Elementare Wahrscheinlichkeitstheorie und Mathematische Statistik

Prof. Dr. Karl-Heinz Fichtner

WS 2004/05

Vorwort

Dieses Dokument wurde als Skript für die auf der Titelseite genannte Vorlesung erstellt und wird jetzt im Rahmen des Projekts „Vorlesungsskripte der Fakultät für Mathematik und Informatik“ weiter betreut. Das Dokument wurde nach bestem Wissen und Gewissen angefertigt. Dennoch garantiert weder der auf der Titelseite genannte Dozent, die Personen, die an dem Dokument mitgewirkt haben, noch die Mitglieder des Projekts für dessen Fehlerfreiheit. Für etwaige Fehler und dessen Folgen wird von keiner der genannten Personen eine Haftung übernommen. Es steht jeder Person frei, dieses Dokument zu lesen, zu verändern oder auf anderen Medien verfügbar zu machen, solange ein Verweis auf die Internetadresse des Projekts `http://uni-skripte.lug-jena.de/` enthalten ist.

Diese Ausgabe trägt die Versionsnummer 2572 und ist vom 4. Dezember 2009. Eine neue Ausgabe könnte auf der Webseite des Projekts verfügbar sein.

Jeder ist dazu aufgerufen, Verbesserungen, Erweiterungen und Fehlerkorrekturen für das Skript einzureichen bzw. zu melden oder diese selbst einzupflegen – einfach eine E-Mail an die Mailingliste `<uni-skripte@lug-jena.de>` senden. Weitere Informationen sind unter der oben genannten Internetadresse verfügbar.

Hiermit möchten wir allen Personen, die an diesem Skript mitgewirkt haben, vielmals danken:

- *Jens Kubieziel `<jens@kubieziel.de>` (2004–2007)*
- *Kim Muckelbauer `<jimmi_usa2000@yahoo.com>` (2007)*
- *Carolin Rabsahl (2007)*
- *Mario Ullrich `<mario-ullrich@gmx.de>` (2007)*

Inhaltsverzeichnis

Auflistung der Theoreme

Sätze

Definitionen und Festlegungen

1. Maß- und allgemeine Integrationstheorie

Diese Vorlesung gliedert sich im wesentlichen in drei Teile. Dazu gehört dieses Kapitel und im späteren Verlauf dann die Wahrscheinlichkeitstheorie. Später wird dann auch auf die Stochastik eingegangen.

Die Maßtheorie stellte früher eine eigenständige 2+4-Vorlesung dar. Für Wirtschaftsmathematiker ist diese jedoch gekürzt. Hauptsächlich werden längere Beweise zu Sätzen nicht geführt. Diese können in der Lektüre nachvollzogen werden.

1.1. Einführung in die Maßtheorie

Die Maßtheorie stellt eine Verallgemeinerung einiger elementargeometrischer Begriffe dar und ermöglicht es, auch komplizierten Mengen ein Maß zuzuordnen. Dabei sollte man diese deutlich von der Messtheorie in der Physik unterscheiden. Wir betrachten hier zunächst allgemeine Maße und beschränken uns dann im späteren Verlauf auf Wahrscheinlichkeitsmaße.

1.1.1. Das Grundmodell der Wahrscheinlichkeitsrechnung

Die Wurzeln der Wahrscheinlichkeitsrechnung liegen zu Zeiten des 30-jährigen Krieges. In den Gefechtspausen veranstalteten die Offiziere oft Glücksspiele. Diese haben naturgemäß einen zufälligen Ausgang und einige Spieler machten sich Gedanken über den Verlauf und wie der Ausgang beeinflusst werden kann. Hiervon stammen einige der ersten praktischen Überlegungen zu dem Themengebiet.

Ein Breifwechsel zwischen Blaise Pascal und Pierre de Fermat aus dem Jahr 1654 wird gemeinhin als Geburtstunde der Wahrscheinlichkeitsrechnung angesehen. Die ersten Veröffentlichungen zur Wahrscheinlichkeitsrechnung enthielten noch keine Formeln. Stattdessen wurden die Rechenschritte in langen Beschreibungen erklärt. Dies hatte den Vorteil, dass man es langsam lesen musste und es dadurch auch besser verstand.

Der Mathematiker David Hilbert stellte am 08. August 1900 eine Liste von 23 bis dahin ungelösten Problemen der Mathematik auf. Dabei formulierte Hilbert als sechstes Problem die Frage, wie die Wahrscheinlichkeitstheorie und die Mechanik axiomatisiert werden kann. Mittlerweile hat man diese Fragestellung auf die gesamte Physik erweitert.

Der russische Mathematiker Andrei Nikolajewitsch Kolmogorow schafft im Jahre 1933 die Axiomatisierung. Seine Axiome werden auch Gegenstand dieser Vorlesung sein. Unser Anliegen ist es zunächst, ein mathematisches Modell zur Beschreibung zufälliger Erscheinungen zu Formen. Als Beispiele dienen uns Münze und Würfel sowie auch Messungen.

1.1.2. Mathematisches Modell für zufällige Vorgänge

Das mathematische Modell besteht aus den drei Faktoren: $[\Omega, \mathcal{F}, P]$. Diese haben folgende Bedeutung:

1) Raum der **Elementarereignisse** - Ω
   Dies ist eine *nichtleere* Menge ($\Omega \neq \emptyset$), die mögliche Versuchsausgänge klassifiziert.
   Beispiel: Würfel $\rightarrow \Omega = \{1,2,3,4,5,6\}$, Messung $\rightarrow M \subseteq \mathbb{R}$, Münzwurf $\rightarrow \Omega = \{W, Z\} = \{0,1\}$

2) System der Ereignisse - $\mathcal{F} \in \mathfrak{P}(\Omega)$
   Die Menge $\mathcal{F}$ beinhaltet zulässige Aussagen über das Versuchsergebnis und es gilt $\mathcal{F} \subseteq 2^{\Omega}$*
   Beispiel: Würfel $\rightarrow A = \{2,4,6\} \subseteq \Omega$ entspricht der Aussage „Ich würfle eine gerade Zahl."

3) **Eintrittswahrscheinlichkeit** - P
   P ist eine Abbildung $P \colon \mathcal{F} \rightarrow [0,1]$ und $P(A)$ bezeichnet die Wahrscheinlichkeit mit der das Ereignis $A \in \mathcal{F}$ eintritt.
   Beispiel: Würfel $\rightarrow \Omega = \{1,2,3,4,5,6\}, A = \{2,4,6\}, P(A) = 1/2, P(\{k\}) = \frac{1}{6}, k = 1, \ldots, 6$

Welche sinnvollen Forderungen sollte man nun an $\Omega, \mathcal{F}, P$ stellen? Daraus ergibt sich der Aufbau der Axiomatik.

1.1.3. Eigenschaften von $\mathcal{F}$

- Grundidee: Logische Operatoren werden zu entsprechenden mengentheoretischen Operatoren: $A_1 \vee A_2 \leftrightarrow A_1 \cup A_2, \neg A \leftrightarrow A^c$†
- $\mathcal{F} \neq \emptyset$
- $\forall A_1, A_2 \in \mathcal{F}(A_1 \cup A_2 \in \mathcal{F} \wedge A_1 \cap A_2 \in \mathcal{F})$
- $\forall A_1, A_2, \ldots \in \mathcal{F}(\bigcup_{k=1}^{\infty} A_k \in \mathcal{F})$
- $\forall A \in \mathcal{F}(A^c \in \mathcal{F})$

*Die **Potenzmenge** wird auch mit $\mathfrak{P}$ bezeichnet. Damit gilt hier: $\mathcal{F} \subseteq \mathfrak{P}(\Omega)$
†**Komplement**

1.1.4. Eigenschaften von $P(A)$

- Beispiel: n-mal Würfeln: $h_n(A) = \frac{\text{Anzahl des Eintretens von } A}{n} \xrightarrow{n\to\infty} =: P(A)$. Man bezeichnet $h_n(A)$ als **relative Häufigkeit** des Eintretens.
- $0 \leq h_n(A) \leq 1 \Rightarrow 0 \leq P(A) \leq 1$
- $h_n(\Omega) = 1 = P(\Omega)$
- $A \cap B = \emptyset \Rightarrow h_n(A \cup B) = h_n(A) + h_n(B) \Rightarrow P(A \cup B) = P(A) + P(B)$
- $A_1, A_2, \ldots \in \mathcal{F} \wedge A_j \cap A_K = \emptyset (k \neq j) \Rightarrow P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$

1.2. σ-Algebra

Im folgenden gilt immer:

- $\Omega \neq \emptyset$
- $\mathfrak{P}(\Omega)$ - Potenzmenge
- $\mathcal{F} \subseteq \mathfrak{P}(\Omega)$

Definition
Die Menge $\mathcal{F}$ heißt **Algebra**, wenn gilt:

(1) $\emptyset \in \mathcal{F}$

(2) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$

(3) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$

$\mathcal{F}$ heißt **σ-Algebra**, wenn die Punkte (1) und (2) gelten und zusätzlich
(3') $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup A_k \in \mathcal{F}$ gilt.

Satz 1.2.1
Jede σ-Algebra ist eine Algebra.

Beweis:
Zum Beweis genügt es zu zeigen, dass eine beliebige σ-Algebra $\mathcal{F}$ den Punkt (3) der Definition erfüllt. Denn die ersten beiden Eigenschaften teilen Algebra und σ-Algebra. Es gelten die Eigenschaften (1), (2) und (3'). Man setzt nun $A_1 = A, A_2 = B, A_n = \emptyset (n = 3, 4, \ldots)$. Dann gilt $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup A_k$ ■

Bemerkung

- Die Umkehrung von Satz 1.2.1 ist in der Regel falsch, d. h. es gibt Algebren, die keine σ-Algebra sind. Das klassiche Beispiel hierfür ist:

$$\Omega = \mathbb{N}; \mathcal{F} :\ = \{A \subseteq \mathbb{N} | A \text{ endlich oder} A^c \text{ endlich}\}$$

  Eigenschaften:

(1) klar

(2) Sei $A \in \mathcal{F}$. Dann ist zu zeigen, dass entweder $A^c \in \mathcal{F}$ oder $(A^c)^c$ endlich ist. Falls A^c endlich, ist nichts weiter zu zeigen. Wenn A endlich ist, dann ist nichts zu A^c ausgesagt. Allerdings ist $(A^c)^c = A$. Somit gilt auch Eigenschaft 2.

(3) Für $A, B \in \mathcal{F}$ ist zu zeigen, dass $(A \cup B)$ endlich bzw. $(A \cup B)^c \in \mathcal{F}$ endlich sind.

1. Fall A, B endlich $\Rightarrow A \cup B$ endlich
2. Fall A^c, B^c endlich $\Rightarrow A^c \cap B^c \subseteq A^c$ endlich $\Rightarrow (A \cup B)^c$ endlich $\Rightarrow A \cup B \in \mathcal{F}$
3. Fall A, B^c endlich $\Rightarrow A^c \cap B^c \subseteq B^c$ endlich $\Rightarrow (A \cup B)^c$ endlich $\Rightarrow A \cup B \in \mathcal{F}$
4. Fall A^c, B endlich $\Rightarrow A^c \cap B^c \subseteq A^c$ endlich $\Rightarrow (A \cup B)^c$ endlich $\Rightarrow A \cup B \in \mathcal{F}$

- Jede endliche Algebra ist eine σ-Algebra.
- Jede σ-Algebra ist bezüglich der (symmetrischen) Differenz abgeschlossen.

