starke Gesetz der großen Zahlen** Eine unabhängige Folge (X_n) quadratisch integrierbarer reeller Zufallsvariablen mit

$$\sum_{n=1}^{\infty} n^{-2} V(X_n) < \infty$$

genügt dem starken Gesetz der großen Zahlen. \times

11.9 **Kolmogorovsches starkes Gesetz der großen Zahlen** Für eine unabhängige Folge $(X_n)_{n \in \mathbb{N}}$ identisch verteilter integrierbarer reeller Zufallsvariablen gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mathbf{E}X_1 \text{ f.s.} \quad \times$$

Der Satz wurde bereits im Kapitel 8 bewiesen, wir wollen nun unter Verwendung der Martingaltheorie einen eleganteren Beweis geben.

Stutzungslemma Sei (X_n) eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit $\mathbf{E}|X_1| < \infty$. Setze $Y_n := X_n \mathbf{1}_{[|X_n| \leq n]}$. Dann gilt

a) $\mathbf{E}Y_n \rightarrow \mathbf{E}X_1,$

b) $P[X_n = Y_n \text{ für fast alle } n] = 1,$

c) (Y_n) erfüllt die Kolmogorov-Bedingung,

$$\sum_{n=1}^{\infty} n^{-2} \mathbf{V}(Y_n) < \infty. \quad \times$$

» a) Sei $Z_n := X_1 \mathbf{1}_{[|X_1| \leq n]}$. Für jedes $n \in \mathbb{N}$ haben X_1 und X_n dieselbe Verteilung also besitzen auch Y_n und Z_n dieselbe Verteilung.

Somit gilt $\mathbf{E}Z_n = \mathbf{E}Y_n$. Ferner ist $|Z_n| \leq |X_n|$ und $Z_n \rightarrow X_1$ f.s. also nach dem Satz von der dominierten Konvergenz

$$\mathbf{E}Y_n = \mathbf{E}Z_n \rightarrow \mathbf{E}X_1 \text{ f.s.}$$

b)

$$\begin{aligned} \sum_{n=1}^{\infty} P[X_n \neq Y_n] &= \sum_{n=1}^{\infty} P[X_n > n] \leq \int_0^{\infty} P[|X_1| > t] dt \\ &\stackrel{\text{Lemma 4.1}}{=} \mathbf{E}|X_1| < \infty. \end{aligned}$$

Mit dem 1. Lemma von Borel und Cantelli folgt somit

$$P[X_n \neq Y_n \text{ für fast alle } n] = 0.$$

c) Für jedes $n \in \mathbb{N}$ gilt nach Partialbruchzerlegung,

$$\frac{1}{n^2} \leq \frac{2}{n(n+1)} \leq 2 \left(\frac{1}{n} - \frac{1}{n+1} \right).$$

Damit erhalten wir eine Abschätzung für den Reihenrest

$$\sum_{n \geq k} \frac{1}{n^2} \leq \frac{2}{k}. \quad (*)$$

Somit folgt

$$\sum_{n=1}^{\infty} \frac{|X_1|^2 \mathbf{1}_{[|X_1| \leq n]}}{n^2} \leq \sum_{n \geq \max\{1, |X_1|\}} \frac{|X_1|^2}{n^2} \leq \frac{2|X_1|^2}{\max\{1, |X_1|\}} \leq 2|X_1|.$$

Also gilt auch für die Summe der Varianzen

$$\sum_{n=1}^{\infty} \frac{\mathbf{V}(Y_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{\mathbf{E}|Y_n|^2}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbf{E}|Z_n|^2}{n^2} \leq 2\mathbf{E}|X_1| < \infty. \quad \ll$$

» *Beweis von Satz 11.9.* Wir betrachten die gestutzten Zufallsvariablen

$$Y_n := X_n \mathbf{1}_{[|X_n| \leq n]}, \quad S_n := \sum_{i=1}^n X_i.$$

Nach Betrachtung b) im obigen Lemma genügt es zu zeigen,

$$\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow \mathbf{E}X_1 \text{ f.s.}$$

Dazu betrachten wir die Zerlegung

$$\frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n \mathbf{E}Y_i + \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{E}Y_i).$$

Nach dem Grenzwertsatz von Caesaro gilt

$$\frac{1}{n} \sum_{i=1}^n \mathbf{E}Y_i \rightarrow \mathbf{E}X_1 \text{ f.s.}$$

und nach Satz 11.8 gilt außerdem

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \mathbf{E}Y_i) \rightarrow 0 \text{ f.s.} \quad \ll$$

11.2 **Korollar** Sei (X_n) eine Folge von unabhängigen identisch verteilten reell-erweiterten Zufallsvariablen mit $\mathbf{E}X_1^- < \infty$ und $\mathbf{E}X_1^+ = \infty$, d.h. $\mathbf{E}X_1 = \infty$. Dann gilt

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \infty \text{ f.s.} \quad \times$$

» Mit Satz 11.9 gilt

$$\frac{1}{n} \sum_{i=1}^n X_i^- \rightarrow \mathbf{E}X_1^- \text{ f.s.}$$

Deshalb können wir ohne Einschränkung davon ausgehen, dass $X_i \geq 0$. Sei $k \in \mathbb{N}$ beliebig aber fest und setze

$$Z_n^{(k)} := X_n \mathbf{1}_{[X_n \leq k]}.$$

So folgt $Z_n^{(k)} \leq k$ und

$$\frac{1}{n} \sum_{i=1}^n X_i \geq \frac{1}{n} \sum_{i=1}^n Z_i^{(k)} \rightarrow \mathbf{E}Z_1^{(k)} \text{ f.s.} \quad (**)$$

Aber $Z_1^{(k)} \uparrow X_1$ und mit dem Satz von der monotonen Konvergenz folgt

$$\mathbf{E}Z_1^{(k)} \uparrow \mathbf{E}X_1.$$

Mit (***) folgt jedoch, dass

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \infty \text{ f.s.} \quad \ll$$

A Begriffe und Sätze der Maß- und Integrationstheorie

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum.