Satz 1.2.2
Sei $\mathcal{F}$ eine Algebra über Ω. Dann ist $\Omega \in \mathcal{F}$ und für alle Folgen $A_1, \dots, A_n \in \mathcal{F} (n \in \mathbb{N})$ gelten:

a) $\bigcup_{k=1}^n A_k \in \mathcal{F}$

b) $\bigcap_{k=1}^n A_k \in \mathcal{F}$

Beweis:

- Ω ist ein Element von $\mathcal{F}$. Denn nach (1) gilt $\emptyset \in \mathcal{F}$ und nach (2) folgt $\emptyset^c = \Omega \in \mathcal{F}$.

zu a) Nach Eigenschaft (3) gilt der Punkt a) für $n = 2$. Also haben wir $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$. Nun gelte: $\bigcup_{k=1}^n A_k \in \mathcal{F} (A_1, \dots, A_n \in \mathcal{F})$. Im Induktionsschritt können wir zeigen: $A_1, \dots, A_{n+1} \in \mathcal{F} (\bigcup_{k=1}^{n+1} A_k = \bigcup_{k=1}^n A_k \cup A_{n+1}) \Rightarrow \bigcup_{k=1}^{n+1} A_{n+1} \in \mathcal{F}$

zu b) Nach Eigenschaft (3) kann folgender Schluss gezogen werden $A_1, \dots, A_n \in \mathcal{F} \Rightarrow A_1^c, \dots, A_n^c \in \mathcal{F} \Rightarrow \bigcup_{k=1}^n A_k^c \in \mathcal{F}$. Nach den deMorganschen Regeln bedeutet dies: $(\bigcup_{k=1}^n A_k^c)^c \Rightarrow \bigcap_{k=1}^n A_k \in \mathcal{F}$ ∎

Satz 1.2.3
Sei $\mathcal{F}$ eine σ-Algebra. Dann gilt

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{F} \qquad A_1, A_2, \dots \in \mathcal{F}$$

(Beweis wie in Satz 1.2.2 mit $n = \infty$)

Bemerkung
Eine σ-Algebra ist i. A. *nicht* gegenüber überabzählbarer Vereinigung bzw. Durchschnitt abgeschlossen.

Beispiel

1. Die „kleinste“ Algebra/σ-Algebra ist $\mathcal{F} = \{\emptyset, \Omega\}$.
2. Die „größte“ σ-Algebra ist $\mathcal{F} = \mathfrak{P}(\Omega)$.
3. Sei $\emptyset \subset A \subset \Omega, \mathcal{F} = \{\emptyset, \Omega, A, A^c\}$. Der Nachweis, dass dies eine σ-Algebra ist, lässt sich leicht führen.

Satz 1.2.4
Sei $(\mathcal{F}_i)_{i\in I}$ eine Familie von σ-Algebren. Dann ist $\bigcap_{i\in I} \mathcal{F}_i$ eine σ-Algebra.

Beweis:
Sei $\bigcap_{i\in I} \mathcal{F}_i \subseteq \mathfrak{P}(\Omega)$.

zu (1) Für $i \in I$ gilt, dass $\emptyset \in \mathcal{F}_i \Rightarrow \emptyset \in \bigcap_{i\in I} \mathcal{F}_i$

zu (2) Sei $A \in \bigcap_{i\in I} \mathcal{F}_i$. Es ist zu zeigen, dass $A^c \in \bigcap_{i\in I} \in \mathcal{F}_i$:
$A \in \mathcal{F}_i \Rightarrow A^c \in \mathcal{F}_i \Rightarrow A^c \in \bigcap_{i\in I} \mathcal{F}_i$

zu (3') Sei $A_1, A_2, \ldots \in \bigcap_{i\in I} \mathcal{F}_i \Rightarrow A_k \in \mathcal{F}_i (k = 1, 2, \ldots) \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}_i \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \bigcap_{i\in I} \mathcal{F}_i$ ■

Definition
Sei $\mathfrak{E} \subseteq \mathfrak{P}(\Omega)$ und $\mathcal{F}$ eine σ-Algebra.

$$\sigma(\mathfrak{E}) := \bigcap_{\mathfrak{E} \subseteq \mathcal{F}} \mathcal{F}$$

heisst die von $\mathfrak{E}$ erzeugte σ-Algebra. Gilt für eine σ-Algebra $\mathfrak{B}$ die Beziehung

$$\mathfrak{B} = \sigma(\mathfrak{E})$$

so heißt $\mathfrak{E}$ **Erzeugendensystem** von $\mathfrak{B}$.

Bemerkung

1. Aus Satz 1.2.4 folgt $\sigma(\mathfrak{E})$ ist eine σ-Algebra. Gleiche Konstruktion ist auch für alle Algebren möglich.
2. $\mathfrak{E}_1 \subseteq \mathfrak{E}_2 \Rightarrow \sigma(\mathfrak{E}_1) \subseteq \sigma(\mathfrak{E}_2)$
3. Sei $\mathfrak{E} \subseteq \mathfrak{B}$ und $\mathfrak{B}$ eine σ-Algebra. Dann folgt, $\sigma(\mathfrak{E}) \subseteq \mathfrak{B}$

Beispiel

$$\sigma(\{\emptyset\}) = \{\emptyset, \Omega\} \qquad \sigma(\{A\}) = \{\emptyset, \Omega, A, A^c\} \qquad (\emptyset \subset A \subset \Omega)$$

Definition
Sei $\Omega = M$ ein metrischer Raum. Wir setzen $\mathfrak{E} = \{A \subseteq M \colon A \text{ offen}\}$. Dann heißt $\sigma(\mathfrak{E})$ die σ-Algebra der **Borelmengen** aus M, $\mathfrak{B}_n$ ist die σ-Algebra der Borelmengen in $\mathbb{R}^n$ und $\mathfrak{B}$ ist die σ-Algebra der Borelmengen in $\mathbb{R}$.

Satz 1.2.5
Jedes der folgenden Mengensysteme von Intervallen ist ein Erzeugendensystem für $\mathfrak{B}$ (Es gilt: $\mathfrak{E}_1, \dots, \mathfrak{E}_8$ sind abzählbar.):

$$\begin{aligned}
\mathfrak{E}_1 \colon &= \{(\alpha, \beta) | \alpha, \beta \in \mathbb{R} \text{ und rational}\} \\
\mathfrak{E}_2 \colon &= \{[\alpha, \beta) | \alpha, \beta \in \mathbb{R} \text{ und rational}\} \\
\mathfrak{E}_3 \colon &= \{(\alpha, \beta] | \alpha, \beta \in \mathbb{R} \text{ und rational}\} \\
\mathfrak{E}_4 \colon &= \{[\alpha, \beta] | \alpha, \beta \in \mathbb{R} \text{ und rational}\} \\
\mathfrak{E}_5 \colon &= \{(-\infty, \beta) | \beta \in \mathbb{R} \text{ und rational}\} \\
\mathfrak{E}_6 \colon &= \{(-\infty, \beta] | \beta \in \mathbb{R} \text{ und rational}\} \\
\mathfrak{E}_7 \colon &= \{(\alpha, \infty) | \alpha \in \mathbb{R} \text{ und rational}\} \\
\mathfrak{E}_8 \colon &= \{[\alpha, \infty) | \alpha \in \mathbb{R} \text{ und rational}\}
\end{aligned}$$

Beweis:
zu $\mathfrak{E}_1$ Sei $\mathfrak{E}_1 \subseteq \mathfrak{E} = \{A \subseteq \mathbb{R} | A \text{ offen}\} \Rightarrow \mathfrak{B} = \sigma(\mathfrak{E})$. Nach der Definition gilt $\mathfrak{E}_1 \subseteq \mathfrak{E} \Rightarrow \sigma(\mathfrak{E}_1) \subseteq \sigma(\mathfrak{E}) = \mathfrak{B}$. Damit ist noch zu zeigen, dass $\mathfrak{B} \subseteq \sigma(\mathfrak{E}_1)$. Sei $A \in \mathfrak{E}$. Setzen $K_r(x) = (x - r, x + r) \Rightarrow A = \bigcup_{K_r(x) \subseteq A} K_r(x)$
$K_r(x) \in \mathfrak{E}_1 \subseteq \sigma(\mathfrak{E}_1) \Rightarrow A \in \sigma(\mathfrak{E}_1) \Rightarrow \mathfrak{E} \subseteq \sigma(\mathfrak{E}_1) \Rightarrow \sigma(\mathfrak{E}) = \mathfrak{B} \subseteq \sigma(\sigma(\mathfrak{E}_1)) = \sigma(\mathfrak{E}_1)$ ∎

1.3. Maße

Im folgenden gilt immer: $\Omega \neq \emptyset$ und $\mathcal{F}$ ist eine σ-Algebra über Ω.

Definition
- Das Tupel $[\Omega, \mathcal{F}]$ heißt **messbarer Raum**.
- Eine nichtnegative Funktion $\mu \colon \mathcal{F} \to [0, +\infty]$ auf $\mathcal{F}$ heißt **Maß** (auf $[\Omega, \mathcal{F}]$), wenn gilt 1. $\mu(\emptyset) = 0$ 2. Für paarweise disjunkte Folgen $(A_k)_{k=1}^{\infty}$ aus $\mathcal{F}$ gilt:
$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$$
- Das Tripel $[\Omega, \mathcal{F}, \mu]$ heißt **Maßraum**.

Beispiel
(1) **Nullmaß** $\mathcal{O}$: $\mathcal{O}(A) = 0$ für $A \in \mathcal{F}$

(2) **Dirac-Maß**: Sei $\omega \in \Omega$. Setzen $\delta_\omega(A) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$

(3) Sei Ω abzählbar und $\mathcal{F} = \mathfrak{P}(\Omega)$. Dann ist das **Zählmaß** $\mu(A) \colon = \#(A)$ die Anzahl der Teilmengen A aus Ω.

Satz 1.3.1
Sei $(\mu_i)_{i \in I}$ eine abzählbare Familie von Maßen auf $[\Omega, \mathcal{F}]$ und $(\alpha_i)_{i \in I}$ eine Familie nichtnegativer reller Zahlen. Durch

$$\mu(A) \colon = \sum_{i \in I} \alpha_i \mu_i(A) \quad (A \in \mathcal{F})$$

ist ein Maß auf $[\Omega, \mathcal{F}]$ definiert.‡

Beweis:
Diese Funktion ist nichtnegativ, da alle α_i nichtnegativ sind und die $\mu_i(A)$ per Definition ebenfalls positiv sind. Die Summenbildung ändert nichts.

zu 1) $\mu(\emptyset) = \sum_{k=1}^{\infty} \alpha_i \mu_i(\emptyset) = 0$. Denn $\mu_i(\emptyset) = 0$.

zu 2) Seien $(A_k)_{k=1}^{\infty}$ paarweise disjunkt aus $\mathcal{F}$.

$$\begin{aligned} \mu(\bigcup_{k=1}^{\infty} A_k) &= \sum_{i \in I} \alpha_i \mu_i \left(\bigcup_{k=1}^{\infty} A_k \right) = \sum_{i \in I} \alpha_i \sum_{k=1}^{\infty} \mu_i(A_k) \\ &= \sum_{k=1}^{\infty} \underbrace{\sum_{i \in I} \alpha_i \mu_i(A_k)}_{\mu(A_k)} = \sum_{k=1}^{\infty} \mu(A_k) \end{aligned}$$

Im letzten Schluss musste der große Umordnungssatz angewendet werden. ∎

Beispiel
Sei $\Omega = \mathbb{N}$ und $\mathcal{F} = \mathfrak{P}(\mathbb{N})$. Dann kann man das Maß wie folgt definieren: $\mu = \sum_{k=1}^{\infty} \alpha_k \delta_k$.

Satz 1.3.2
Sei μ ein Maß auf $[\Omega, \mathcal{F}]$. Für alle Folgen $A_1, \ldots, A_n \in \mathcal{F}$, die paarweise disjunkt sind, gilt:

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

Diese Eigenschaft heißt **endliche Additivität**.