Satz von der monotonen Konvergenz (B. Levi) Für erweitert reellwertige messbare Funktionen f_n mit $f_n \geq 0$ und $f_n \uparrow f$ existiert $\lim_n \int f_n \, d\mu$ und es gilt

$$\lim_n \int f_n \, d\mu = \int \lim_n f_n \, d\mu. \quad \times$$

Lemma von Fatou Für jede Folge (f_n) von erweitert reellwertigen messbaren Funktionen mit $f_n \geq 0$ μ -f.ü. gilt

$$\int \liminf f_n \, d\mu \leq \liminf \int f_n \, d\mu. \quad \times$$

Satz von der dominierten Konvergenz (Lebesgue) Für erweitert-reellwertige messbare Funktionen f_n, f und g mit $f_n \rightarrow f$ μ -f.ü., $|f_n| \leq g$ μ -f.ü. für alle n und $\int g \, d\mu < \infty$ existiert $\lim_{n \rightarrow \infty} \int f_n \, d\mu$ und es gilt

$$\lim_{n \rightarrow \infty} \int f_n \, d\mu = \int f \, d\mu. \quad \times$$

Definition 1.) Das Maß μ heißt σ -endlich, wenn es ein Folge von Mengen $A_n \in \mathcal{A}$ gibt mit $A_n \uparrow \Omega$ und $\mu(A_n) < \infty$.

2.) Ein Maß ν auf \mathcal{A} heißt μ -stetig, falls

$$\forall A \in \mathcal{A} : \mu(A) = 0 \Rightarrow \nu(A) = 0. \quad \times$$

Satz von Radon-Nikodym Seien (Ω, \mathcal{A}) Messraum, μ und ν σ -endliche Maße auf \mathcal{A} und ν sei μ -stetig. Dann existiert eine Funktion $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}_+, \mathbb{B}_+)$ mit

$$\forall A \in \mathcal{A} : \nu(A) = \int_A f \, d\mu$$

f ist eindeutig bis auf Äquivalenz “= μ -f.ü.”.

f ist die sog. **Radon-Nikodym-Ableitung** von ν nach μ und wird häufig kurz durch $\frac{d\nu}{d\mu}$ angegeben. \times

» Zum Beweis siehe Elstrodt. «

BSP 1 Sei Q ein W -Maß auf \mathcal{B} , λ das LB-Maß auf \mathcal{B} und Q sei λ -stetig. Dann ist die Radon-Nikodym-Ableitung

$$\frac{dQ}{d\lambda}$$

die gewöhnliche Dichte von Q . ■

Zusatz Seien μ, ν, φ σ -endliche Maße, φ sei ν -stetig und ν sei μ -stetig. Dann gilt

$$\frac{d\varphi}{d\mu} = \frac{d\varphi}{d\nu} \cdot \frac{d\nu}{d\mu}. \quad \times$$

Literaturverzeichnis

- [1] Bauer, H. Wahrscheinlichkeitstheorie; de Gruyter, Berlin (1990), 4. Aufl.
- [2] Bauer, H. Maß- und Integrationstheorie; de Gruyter, Berlin (1998), 2. Aufl.
- [3] Billingsley, P. Probability and Measure; Wiley, New York (1995), 3rd ed.
- [4] Capiński, M., Kopp, E. Measure Integral and Probability; Springer (2004), 2nd ed.
- [5] Dudley, R.M. Real Analysis and Probability; Cambridge University Press (2002), 2nd ed.
- [6] Elstrodt, J. Maß- und Integrationstheorie; Springer (2009), 6. Aufl.
- [7] Feller, W. An Introduction to Probability and Its Applications I,II; Wiley, New York (1970/71), 3rd ed./2nd ed.
- [8] Fisz, M. Wahrscheinlichkeitsrechnung und mathematische Statistik; VEB Deutscher Verlag der Wissenschaften, Berlin (1970), 5. Aufl.
- [9] Gänsler, P., Stute, W. Wahrscheinlichkeitstheorie; Springer, Berlin (1977).
- [10] Henze, N. Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls; Vieweg+Teubner (2008), 7. Aufl.
- [11] Hesse, C. Wahrscheinlichkeitstheorie — Eine Einführung mit Beispielen und Anwendungen. Vieweg+Teubner (2009), 2. Aufl.
- [12] Hinderer, K. Grundbegriffe der Wahrscheinlichkeitstheorie; Springer, Berlin (1985), 3. korrigierter Nachdruck.
- [13] Jacod, J., Protter, P. Probability Essentials. Springer, Berlin (2004), 2nd ed.
- [14] Kallenberg, O. Foundations of Modern Probability. Springer, New York (2001), 2nd ed.

- [15] Klenke, A. Wahrscheinlichkeitstheorie; Springer, Heidelberg (2006)
- [16] Krengel, U. Einführung in die Wahrscheinlichkeitstheorie und Statistik; Vieweg, Braunschweig (2003), 7. Aufl.
- [17] Loève, M. Probability Theory I,II; Springer, Berlin (1977/78), 4th ed.
- [18] Meintrup D., Schäffler S. Stochastik — Theorie und Anwendungen; Springer (2004).
- [19] Williams, D. Probability with Martingales; Cambridge University Press (1991).
- [20] Wengenroth, J. Wahrscheinlichkeitstheorie. De Gruyter (2008).
- [21] Williams, D. Weighing the Odds — A Course in Probability and Statistics; Cambridge University Press (2001).