Beweis:
Wir setzen $B_k \colon = \begin{cases} A_k & k = 1, \ldots, n \\ \emptyset & k > n \end{cases}$. Damit ist folgendes bekannt:

‡Schreibweise: $\mu = \sum_{i \in I} \alpha_i \mu_i$

1. $\forall k \in \mathbb{N}(B_k \in \mathcal{F})$
2. $\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{n} A_k$
3. $A_i \cap A_j = A_i \cap B_j = \emptyset$ für $i \neq j$ todo: stimmen die Indizes?
4. $\mu(A_k) = \mu(B_k)$ für $k = 1, \dots, n$
5. $\mu(B_k) = 0$ für $(k > n)$

Damit folgt nun: $\mu(\bigcup_{k=1}^{n} A_k) \overset{(2)}{=} \mu(\bigcup_{k=1}^{\infty} B_k) \overset{(1),(3)}{=} \sum_{k=1}^{\infty} \mu(B_k) \overset{(5)}{=} \sum_{k=1}^{n} \mu(B_k) \overset{(4)}{=} \sum_{k=1}^{n} \mu(A_k)$ ■

Bemerkung

1. Für $n = \infty$ im Satz 1.3.2 wird die Additivität als **σ-Additivität** bezeichnet.
2. Satz 1.3.2 impliziert, dass aus der σ-Additivität die endliche Additivität folgt. Die Umkehrung ist jedoch falsch.
3. **Subadditivität**: Für alle $A_1, \dots, A_n \in \mathcal{F}$ gilt: $\mu(\bigcup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} \mu(A_k)$. Der Beweis wird per Induktion geführt.

Beispiel
Sei $\Omega = \mathbb{N}, \mathcal{F} = \mathfrak{P}(\mathbb{N})$ und das Maß ist definiert: $\mu(A) = \begin{cases} 0 & A < \infty \\ +\infty & A = \infty \end{cases}$. Sei weiter $A_k = \{2k\}$. Dann gilt $\mu(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mu(A_k) = 0$. Denn auch die Vereinigung ist wieder endlich. Wenn wir zur unendlichen Vereinigung übergehen, ändert sich die Situation. Die Menge, die durch die Vereinigung entsteht, ist die Menge der geraden Zahlen. Diese ist unendlich als gilt: $\mu(\bigcup_{k=1}^{\infty} A_k) = +\infty \neq 0 = \sum_{k=1}^{\infty} A_k$.

Satz 1.3.3 (elementare Eigenschaften)
Sei $[\Omega, \mathcal{F}, \mu]$ ein Maßraum und $A, B \in \mathcal{F}$. Dann gilt:

(a) $\mu(A \cup B) \leq \mu(A) + \mu(B)$

(b) $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$ (Monotonie des Maßes)

(c) $A \subseteq B, \mu$ endlich $\Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$

Beweis:
zu (b) und (c) $A \subseteq B \Rightarrow B = A \cup (B \setminus A) \Rightarrow \mu(B) = \mu(A \cup (B \setminus A)) \overset{\text{Satz 1.3.2}}{=} \mu(A) + \mu(B \setminus A)$ Wir haben $\mu(B \setminus A) \geq 0 \Rightarrow \mu(A) \leq \mu(B)$ und Teil (b) ist gezeigt.
Durch Umstellen erhält man $\mu(B \setminus A) = \mu(B) - \mu(A)$. Hierzu muss $\mu(A) < +\infty$ sein.

zu (a) Wie oben haben wir: $A \cup B = A \cup (B \setminus A) \Rightarrow \mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A)$.
Es gilt: $B \setminus A \subseteq B \overset{(b)}{\Rightarrow} \mu(B \setminus A) \leq \mu(B) \Rightarrow \mu(A \cup B) - \mu(A) \leq \mu(B) \Rightarrow \mu(A \cup B) \leq \mu(A) + \mu(B)$ ■

Satz 1.3.4
Seien μ ein endliches Maß ($\mu(\Omega) < +\infty$) und $(A_n)_{n=1}^{\infty}$ eine Folge aus $\mathcal{F}$. Dann gelten:

(a) (Stetigkeit von unten): Ist $(A_n)_{n=1}^{\infty}$ monoton wachsend, so gilt:

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

(b) (Stetigkeit von oben): Ist $(A_n)_{n=1}^{\infty}$ monoton fallend, so gilt:

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

Beweis:
zu a) Für den Beweis ist die Forderung $\mu < \infty$ nicht notwendig. Setzen $B_1 : = A_1, B_{n+1} = A_{n+1} \setminus A_n$. Damit gelten folgende Tatsachen:

1. Alle B_k sind paarweise disjunkt.
2. $B_k \in \mathcal{F}$
3. $\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k$
4. $\bigcup_{k=1}^{\infty} B_k = A_n$

Somit folgt: $\mu(\bigcup_{k=1}^{\infty} A_k) \stackrel{3}{=} \mu(\bigcup_{k=1}^{\infty} B_k) \stackrel{1,2}{=} \sum_{k=1}^{\infty} \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) \stackrel{1}{=} \lim_{n\to\infty} \mu(\bigcup_{k=1}^{n} B_k) \stackrel{4.}{=} \lim_{n\to\infty} \mu(A_n)$

zu b) $A_n \supseteq A_{n+1} \Rightarrow A_n^c \subseteq A_{n+1}^c$, d. h. die Folge der Komplemente ist monoton wachsend.
Nach (a) gilt nun: $\mu(\bigcup_{n=1}^{\infty} A_n^c) = \lim_{n\to\infty} \mu(A_n^c)$
$\mu(\bigcup_{n=1}^{\infty} A_n^c) = \mu((\bigcap_{n=1}^{\infty} A_n)^c) = \mu(\Omega \setminus (\bigcap_{n=1}^{\infty} A_n)) \stackrel{\text{Satz 1.3.3 c}}{=} \mu(\Omega) - \mu(\bigcap A_n)$
$\mu(A_n^c) = \mu(\Omega \setminus A_n) = \mu(\Omega) - \mu(A_n) \Rightarrow \lim_{n\to\infty} \mu(A_n)$.
In der Literatur werden z. T. schwächere Forderungen angesetzt. ∎

1.4. Produkte von Maßräumen

Im folgenden gilt immer: $[\Omega_1, \mathcal{F}_1, \mu_1], \ldots, [\Omega_n, \mathcal{F}_n, \mu_n]$ eine Folge von Maßräumen. Wir betrachten direkte Produkte von Mengen: $\times_{i=1}^{n} A_i : = \{[w_1, \ldots, w_n] | w_i \in A_i (i = 1, \ldots, n)\}$. Das Ziel ist es, einen Maßraum mit der Grundmenge $\Omega = \times_{i=1}^{n} \Omega_i$ zu bauen.

Definition
- Die Menge
$$\otimes_{i=1}^{n} \mathcal{F}_i : = \sigma(\{\times_{i=1}^{n} A_i | A_i \in \mathcal{F}_i (i = 1, \ldots, n)\})$$
heißt **Produkt** der σ-Algebren $\mathcal{F}_1, \ldots, \mathcal{F}_n$.

- Die Menge
$$\times_{i=1}^n[\Omega_i,\mathcal{F}_i]\colon=[\times_{i=1}^n\Omega_i,\otimes_{i=1}^n\mathcal{F}_i]$$
heißt **Produkt** der messbaren Räume $[\Omega_1,\mathcal{F}_1,\dots,\Omega_n,\mathcal{F}_n]$.

Bemerkung

1. Das Produkt der σ-Algebren $\otimes\mathcal{F}_i$ ist eine σ-Algebra auf $\times_{i=1}^n\Omega_i$.
2. Für $n=2$ nutzen wir die Schreibweise $\Omega_1\times\Omega_2,\mathcal{F}_1\otimes\mathcal{F}_2$.

Beispiel

1. $\mathcal{F}_i=\{\emptyset,\Omega_i\}\Rightarrow\otimes_{i=1}^n\mathcal{F}_i=\{\emptyset,\times_{i=1}^n\Omega_i\}$
2. Sei Ω_1 bis Ω_n abzählbar: $\mathcal{F}_i=\mathfrak{P}(\Omega_i)\Rightarrow\otimes_{i=1}^n\mathcal{F}_i=\mathfrak{P}(\times_{i=1}^n\Omega_i)$. Denn wegen der Abzählbarkeit der Ω_i gilt, dass auch $\times_{i=1}^n\Omega_i$ abzählbar sind. Diese sind eine Obermenge von $A=\bigcup_{\omega\in A}\{\omega\}$ und die Vereinigung ist wieder abzählbar. Nun genügt es zu zeigen, dass $\{\underbrace{[\omega_1,\dots,\omega_n]}_{\times_{i=1}^n\{\omega_i\}}\}\in\otimes_{i=1}^n\mathcal{F}_i$ für $\omega_i\in\Omega_i,i=1,\dots,n$. Der Rest ist klar, denn $\{\omega_i\}\in\mathcal{F}_i=\mathfrak{P}(\Omega_i)$
3. Lässt man oben den Zusatz „abzählbar" weg, gibt es Probleme.

Definition

Ein Maß μ auf $\times_{i=1}^n[\Omega_i,\mathcal{F}_i]$ heißt **Produkt der Maße** $\mu_1,\dots,\mu_n$[§], wenn gilt
$$\mu(\times_{i=1}^nA_i)=\prod_{i=1}^n\mu_i(A_i)$$

Satz 1.4.1

Für endliche Maße $\mu_1,\dots,\mu_n$ existiert das Produktmaß und ist eindeutig bestimmt.

Bemerkung

1. Die Forderung, dass die $\mu_1,\dots,\mu_n$ endlich sein sollen, kann man abschwächen (siehe hierzu auch Lebesguesches Maß).
2. Der Fall $n=+\infty$ ist für Wahrscheinlichkeitsmaße interessant.

Definition

Sei $\mathfrak{B}_n$ die σ-Algebra der Borelmengen auf $\mathbb{R}^n$. Ein Maß ℓ_n auf $[\mathbb{R}^n,\mathfrak{B}_n]$ heißt n-dimensionales **Lebesguesches Maß**, wenn gilt:
$$\ell_n\left(\times_{i=1}^n[a_i,b_i)\right)=\prod_{i=1}^n(b_i-a_i)\quad(a_i<b_i,i=1,\dots,n)$$

Satz 1.4.2

Das Lebeguesche Maß ℓ_n existiert und ist eindeutig bestimmt. Weiterhin gilt:
$$[\mathbb{R}^n,\mathfrak{B}_n,\ell_n]=\times_{i=1}^n[\mathbb{R},\mathfrak{B},\ell]=[\mathbb{R},\mathfrak{B},\ell]^{nx}=[\mathbb{R}^n,\mathfrak{B}_n,\ell^{nx}]$$

[§] $\mu=\times_{i=1}^n\mu_i$

Bemerkung
Die Bemerkungen sind gleichzeitig Beweisideen für den oben angegebenen Satz.

- $\mathfrak{B}_n = \otimes_{i=1}^n \mathfrak{B} = \mathfrak{B}^{n\otimes}$. Weiterhin ist klar, dass $\mathbb{R} \supseteq A_1, \dots, A_n$ offen. Damit ist auch $\times_{i=1}^n A_i$ offen.

- Wir haben $\ell_n(\times_{i=1}^n A_i) = \ell^{n\otimes}(\times_{i=1}^n A_i)$ für Mengen $A_i = [a_i, b_i)$. Ferner wissen wir, dass diese Mengen $\mathfrak{B}$ erzeugen (siehe auch Satz 1.2.5). Die allgemeine Theorie führt zur Behauptung des oben stehenden Satzes.

Satz 1.4.3
Sei $x \in \mathbb{R}$. Dann gilt: $\ell(\{x\}) = 0$.

Beweis:
Wir haben $\ell([a,b)) = b - a$ mit $a < b$. Nun betrachten wir die Mengenfolge $B_k := [x, x + \frac{1}{n}) \Rightarrow B_k \supseteq B_{k+1} \Rightarrow \bigcap_{k=1}^\infty B_k = \{x\}$. Wegen der Stetigkeit von oben (siehe Satz 1.3.4) folgt nun: $\ell(\{x\}) = \ell(\bigcap_{k=1}^\infty B_k) = \lim_{k\to\infty} \ell(B_k) = \lim_{k\to\infty} \underbrace{\ell([x, x + \frac{1}{k}))}_{\frac{1}{k}} =$
$\lim_{k\to\infty} \frac{1}{k} = 0$ ■

Satz 1.4.4 (Folgerung aus Satz 1.4.3)
Seien $a < b \Rightarrow \ell([a,b]) = \ell((a,b)) = \ell((a,b]) = \ell([a,b)) = b - a$

Beweis:

- Es gilt $[a,b] = [a,b) \cup \{b\}$. Der Durchschnitt beider Mengen ist leer ($[a,b) \cap \{b\} = \emptyset$). Damit folgt: $\ell([a,b]) = \ell([a,b) \cup \{b\}) = \ell([a+b)) + \ell(\{b\}) = b - a + 0$

- $(a,b] = [a,b] \setminus \{a\}$
  Es ist klar, dass $\{a\}$ in $[a,b]$ liegt. Damit folgt nach Punkt c von Satz 1.3.3: $\ell((a,b]) = \ell([a,b]) - \ell(\{a\}) = (b-a) - 0 = b - a$

- $(a-b) = (a,b] \setminus \{b\}$
  Wie schon oben ist klar, dass $\{b\}$ in $(a,b]$ liegt. Nach gleichem Schluss folgt auch hier: $\ell((a,b)) = \ell((a,b]) - \ell(\{b\}) = (b-a) - 0 = b - a$ ■

Satz 1.4.5
Für alle abzählbaren $A \subseteq \mathbb{R}^n$ gilt:

$$\ell^{(n)}(A) = 0$$

Beweis:
$A = \bigcup_{x\in A}\{x\} = \bigcup_{[x_1,\dots,x_n]\in A}\{[x_1, \dots, x_n]\}$
Da A abzählbar ist, folgt $\ell^{(n)}(A) = \sum_{x\in A} \ell^{(n)}\{x\}$. Nunmehr ist noch zu zeigen, dass $\ell^{(n)}(\{[x_1, \dots, x_n]\}) = 0$. Dies ist gleich: $\ell^{(n)} \times_{i=1}^n \{x_i\} = \prod_{i=1}^n \ell(\{x_i\}) = 0$. ■

1.5. Wahrscheinlichkeitsmaße und Verteilungsfunktion

Definition
Ein Maß P auf einem meßbaren Raum $[\Omega, \mathcal{F}]$ heißt **Wahrscheinlichkeitsmaß**, wenn gilt:

$$P(\Omega) = 1$$

In diesem Fall heißt $[\Omega, \mathcal{F}, P]$ **Wahrscheinlichkeitsraum**.

Satz 1.5.1
Sei P ein Wahrscheinlichkeitsmaß auf $[\Omega, \mathcal{F}]$. Dann gelten:

(a) $0 \leq P(B) \leq 1 \quad (B \in \mathcal{F})$

(b) $P(B^c) = 1 - P(B) \quad (B \in \mathcal{F})$

Beweis:
(a) klar, da $P(\emptyset) \leq P(B) \leq P(\Omega) \Rightarrow 0 \leq P(B) \leq 1$

(b) $B^c = \Omega \setminus B, B \subseteq \Omega \Rightarrow P(B^c) = P(\Omega) - P(B) \Rightarrow P(B^c) = 1 - P(B)$ ■

Definition
Sei P ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}^n, \mathfrak{B}_n]$ mit $n \geq 1$. Eine Abbildung $F\colon \mathbb{R}^n \to \mathbb{R}$ heißt **Verteilungsfunktion** von (zu) P, wenn gilt:

$$F(x_1, \ldots, x_n) = P(\times_{i=1}^n (-\infty, x_i)) \qquad (x_1, \ldots, x_n \in \mathbb{R})$$

Satz 1.5.2
Ein Wahrscheinlichkeitsmaß P ist durch seine Verteilungsfunktion F eindeutig bestimmt. Dabei sind die P und F wie oben definiert.

Beweis:
Beweisidee: Wenn man ein durchschnittsabgeschlossenes Erzeugendensystem hat und darauf ein Wahrscheinlichkeitsmaß kennt, kennt man es auf der gesamten σ-Algebra. ■

Satz 1.5.3
Eine Verteilungsfunktion F eines Wahrscheinlichkeitsmaßes P auf $[\mathbb{R}, \mathfrak{B}]$ hat folgende Eigenschaften:

(F1) $F(x) \leq F(y) \quad \forall (x \leq y)$ (monoton nicht fallend bzw. wachsend)

(F2) $\lim_{n\to\infty} F(x_n) = F(x)$ mit $x \in \mathbb{R}, x_n \uparrow x$ (linksseitig stetig)

(F3) $\lim_{x\to+\infty} F(x) = 1$

(F4) $\lim_{x\to-\infty} F(x) = 0$

Beweis:
(F1) Sei $F(x) := P((-\infty, x))$ mit $x \in \mathbb{R}$. Dann gilt $(-\infty, x) \subseteq (-\infty, y)$, da $x \leq y$. Nach Satz 1.3.3 Punkt b hat man $F(x) \leq F(y) \Rightarrow P((-\infty, x)) \leq P((-\infty, y)) \Rightarrow F(x) \leq F(y)$

(F2) Seien $x \in \mathbb{R}$ und $x_n \leq x_{n+1}, \lim x_n = x$. Damit folgt: $(-\infty, x_n) \subseteq (-\infty, x_{n+1})$ und $(-\infty, x) = \bigcup_{n=1}^{\infty}(-\infty, x_n)$. Unter Zuhilfenahme von Satz 1.3.4 a erhält man $F(x) = P((-\infty, x)) = P(\bigcup_{n=1}^{\infty}(-\infty, x_n)) = \lim_{n\to\infty} P((-\infty, x_n)) = \lim F(x_n)$.

(F3) Sei $x_n \uparrow +\infty$[¶]. Damit gilt wieder die schon oben angeführte Teilmengenbeziehung und es folgt: $\bigcup_{n=1}^{\infty}(-\infty, x_n) = \mathbb{R}$. Wie bereits oben erhält man nun durch den Satz 1.3.4 a, dass $1 = P(\mathbb{R}) = P(\bigcup_{n=1}^{\infty}(-\infty, x_n)) = \lim P((-\infty, x_n))$

(F4) Sei nun $x_n \downarrow -\infty \Rightarrow (-\infty, x_n) \supseteq (-\infty, x_{n+1}) \Rightarrow \bigcap_{n\in\mathbb{N}}(-\infty, x_n) = \emptyset$. Damit ist nun $P(\emptyset) = 0 = \lim F(x_n)$ ∎

Satz 1.5.4
Sei $F\colon \mathbb{R} \to \mathbb{R}$ mit den Eigenschaften (F1) bis (F4). Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $[\mathbb{R}^n, \mathfrak{B}_n]$ mit $F(x) = P((-\infty, x))$. Dies bedeutet, dass F die Verteilungsfunktion von P ist.

Bemerkung

1. Sei F die Verteilungsfunktion zu einem Wahrscheinlichkeitsmaß P auf $[\mathbb{R}^n, \mathfrak{B}_n]$. Dann gelten folgende Eigenschaften:

   ($\hat{\text{F}}$1) $F(x_1, \ldots, x_n) \leq F(y_1, \ldots, y_n)$ für $(x_k \leq y_k, k = 1, \ldots, n)$

   ($\hat{\text{F}}$2) $x_k^{(m)} \uparrow x_k \Rightarrow \lim_{m\to\infty} F(x_1^{(m)}, \ldots, x_n^{(m)}) = F(x_1, \ldots, x_m)$

   ($\hat{\text{F}}$3) $x_k^{(m)} \uparrow +\infty \Rightarrow \lim_{m\to\infty} F(x_1^{(m)}, \ldots, x_n^{(m)}) = 1$

   ($\hat{\text{F}}$4) $x_k^{(m)} \uparrow -\infty \Rightarrow \lim_{m\to-\infty} F(x_1^{(m)}, \ldots, x_n^{(m)}) = 0$

2. Sei nun $F\colon \mathbb{R}^n \to [0, 1]$ mit den Eigenschaften ($\hat{\text{F}}$1) bis ($\hat{\text{F}}$4). Dann muss nicht notwendigerweise ein Wahrscheinlichkeitsmaß P auf $[\mathbb{R}^n, \mathfrak{B}_n]$ mit der Verteilungsfunktion P existieren. Somit kann der obenstehende Satz nicht auf $\mathbb{R}^n$ verallgemeinert werden.

Beispiel
Sei $\lambda > 0$. Wir setzen $F_\lambda(x) \colon = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$. Die Eigenschaften (F1) bis (F4) gelten hier und sind trivial nachzuweisen. Das Wahrscheinlichkeitsmaß Ex_λ mit der Verteilungsfunktion F_λ heißt **Exponentialverteilung** mit dem Parameter λ.

Beweis:
Für die Eigenschaft (F1) nehmen wir o. B. d. A. an, dass $x \leq y$ gilt. Dann können wir folgende Fälle unterscheiden:

1. Fall $y \leq 0 \Rightarrow F_\lambda(x) = F_\lambda(y)$

2. Fall $y > 0, x \leq 0 \Rightarrow F_\lambda(x) = 0, F_\lambda(y) = 1 - e^{-\lambda y} \geq 0 \Rightarrow F_\lambda(x) \leq F_\lambda(y)$

3. Fall $0 < x \leq y \Rightarrow e^{\lambda x} \leq e^{\lambda y} \Rightarrow e^{-\lambda x} \geq e^{-\lambda y} \Rightarrow 1 - e^{-\lambda x} \leq 1 - e^{-\lambda y} \Rightarrow F_\lambda(x) \leq F_\lambda(y)$.

[¶]d. h. $x_n \leq x_{n+1}, x_n \xrightarrow{n\to\infty} +\infty \Rightarrow (-\infty, x_n) \leq (-\infty, x_{n+1})$

Die Eigenschaften (F2) und (F4) sind sofort klar. Zu (F3) ist $e^{\lambda x} \xrightarrow{x \to +\infty} +\infty \Rightarrow e^{-\lambda x} \xrightarrow{x \to +\infty} 0 \Rightarrow 1 - e^{-\lambda x} \xrightarrow{x \to +\infty} 1$. Somit gelten alle vier Bedingungen. ■

Beispiel
Sei $A \in \mathfrak{B}_n$ mit $0 < \ell^{(n)}(A) \leq +\infty$. Wir setzen:

$$\mathrm{Gl}_A(B) :\, = \frac{\ell^{(n)}(A \cap B)}{\ell^{(n)}(A)} \qquad (B \in \mathfrak{B}_n)$$

Dann heißt Gl_A die **Gleichverteilung** zu A. Es gilt natürlich, dass Gl_A ein Wahrscheinlichkeitsmaß ist.

Für den Spezialfall $A = [0,1]$ mit $\mathrm{Gl}_A(B) = \ell([0,1] \cap B)$ ist als Übungsaufgabe zu betrachten.

Bemerkung
Die Gleichverteilung ist ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}^n, \mathfrak{B}_n]$. Denn es gilt für $(A_k)_{k=1}^{\infty}$ paarweise disjunkte Mengen aus $\mathfrak{B}$:

$$\begin{aligned}
\mathrm{Gl}_A(\emptyset) &= \frac{\ell^{(n)}(A \cap \emptyset)}{\ell^{(n)}(A)} = 0 \\
\mathrm{Gl}_A(\mathbb{R}^n) &= \frac{\ell^{(n)}(A \cap \mathbb{R})}{\ell^{(n)}(A)} = \frac{\ell^{(n)}(A)}{\ell^{(n)}(A)} = 1 \\
\mathrm{Gl}_A\left(\bigcup_k B_K\right) &= \frac{\ell^{(n)}((\bigcup B_k) \cap A)}{\ell^{(n)}(A)} = \frac{\ell^{(n)}(\bigcup (B_k \cap A))}{\ell^{(n)}(A)} = \frac{\sum_k \ell^{(n)}(B_k \cap A)}{\ell^{(n)}(A)} \\
&= \qquad \sum_k \mathrm{Gl}_A(B_k)
\end{aligned}$$

Satz 1.5.5
Sei P ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$ für $k = 1, \ldots, n$ mit der Verteilungsfunktion F_k und $P = \times_{k=1}^{n} P_k$. Dann ist P ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}^n, \mathfrak{B}_n]$ und die Verteilungsfunktion F zu P hat die Gestalt:

$$F(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_k(x_k) \quad (x_1, \ldots, x_n \in \mathbb{R})$$

Die Folge wird auch **Tensorprodukt** genannt.

Beweis:
Wir müssen zeigen, dass P ein Wahrscheinlichkeitsmaß ist. Dazu wissen wir, dass P ein Maß ist. Daher müssen wir noch zeigen, dass $P(\mathbb{R}^n) = 1$. Es gilt, $P(\mathbb{R}^n) = P(\times_{k=1}^{n} \mathbb{R}^n) = (\times_{k=1}^{n} P_k)(\mathbb{R}^{nx}) = \prod_{k=1}^{n} \underbrace{P_k(\mathbb{R})}_{=1} = 1$.

Nun ist noch zu zeigen, dass F eine Verteilungsfunktion zu P ist, d. h. $F(x_1, \ldots, x_n) = P(\times_{k=1}^{n}(-\infty, x_k)) = \times_{k=1}^{n} P_k(\times_{k=1}^{n}(-\infty, x_k)) = \prod_{k=1}^{n} P_k((-\infty, x_k)) = \prod_{k=1}^{n} F_k(x_k)$. ■

Satz 1.5.6
Sei F eine Verteilungsfunktion eines Wahrscheinlichkeitsmaßes P auf $[\mathbb{R}, \mathfrak{B}]$. Dann gelten:

(1) $P([a,b)) = F(b) - F(a)$ für $(a < b)$

(2) $P(\{a\}) = \underbrace{F(a+0)}_{\lim_{x_n \downarrow a} F(a)} - F(a)$

Beweis:

zu (1) Die Frage ist zunächst, wie sich das Intervall durch Halbstrahlen darstellen lässt. Wir wissen, $(-\infty, b) \setminus (-\infty, a) = [a, b)$ und $(-\infty, a) \subseteq (-\infty, b)$. Nach Satz 1.3.3 c ist $P([a,b)) = P((-\infty, b)) \setminus (-\infty, a)) = P((-\infty, b)) - P((-\infty, a)) = F(b) - F(a)$

zu (2) Es gilt, $\{a\} = \bigcap_{n \geq 1} [a, a + \frac{1}{n})$ und $[a, a + \frac{1}{n}) \supseteq [a, a + \frac{1}{n+1})$. Wegen Satz 1.3.4 Punkt b ist $P(\{a\}) = P(\bigcap_{n \geq 1} [a, a + \frac{1}{n})) = \lim P([a, a + \frac{1}{n})) = \lim_{n \to \infty} F(a + \frac{1}{n}) - F(a) = F(a+0) - F(a)$. ∎

Satz 1.5.7
Seien I abzählbar, P_i ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$, $\alpha_i \geq 0$ mit $i \in I$ und $P = \sum_{i \in I} \alpha_i P_i$. Dann ist P genau dann ein Wahrscheinlichkeitsmaß, wenn gilt

$$\sum_{i \in I} \alpha_i = 1$$

In diesem Fall gilt für die Verteilungsfunktion F von P

$$F(x) = \sum_{i \in I} \alpha_i F_i(x) \quad (x \in \mathbb{R})$$

wobei F_i die Verteilungsfunktion zu P_i ist.

Beweis:
Es ist klar, dass P ein Maß auf $[\mathbb{R}, \mathfrak{B}]$ ist (Folgerung aus Satz 1.3.1). Nun verbleibt zu zeigen, dass $P(\mathbb{R}) = 1$. Dies ist genau dann der Fall, wenn gilt $\sum \alpha_i = 1$: $P(\mathbb{R}) = \sum_{i \in} \alpha_i P_i(\mathbb{R}) = \sum_{i \in I} \alpha_i = 1$. Nach der Definition gilt $F(x) = P((-\infty, x)) = \sum_{i \in I} \alpha_i P_i((-\infty, x)) = \sum_{i \in I} \alpha_i F_i(x)$. ∎

Beispiel
Dies sind im wesentlichen Anwendungen zum oben genannten Satz:

- Für jedes $y \in \mathbb{R}$ ist δ_y ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$.
- **Binomialverteilung**: Gegeben ist $0 \leq p \leq 1, n \in \mathbb{N}$. Hierzu bilden wir

$$B_{n,p} := \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} \delta_i$$

Aus dem oben angeführten Satz folgt, dass die Binomialverteilung ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$ ist.[‖]

[‖] $(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}; a = p, b = 1 - p \Rightarrow (a+b)^n = 1$

- **Poissonverteilung** Gegeben ist $\lambda > 0$. Wir definieren:

$$\pi_\lambda := \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} \delta_i$$

  Wieder aus dem obigen Satz folgt und aus $\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^\lambda \Rightarrow e^\lambda e^{-\lambda} = 1$ folgt, dass das ein Wahrscheinlichkeitsmaß ist.

- **geometrische Verteilung**: Gegeben ist $0 < q < 1$.

$$G_q \colon = \sum_{i=1}^{\infty} q^{i-1}(1-q)\delta_i$$

  Auch dies ist ein Wahrscheinlichkeitsmaß, da $\lim \sum q^{i-1} = \frac{1}{1-q}$.

2. Integrationstheorie

Hier soll ein allgemeiner Integralbegriff eingeführt werden. Wir brauchen das später für das Rechnen mit dem Erwartungswert.

Man stelle sich vor, jemand führt n Würfe mit einem Würfel aus und berechnet $1/n \sum_{i=1}^{n} x_i$. Für unendlich viele Würfe konvergiert das gegen den Erwartungswert $E\xi = \int \xi(\omega)P(d\omega)$.

Es wird zwei wichtige Spezialfälle geben:

1. $\sum_k k\alpha_k$
2. $\int_{-\infty}^{\infty} f(x)\, dx$

2.1. Messbare Funktionen (Abbildungen)

Im folgenden gilt immer, dass $[\Omega, \mathcal{F}, \mu]$ ein Maßraum und $[X, \mathfrak{X}]$ ein messbarer Raum ist. Wir betrachten $f\colon \Omega \to X$ und definieren das **Urbild** einer Menge: $f^{-1}(A)\colon = \{\omega \in \Omega | f(\omega) \in A\}$ mit $A \subseteq X$. Dies ist nicht zu verwechseln mit inversen Abbildungen.*

Satz 2.1.1 (Elementare Eigenschaften des Urbilds)
Es gelten

a) $f^{-1}(\emptyset) = \emptyset, f^{-1}(X) = \Omega$

b) $f^{-1}(A^c) = (f^{-1}(A))^c \quad (\forall A \subseteq X)$, dabei ist $f^{-1}(A^c)$ das Komplement in X und $(f^{-1}(A))^c$ das Komplement in Ω.

c) $f^{-1}(\bigcup_{i\in I} A_i) = \bigcup_{i\in I} f^{-1}(A_i) \quad (A_i \subseteq X, i \in I)$

d) $f^{-1}(\bigcap_{i\in I} A_i) = \bigcap_{i\in I} f^{-1}(A_i)$ (Für beide Fälle gilt, dass die $(A_i)_{i\in I}$ Folgen von Teilmengen aus X sind.)

Beweis:
zu a) $f(\omega) \in \emptyset$ für jedes $\omega \in \Omega \Rightarrow f^{-1}(\emptyset) = \emptyset$
$f(\omega) \in X$ ist richtig $\forall \omega \in \Omega \Rightarrow f^{-1}(X) = \Omega$

zu b) $\omega \in f^{-1}(A^c) \Leftrightarrow f(\omega) \in A^c \Leftrightarrow \neg f(\omega) \in A \Leftrightarrow \neg \omega \in f^{-1}(A) \Leftrightarrow \omega \notin f^{-1}(A) \Leftrightarrow \omega \in (f^{-1}(A))^c \Rightarrow f^{-1}(A^c) = (f^{-1}(A))^c$

*Falls es eine inverse Abbildung gibt, gilt $f^{-1}(A) = \{f^{-1}(x) | x \in A\}$

zu c) selbst in Übung

zu d) selbst in Übung ■

Definition
Sei $f\colon \Omega \to X$. Die Abbildung f heißt **($\mathcal{F}, \mathfrak{X}$)-messbar**, wenn gilt:

$$f^{-1}(A) \in \mathcal{F} \qquad (A \in \mathfrak{X})^{\dagger}$$

In diesem Fall heißt das Maß $\mu \circ f^{-1}$, definiert durch $\mu \circ f^{-1}(A) = \mu(f^{-1}(A))$ für alle $A \in \mathfrak{X}$, **Bildmaß** von μ bezüglich f.

Bemerkung

1.) Die Definition von $\mu \circ f^{-1}(A)$ ist sinnvoll, weil $f^{-1}(A) \in \mathcal{F}$ für alle $A \in \mathfrak{X}$.

2.) Es existiert eine formale Ähnlichkeit zwischen der Definition der Stetigkeit und der der Messbarkeit.

3.) Im Fall das Ω, X metrische Räume sind und $\mathcal{F}, \mathfrak{X}$ der σ-Algebra der Borelmengen entsprechen, folgt aus der Stetigkeit von f, dass f auch messbar ist. Um das zu zeigen, wähle man $\mathfrak{E} = \{A \subseteq X : A \text{ offen}\}$. Nach der Definition ist $\mathfrak{X} = \sigma(\mathfrak{E}) \Rightarrow f^{-1}(A) \in \mathcal{F}$ für ein $A \in \mathfrak{E}$. Somit ist klar, dass, wenn f stetig ist, folgt, dass $f^{-1}(A)$ offen ist und damit ist $f^{-1}(A) \in \mathcal{F}$. (siehe auch den unten stehenden Satz)

4.) Sei $f^{-1}\colon \mathfrak{P}(X) \to \mathfrak{P}(\Omega)$ mit $\mu\colon \mathcal{F} \subset \mathfrak{P}(\Omega) \to [0, +\infty]$
Die Definition von $\mu \circ f^{-1}$ ist *nur möglich*, wenn f als $(\mathcal{F}, \mathfrak{X})$-messbar vorausgesetzt wird.

5.) Ist $\mu \circ f^{-1}$ ein Maß auf $[X, \mathfrak{X}]$. Hierzu ist klar, dass $\mu \circ f^{-1}\colon \mathfrak{X} \to [0, +\infty]$. Es ist $\mu \circ f^{-1}(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$. Im folgenden ist die zweite Maßeigenschaft zu prüfen. Sei dazu $(A_i)_{i=1}^{\infty} \subseteq \mathfrak{X}, A_i \cap A_j = \emptyset, i \neq j \Rightarrow f^{-1}(A_i) \cap f^{-1}(A_j) = f^{-1}(\emptyset) = \emptyset \Rightarrow \mu \circ f^{-1}(\bigcup A_i) = \mu(f^{-1}(\bigcup A_i)) = \mu(\bigcup f^{-1}(A_i)) = \sum \mu(f^{-1}(A_i)) = \sum \mu \circ f^{-1}(A_i)$

Satz 2.1.2
Sei $f\colon \Omega \to X$. f ist $(\mathcal{F}, \mathfrak{X})$-messbar, wenn für ein $\mathfrak{E} \subseteq \mathfrak{P}(X)$ mit $\sigma(\mathfrak{E}) = \mathfrak{X}$ gilt:

$$f^{-1}(A) \in \mathcal{F} \qquad (A \in \mathfrak{E})$$

Beispiel

1. Die „Konstanten" sind messbar, d. h. wir haben $f\colon \Omega \to X$ mit $f(\omega) = c$. Dann ist $f^{-1}(A) = \begin{cases} \emptyset & c \notin A \\ \Omega & c \in A \end{cases}$. Da nun die leere Menge und auch Ω Element von $\mathcal{F}$ sind, gilt $f^{-1}(A) \in \mathcal{F}$. Damit ist die Funktion immer $(\mathcal{F}, \mathfrak{X})$-messbar.
2. Sei $\mathcal{F} = \{\emptyset, \Omega\}$ und $[X, \mathfrak{X}] = [\mathbb{R}, \mathfrak{B}]$. Zu zeigen ist, ob f $(\mathbb{R}, \mathfrak{B})$-messbar ist. Dies ist genau dann der Fall, wenn ein $c \in \mathbb{R}$ mit $f(\omega) = c$ existiert ...

†Schreibweise: $f\colon [\Omega, \mathcal{F}] \to [X, \mathfrak{X}]$

3. Sei $\mathcal{F} = \mathfrak{P}(\Omega)$. Dann ist jede Abbildung $f\colon \Omega \to X$ messbar. $\mathcal{F} = \mathfrak{P}(\Omega) \Rightarrow f^{-1}(A) \in \mathcal{F}$ $(A \subseteq X)$. Für beliebige $\mathfrak{X}$ ist f $(\mathfrak{P}(\Omega), \mathfrak{X})$-messbar.

Satz 2.1.3
Seien $f_1\colon [\Omega_1, \mathcal{F}_1] \to [\Omega_2, \mathcal{F}_2]$ und $f_2\colon [\Omega_2, \mathcal{F}_2] \to [\Omega_3, \mathcal{F}_3]$. Dann gilt $f_2 \circ f_1\colon [\Omega_1, \mathcal{F}_1] \to [\Omega_3, \mathcal{F}_3]$. Die Hintereinanderausführung messbarer Funktionen ist wieder messbar.

Beweis:
Im ersten Schritt zeigen wir: $(f_2 \circ f_1)^{-1}(A) = f_1^{-1}(f_2^{-1}(A))$. Sei dazu $\omega \in (f_2 \circ f_1)^{-1}(A) \Leftrightarrow (f_2 \circ f_1)(\omega) \in A \Leftrightarrow f_2(f_1(\omega)) \in A \Leftrightarrow f_1(\omega) \in f_2^{-1}(A) \Leftrightarrow \omega \in f_1^{-1}(f_2^{-1}(A))$. Dieses Ergebnis verwenden wir nun für den weiteren Beweis.

Sei $A \in \mathcal{F}_3$. Es ist nun zu zeigen, dass $(f_2 \circ f_1)^{-1}(A) \in \mathcal{F}_1$. Dazu stellen wir fest, dass $(f_2 \circ f_1)^{-1}(A) = f_1^{-1}(f_2^{-1}(A))$. Da aber f_2 messbar ist, folgt $f_2^{-1}(A) \in \mathcal{F}_2 \Rightarrow f_1^{-1}(f_2^{-1}(A)) \in \mathcal{F}_1$. Das ist aber nichts anderes als $(f_2 \circ f_1)^{-1}(A) \in \mathcal{F}_1$. ∎

Für die folgenden Aussagen gilt immer $[X, \mathfrak{X}] = [\mathbb{R}, \mathfrak{B}]$ und wir bezeichnen $(\mathcal{F}, \mathfrak{B})$-messbar als messbar.

Satz 2.1.4
Die Abbildung $f\colon \Omega \to \mathbb{R}$ ist genau dann messbar, wenn eine der folgenden Aussagen richtig ist (x, x_1, x_2 sind dabei immer rational):

(1) $\{\omega \in \Omega | f(\omega) < x\} \in \mathcal{F}$

(2) $\{\omega \in \Omega | f(\omega) \leq x\} \in \mathcal{F}$

(3) $\{\omega \in \Omega | f(\omega) > x\} \in \mathcal{F}$

(4) $\{\omega \in \Omega | f(\omega) \geq x\} \in \mathcal{F}$

(5) $\{\omega \in \Omega | x_1 \leq f(\omega) < x_2\} \in \mathcal{F}$

(6) $\{\omega \in \Omega | x_1 < f(\omega) < x_2\} \in \mathcal{F}$

(7) $\{\omega \in \Omega | x_1 \leq f(\omega) \leq x_2\} \in \mathcal{F}$

(8) $\{\omega \in \Omega | x_1 < f(\omega) < x_2\} \in \mathcal{F}$

Beweis:
Die ist eine Beweisidee für (1). Die weiteren Beweise sind analog zu führen: Nach dem Satz 1.2.5 wissen wir, dass die Menge $\{(-\infty, x) | x \in \mathbb{Q}\}$ ein durchschnittsabgeschlossenes Erzeugendensystem für $\mathfrak{B}$ ist. Wenn $\mathfrak{E}$ ein solches System mit $f^{-1}(B) \in \mathcal{F}$ ist, dann ist f messbar. Es gilt: $\{\omega \in \Omega | f(\omega) < x\} = f^{-1}((-\infty, x))$. Aus der Kombination beider Aussagen folgt der Satz. ∎

Satz 2.1.5
Seien $f, g\colon [\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}]$ messbare Funktionen. Dann sind die Funktionen $f + g$ und fg ebenfalls messbare Funktionen.

Beweis:
Seien $f, g\colon \Omega \to \mathbb{R}^2$. Diese Funktionen werden wie folgt umgewandelt: $h\colon = [f,g]\colon \Omega \to \mathbb{R}^2, [f,g](\omega) = [f(\omega), g(\omega)]$. Ist nun h messbar?

$$\begin{aligned} h^{-1}(A \times B) &= \{\omega \in \Omega | [f(\omega), g(\omega)] \in A \times B\} \\ &\Rightarrow f(\omega) \in A \wedge g(\omega) \in B \\ &= \{\omega \in \Omega | f(\omega) \in A\} \cap \{\omega \in \Omega | g(\omega) \in B\} \\ &= f^{-1}(A) \cap g^{-1}(B) \end{aligned}$$

Damit ist gezeigt, dass f, g messbar sind und es gilt, dass $f^{-1}(A) \in \mathcal{F} \wedge g^{-1}(B) \in \mathcal{F}$. Wegen der Abgeschlossenheit gegenüber Durchschnitten $((f^{-1}(A) \cap g^{-1}(B)) \in \mathcal{F})$ gilt für alle $A, B \in \mathfrak{B}\colon h^{-1}(A \times B) \in \mathcal{F}$. Und es folgt weiter $\sigma(\{A \times B | A \in \mathfrak{B}, B \in \mathfrak{B}\}) = \mathfrak{B}_2$. Damit ist h messbar.
Wir setzen nun $a(x_1, x_2) = x_1 + x_2, b(x_1, x_2) = x_1 x_2, a, b\colon \mathbb{R}^2 \to \mathbb{R}$, stetig. Also ist $f + g = a \circ h$ und $fg = b \circ h$. Da a, b stetig sind, sind sie auch messbar und nach Satz 2.1.3 ist die Verknüpfung beider Abbildungen wieder messbar. ■

Satz 2.1.6
Sei $(f_i)_{i \in I}\colon [\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}]$ eine abzählbare Familie messbarer Funktionen, so dass die Mengen $\{\omega \in \Omega | \sup f_i(\omega) < +\infty\} = \Omega$ bzw. $\{\omega \in \Omega | \inf f_i(\omega) < -\infty\} = \Omega$ sind. Dann ist $\sup f_i$ bzw. $\inf f_i$ messbar.‡

Beweis:
1. Fall Sei $f\colon = \sup f_i$. Dann ist f eine Abbildung von Ω nach $\mathbb{R}$. Wegen des Satz 2.1.4 genügt es zu zeigen, dass $\{\omega \in \Omega | f(\omega) \leq x\} \in \mathcal{F}$. Dazu $f(\omega) \leq x \Leftrightarrow f_i(\omega) \leq x \Rightarrow \{\omega \in \Omega | f(\omega) \leq x\} = \bigcap_{i \in I} \{\underbrace{\omega \in \Omega | f_i(\omega) \leq x}_{f_i^{-1}((-\infty, x)) \in \mathcal{F}}\}$. Damit folgt, dass f_i messbar ist.
Nach Satz 1.2.3 ist es gegenüber Vereinigung abgeschlossen und es folgt somit: $\{\omega \in \Omega\colon f(x) \leq x\} = \bigcap_{i \in I} f_i^{-1}((-\infty, x)) \in \mathcal{F}$

2. Fall Sei $f\colon = \inf f_1 \Rightarrow f\colon \Omega \to \mathbb{R}$. Wegen Satz 2.1.4 (4) genügt es zu zeigen, dass $\bigcap_{i \in I} \{\omega | f_i(\omega) \geq x\} = \{\omega \in \Omega | f(\omega) \geq x\} \in \mathcal{F}$. Dazu $f(\omega) \geq x \Leftrightarrow f_i(\omega) \geq x (i \in I) \Rightarrow \{\omega \in \Omega | f(\omega) \geq x\} = \bigcup_{i \in I} \{\omega \in \Omega | f_i(\omega) \geq x\} = \bigcup_{i \in I} \{f_i^{-1}([x, +\infty)) \in \mathfrak{B}\} \Rightarrow \{\omega \in \Omega | \inf_i f_i(\omega) \geq x\} \in \mathfrak{B}$ für $x \in \mathbb{R}$. Dann folgt nach Satz 2.1.4 Punkt 4, dass $\inf_i f_i$ messbar ist. ■

Satz 2.1.7
Sei f eine messbare Funktion ($f\colon [\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}]$). Dann gelten:

(1) $af + b$ ist messbar. $(a, b \in \mathbb{R})$

(2) $f^+\colon = \max(f, 0)$ ist messbar und der positive Ast der Funktion.

(3) $f^-\colon = \min(f, 0)$ ist messbar und der negative Ast der Funktion.

‡Erläuterung: $(\sup f_i)(\omega)\colon = \sup f_i(\omega)$ mit $\omega \in \Omega$

(4) $|f|$ ist messbar.

Beweis:

zu (1) Wir wissen, dass „konstante“ Funktionen immer messbar sind. Damit wie auch nach Satz 2.1.5 gilt die Behauptung.

zu (2) Wir setzen $f_1 = f, f_2 = 0$. Dann folgt, dass $f^+ = \max(f_1, f_2)$. Beide Funktionen sind messbar und wegen des obigen Satzes ist auch $\sup f_i = f^+$ messbar.

zu (3) Beweis erfolgt analog zu (2).

zu (4) $|f| = f^+ - f^- = f^+ + (-1)f^- \Rightarrow$ messbar. ∎

Satz 2.1.8
Sei $(f_n)_{n=1}^{\infty}$ eine Folge messbarer Funktionen und $\limsup f_n$ oder $\liminf f_n$ existieren als reelle Funktionen. Dann sind diese messbar. Falls $\lim f_n$ existiert, so ist auch er messbar.

Beweis:
selbst ∎

Satz 2.1.9
Sei $A \subseteq \Omega$. Die Indikatorfunktion $\chi_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$ ist genau dann messbar, wenn $A \in \mathcal{F}$.

Beweis:
Es ist $\chi_A^{-1}(B)$ zu prüfen:

$$\chi_A^{-1}(B) = \begin{cases} \emptyset & 0, 1 \notin B \\ A & 1 \in B \wedge 0 \notin B \\ A^c & 1 \notin B \wedge 0 \in B \\ \Omega & 1 \in B \wedge 0 \in B \end{cases}$$

Für die Gleichheit müssen wir beide Richtungen zeigen: Sei χ_A messbar. Dann ist $\{1\} \in \mathfrak{B} \Rightarrow \chi_A^{-1}(\{1\}) = A \in \mathcal{F}$.
Sei $A \in \mathcal{F}$. Dann ist zu zeigen, dass χ_A messbar ist, d. h. $\chi_A^{-1}(B) \in \mathcal{F}$ für $B \in \mathfrak{B}$. Es folgt, dass $\chi_A^{-1}(B) \in \{\emptyset, \Omega, A, A^c\}$ und somit genügt es zu zeigen, dass $\{\emptyset, \Omega, A, A^c\} \subseteq \mathcal{F}$ ist. Wir wissen, dass $\emptyset, \Omega \in \mathcal{F}$ sind. Weiterhin ist nach Voraussetzung $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ und es folgt die Behauptung. ∎

Beispiel
$\Omega = \mathbb{R}, A = \{x \in \mathbb{R} | x \text{ rational}\} \Rightarrow A$ abzählbar. Da alle abzählbaren Mengen in der Borelmenge liegen.

Definition
Sei $f\colon \Omega \to \mathbb{R}$. Die Funktion f heißt **einfach**, wenn existieren: $n \geq 1, a_1, \ldots, a_n \in \mathbb{R}, A_1, \ldots, A_n \in \mathcal{F}$ mit

$$f = \sum_{i=1}^{n} a_i \chi_{A_i}$$

Bemerkung

1. Das Wort „einfach“ bezieht sich auf die gegebenen σ-Algebren $\mathcal{F}$ und $\mathfrak{B}$.
2. Einfache Funktionen sind messbar. Denn nach dem vorhergehenden Satz folgt, dass χ_{A_i} messbar sind und die Linearkombinationen messbarer Funktionen sind wieder messbar. (Zum Beweis siehe Satz 2.1.5 und Satz 2.1.9).

Satz 2.1.10
Sei $f\colon \Omega \to \mathbb{R}$. Dann gelten

(1) Wenn f einfach ist, dann existieren $n \geq 1, b_1, \dots, b_n \in \mathbb{R}, B_1, \dots, B_n \in \mathcal{F}$ mit $b_i \neq b_j (i \neq j), B_i \cap B_j = \emptyset, \bigcup_{i=1}^n B_i = \Omega$[§], $f = \sum_{i=1}^n b_i \chi_{B_i}$. Diese Darstellung von f ist bis auf Permutation des Indexbereichs eindeutig bestimmt.

(2) f ist genau dann einfach, wenn f messbar ist und die Anzahl des Wertebereichs [¶] endlich ist.

Beweis:
zu (1) Der Beweis hierzu ist im untenstehenden Beweis enthalten.

zu (2) „$\Rightarrow$“ Wegen der Bemerkung und der Definition folgt aus einfach auch die Messbarkeit. Auch die zweite Bedingung geht sofort aus der Definition hervor.

„$\Leftarrow$“ Setzen $\{b_1, \dots, b_n\} = \{f(\omega) | \omega \in \Omega\}$. Die b_i sind paarweise verschieden. Weiter setzen wir $B_i = f^{-1}(\{b_i\})$ für $i = 1, \dots, n$. Es gilt, dass alle B_i paarweise disjunkt sind und das deren Vereinigung Ω ergibt. Damit folgt $\{b_i\} \in \mathfrak{B}$. Wegen der Messbarkeit gilt weiter $B_i = f^{-1}(\{b_i\}) \in \mathcal{F}$. Es ist trivial, dass $B_i \cap B_j = \emptyset$ und klar, dass $f = \sum_{i=1}^n b_i \chi_{B_i}$. Daraus folgt, dass f einfach ist. ■

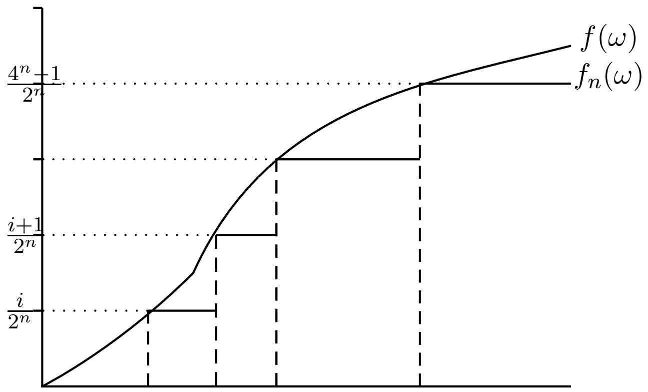
Satz 2.1.11
Sei f messbar und nichtnegativ[‖]. Dann existiert eine Folge nichtnegativer einfacher Abbildungen $(f_n)_{n=1}^\infty$ mit $f_n \leq f_{n+1}$ (monoton wachsend) und $f = \lim_{n\to\infty} f_n (f_n \uparrow f)$.

Beweis:
Der Beweis erfolgt konstruktiv, d. h. man konstruiert eine spezielle Folge $(f_n)_{n=1}^\infty$ mit den Eigenschaften aus dem Satz:

$$f_n \colon = \sum_{i=0}^{4^n-1} \frac{i}{2^n} \chi_{f^{-1}([\frac{i}{2^n}, \frac{i+1}{2^n}))} + 2^n \chi_{f^{-1}([2^n, +\infty))}$$

$$f_n(\omega) \colon = \begin{cases} i/2^n & i/2^n \leq f(\omega) < \frac{i+1}{2^n} \\ 2^n & f(\omega) \geq 2^n \end{cases} \qquad i = 0, \dots, 4^n - 1$$

[§] B_i sind messbare Zerlegungen von Ω.
[¶] $\#\{f(\omega) | \omega \in \Omega\} < +\infty$
[‖] $f = f^+$



Damit ergibt sich $0 \leq f_n \leq f_{n+1} \leq f$. Wenn f messbar ist, folgt, dass f auch einfach ist. Die Differenz $f(\omega) - f_n(\omega) \leq \frac{1}{2^n}$** und es folgt, dass $\lim f_n(\omega) = f(\omega)$ ■

2.2. Einführung des Integrals

Im folgenden ist immer $[\Omega, \mathcal{F}, \mu]$ ein Maßraum und wir betrachten $f\colon [\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}]$. Das Integral $\int f(\omega)\mu(d\omega) \hat{=} \int f d\mu$ wird in drei Stufen eingeführt:

Definition (1. Stufe)
Sei $f \geq 0$ und einfach, d. h. wir haben eine Darstellung der Form $f = \sum_{i=1}^n a_i \chi_{A_i}$ mit $n = 1, a_1, \ldots, a_n \geq 0, A_i \in \mathcal{F}$. Dann setzen wir

$$\int f(\omega)\mu(d\omega) \colon = \sum_{i=1}^n a_i \mu(A_i)^{\dagger\dagger}$$

Bemerkung

1. $\int f d\mu \geq 0$

2. Die Definition des Integrals ist eindeutig, d. h. unabhängig von der betrachteten Darstellung als Linearkombination von messbaren Indikatorfunktionen:
$f = \sum_{i=1}^n a_i \chi_{A_i} = \sum_{k=1}^m b_k \chi_{B_k} \Rightarrow \sum_{i=1}^n a_i \mu(A_i) = \sum_{k=1}^m b_k \mu(B_k)$

3. Das Integral ist linear, d. h. $f, g \geq 0$ und einfach, $\alpha, \beta \geq 0 \Rightarrow \int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu$. Denn sei $f = \sum a_i \chi_{A_i}$ und $g = \sum b_j \chi_{B_j}$. Dann ist $\alpha f + \beta g = \sum \alpha a_i \chi_{A_i} + \sum \beta b_j \chi_{B_j} \Rightarrow \int (\alpha f + \beta g) d\mu = \sum \alpha a_i \mu(A_i) + \sum \beta b_j \mu(B_j) = \alpha \int f d\mu + \beta \int g d\mu$

** $\omega \in \Omega$ mit $f(\omega) \leq 2^n$
†† Vereinbarung: Es kann sein, dass $\mu = \infty$ und es kann $0 \cdot \infty$ auftauchen. Dann setzen wir $0 \cdot \infty = 0$

4. Wenn $f, g \geq 0$ einfach und $f \leq g$, dann folgt, dass $\int f d\mu \leq \int g d\mu$. Das Integral ist monoton. Denn da $f \leq g$ ist, ist auch $g - f \geq 0$ und da beide Funktionen einfach sind, ist auch die Differenz wieder einfach. Es gilt: $g = f + (g - f) \Rightarrow \int g d\mu = \int f d\mu + \underbrace{\int (g - f) d\mu}_{\geq 0} \Rightarrow \int g d\mu \geq \int f d\mu$

5. Sei $(f_n)_{n=1}^{\infty}$ eine Folge nichtnegativer einfacher Funktionen mit $f_n \leq f_{n+1}$. Dann existiert $\lim_{n\to\infty} \int f_n d\mu$. Dabei ist auch $+\infty$ möglich. Aus dem vierten Punkt folgt $0 \leq \int f_n d\mu \leq \int f_{n+1} d\mu$

6. Seien $(f_n), (g_n)$ Folgen nichtnegativer einfacher Funktionen mit $f_n \leq f_{n+1}, g_n \leq g_{n+1}, \lim f_n \leq \lim g_n$. Dann gilt $\lim \int f_n d\mu \leq \lim \int g_n d\mu$. Zum Beweis muss die zusätzliche Voraussetzung $f_n \leq g_n$ benutzt werden. Denn nach dem vierten Punkt folgt: $\int f_n d\mu \leq \int g_n d\mu$ und wegen fünftens existieren $\lim \int f_n d\mu, \lim \int g_n d\mu$. Damit ist $\lim \int f_n d\mu \leq \lim \int g_n d\mu$.

Definition (2. Stufe)
Sei $f \geq 0$ und messbar. Nach Satz 2.1.11 existieren $(f_n)_{n=1}^{\infty}, 0 \leq f_n \leq f_{n+1}, f = \lim f_n, f_n$ einfach. Wir setzen:

$$\int f d\mu :\, = \lim_{n\to\infty} \int f_n d\mu$$

Bemerkung

1. Die Definition ist „sinnvoll“, weil das Integral nach der Definition erster Stufe definiert ist und der Limes nach der obigen Bemerkung 5 existiert.

2. Die Definition ist auch eindeutig, d. h. wenn wir haben: $0 \leq f_n \leq f_{n+1}, 0 \leq g_n \leq g_{n+1}, f = \lim f_n = \lim g_n$ und beide sind einfach, dann folgt daraus: $\lim \int f_n d\mu = \lim \int g_n d\mu$. Dies ist eine direkte Folgerung aus der Bemerkung 6 oben.

3. Die Definition zweiter Stufe ist eine Erweiterung der Definition erster Stufe. Denn sei $f \geq 0$ einfach. Dann setzen wir $f_n :\, = f$ mit $n = 1, 2, \ldots \Rightarrow 0 \leq f_n \leq f_{n+1}$. Die f_n seien messbar. Dann folgt nach der Definition zweiter Stufe: $\int f\, d\mu = \lim \int f_n\, d\mu = \int f\, d\mu$.

4. Das so definierte Integral ist linear.
Beweis: Sei $\alpha, \beta \in \mathbb{R} \geq 0, f, g \geq 0$ und messbar. Weiter seien $f = \lim f_n, g = \lim g_n, 0 \leq f_n \leq f_{n+1}, 0 \leq g_n \leq g_{n+1}$ alle einfach, d. h. $\int f d\mu = \lim_{n\to\infty} \int f_n d\mu$ und $\int g d\mu = \lim_{n\to\infty} \int g_n d\mu$.
$\alpha f + \beta g = \lim_{n\to\infty}(\alpha f_n + \beta g_n) \Rightarrow \int (\alpha f + \beta g) d\mu = \lim_{n\to\infty} \int (\alpha f_n + \beta g_n) d\mu = \lim_{n\to\infty}(\alpha \int f_n d\mu + \beta \int g_n d\mu) = \alpha \int f d\mu + \beta \int g d\mu$

5. Für $0 \leq f \leq g$ gilt: $0 \leq \int f d\mu \leq \int g d\mu$, denn $g = f + (g - f) \Rightarrow \int g d\mu = \int f d\mu + \int (g - f) d\mu \Rightarrow \int g d\mu \geq \int f d\mu$. Dies folgt aus $\int f\, d\mu \geq 0$ für die erste Stufe.

6. Monotonie: $0 \leq f \leq g \Rightarrow \int f\, d\mu \leq \int g\, d\mu$. Siehe hierzu die vierte Bemerkung bei der Definition der ersten Stufe.

Definition (3. Stufe)
Sei f messbar und bezüglich μ integrierbar, d. h. $\int f^+ d\mu$ und $\int(-f^-)d\mu < +\infty$. Dann gilt:

$$\int f d\mu :=\int f^+ d\mu - \int (-f^-) d\mu$$

Bemerkung

1. klar: $|f| = f^+ + (-f^-) \Rightarrow \int |f| d\mu = \int f^+ d\mu + \int (-f^-) d\mu$

$$\boxed{f \text{ ist genau dann } \mu\text{-integrierbar, wenn } \int |f| d\mu < \infty}$$

2. $\int f d\mu$ für $f \geq 0$ und messbar existiert immer. Es ist allerdings der Fall $+\infty$ möglich. Das Integral $\int f d\mu$ im Falle f messbar und nicht notwendigerweise ≥ 0 ist nach der Definition dritter Stufe nur für μ-integrierbare Funktionen erklärt. Ohne die Bedingung $\int f\, d\mu < +\infty$ ist das Integral nicht erklärt.

3. Die Eigenschaften Monotonie und Linearität des Integrals übertragen sich von der zweiten zur dritten Stufe. Für den Beweis ist zu zeigen, dass für f, g μ-integrierbar und für $a, b \in \mathbb{R}$ gilt, dass $af + bg$ wieder μ-integrierbar sind und dass $\int (af+bg) d\mu = a \int f d\mu + b \int g d\mu$ ist. Da f, g messbar sind, folgt dass auch $af + bg$ messbar sind. Somit ist $af + bg$ natürlich auch μ-integrierbar. Es gilt $\int (af + bg) d\mu \leq \int (|a||f| + |b||g|) d\mu$ wegen der Monotonie zweiter Stufe. Nach der Linearität zweiter Stufe ist das aber gleich $|a| \int |f| d\mu + |b| \int |g| d\mu$. Sowohl $|f|$ wie auch $|g|$ sind kleiner $+\infty$. Somit ist auch: $|a| \int |f| d\mu + |b| \int |g| d\mu < +\infty$.

Satz 2.2.1 (Integral bezüglich diskreter Maße)
Seien $(\alpha_i)_{i \in I}$ eine abzählbare Familie nichtnegativer reeller Zahlen, $(\omega_i)_{i\in I} \subseteq \Omega$ und $\mu = \sum_{i\in I} \alpha_i \delta_{\omega_i}$[‡‡]. Eine messbare Funktion f (d. h. $f \colon [\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}]$) ist genau dann **$\mu$-integrierbar**, wenn gilt $\sum_{i \in I} \alpha_i |f(\omega_i)| < +\infty$. In diesem Fall gilt:

$$\int f d\mu = \sum_{i \in I} \alpha_i f(\omega_i)$$

Beweis:

1. Fall Sei $f = \chi_A$ mit $(A \in \mathcal{F})$. Dann folgt nach der Definition des Integrals erster Stufe, dass für $g = \sum a_i \chi_{A_i}$ (einfach und größergleich 0) gilt, $\int g\, d\mu_i = \sum a_i \mu(A_i)$. Damit folgt, $\int f\, d\mu = \int \chi_A\, d\mu = \mu(A) = \sum_{i \in I} \alpha_i \delta_{\omega_i}(A) = \sum_{i\in I} \alpha_i \chi_A(\omega_i) = \sum_{i \in I} \alpha_i f(\omega_i)$

2. Fall Sei $f = \sum_{k=1}^n a_k \chi_{A_k}$ mit $a_k \geq 0, A_k \in \mathcal{F}$. Dann ist $f \geq 0$, einfach und $\int f d\mu = \sum_{k=1}^n a_k \mu(A_k) = \int \sum_{k=1}^n a_k \chi_{A_k} d\mu = \sum_{k=1}^n a_k \int \chi_{A_k} d\mu = \sum_{k=1}^n a_k \sum_{i\in I} \alpha_i \chi_{A_k}(\omega_i) = \sum_{i\in I} \alpha_i \underbrace{\sum_{k=1}^n a_k \chi_{A_k}(\omega_i)}_{f(\omega_i)} = \sum_{i \in I} \alpha_i f(\omega_i)$

[‡‡] Maße von diesem Typ heißen **diskrete Maße**

3. Fall Seien $f \geq 0$ und messbar, $0 \leq (f_m)_{m=1}^{\infty} \uparrow f$ einfach. Nach der Definition zweiter Stufe folgt nun, dass $\int f \, d\mu = \lim_{n\to\infty} \int f_m \, d\mu$. Nach dem vorigen Fall folgt: $\lim_{n\to\infty} \sum_{i\in I} \alpha_i f_m(\omega_i)$. Da die Funktion monoton wachsend ist, kann der Limes und die Summe vertauscht werden: $\sum_{i\in I} \alpha_i \lim_{n\to\infty} f_m(\omega_i) = \sum_{i\in I} \alpha_i f(\omega_i)$.

4. Fall Sei f messbar und gelte $f = f^+ - f^-$. Wir wissen nach der Definition des Integrals zweiter Stufe:

$$\int f^+ \, d\mu = \int\limits_{i\in I} \alpha_i f^+(\omega_i)$$
$$\int f^- \, d\mu = \int\limits_{i\in I} \alpha_i f^-(\omega_i)$$

Weiterhin wissen wir, dass f genau dann μ-integrierbar ist, wenn $\int |f| \, d\mu < \infty$. Damit sind auch die Integrale der beiden Teilfunktionen kleiner als unendlich. Wenn f μ-integrierbar ist, dann folgt aus der Definition dritter Stufe:

$$\int f \, d\mu = \int f^+ - (-f^-) d\mu = \sum_{i\in I} \alpha_i f^+(\omega_i) - \sum_{i\in I} \alpha_i (-f^-)(\omega_i)$$
$$= \sum_{i\in I} \alpha_i \big(f^+(\omega_i) - (-f^-)(\omega_i)\big) = \sum_{i\in I} \alpha_i \big(f^+(\omega_i) + f^-(\omega_i)\big) = \sum_{i\in I} \alpha_i f(\omega_i) \quad \blacksquare$$

Definition
Der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mu)$ heißt **diskret**, wenn Ω abzählbar, $\mathcal{F} = \mathfrak{P}(\Omega)$ und μ diskret ist.

Beispiel

1. Sei $\mu = \delta_{\omega_0}$ und ein festes $\omega_0 \in \Omega$ gegeben. Dann ist jede messbare Funktion f μ-integrierbar und es gilt: $\int f d\mu = \int f d\delta_{\omega_0} = f(\omega_0)$
2. Wir betrachten die Reihe $\sum_{k=0}^{m} a_k$. Dabei ist auch der Fall $m = \infty$ möglich. Nun wählen $\Omega = \{0, 1, \ldots, n\}, \mathcal{F} = \mathfrak{P}(\Omega), \mu = \sum_{i=0}^{n} \delta_i$ und das Zählmaß $\mu(A) = \#A$. Setzen $f\colon \Omega \to \mathbb{R}$ mit $f(k)\colon = a_k$ für $k \in \Omega$. Damit folgt, $\int f d\mu = \sum_{k=0}^{m} a_k$ und f ist bezüglich μ genau dann integrierbar, wenn $\sum_{k=0}^{m} |a_k| < \infty$. Für den Fall $m = \infty$ muss die Reihe absolut konvergent sein.

Satz 2.2.2 (Riemannintegral stetiger Funktionen)
Seien $f\colon \mathbb{R} \to \mathbb{R}$ stetig, f stetig mit Stammfunktion F, $a, b \in \mathbb{R}$ und $a < b$. Dann ist $f\chi_{[a,b]}$ bezüglich ℓ integrierbar und es gilt:

$$\int f\chi_{[a,b]} \, d\ell = \int\limits_a^b f(x) \, dx = F(b) - F(a)$$

Beweis:
Der Fall, dass die Funktion einfach ist, wird nicht betrachtet. Denn alle stetigen Funktionen (außer Konstanten) sind nicht einfach.

Sei nun $f \geq 0$ und stetig. Dann ist f auch messbar und $f\chi_{[a,b]}$ ebenfalls. Wir konstruieren einen Funktionenfolge $(f)_{n=1}^{\infty}$. Nach der Definition zweiter Stufe wissen wir, dass $\int f\chi_{[a,b]}d\ell = \lim_{n\to\infty} \int f_n d\ell$, falls $0 \leq f_n \uparrow f$ und f_n einfach. Im folgenden wird f_n als untere Darbouxsche Summe konstruiert. Dabei wird das Intervall $[a,b]$ in Intervalle $z_k^{(n)}$ zerlegt.

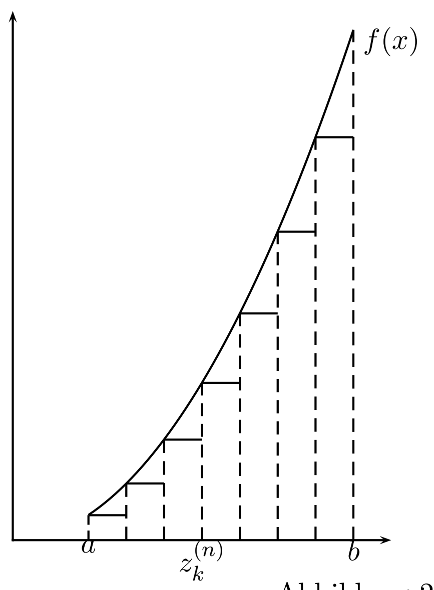


Abbildung 2.1.: Untere Darbouxsche Summen

todo: Ergänzung von Carolin einfügen

$$\begin{aligned}
z_{k+1}^{(n)} - z_k^{(n)} &= \frac{b-a}{2^n} f_n(x) = 0 \quad (x \notin [a,b]) \\
b_k^{(n)} \colon &= \inf_{z_k^{(n)} \le x < z_{l+1}^{(n)}} \\
f_n(x) &= \sum_k b_k^{(n)} \chi_{[z_k^{(n)}, z_{k+1}^{(n)})}(x) \\
&\Rightarrow f_n \ge 0, f_n \le f_{n+1}, \lim f_n = f\chi_{[a,b]}, f_n \text{ einfach} \\
\Rightarrow \int f\chi_{[a,b]}\, d\ell &= \lim_{n\to\infty} \int f_n \, d\ell \\
&= \lim_{n\to\infty} \sum_k b_k^{(n)} \underbrace{\ell([z_k^{(n)}, z_{k+1}^{(n)}))}_{\frac{b-a}{2^n}} \\
&= \lim_{n\to\infty} \sum_k b_k^{(n)} \frac{b-a}{2^n} = \int\limits_a^b f(x)\, dx = F(b) - F(a)
\end{aligned}$$

Damit ist die Behauptung für $f \ge 0$ gültig. Sei nun f „beliebig“ (insbesondere aber stetig, beachte die Voraussetzungen). Damit sind $f^+, (-f^-)$ stetig und das Integral $\int f^+\chi_{[a,b]}\, d\ell = \int (f\chi_{[a,b]})^+ = \int_a^b f^+(x)\, dx$. Selbiges gilt natürlich für $(-f^-)$. Es ist klar, dass $\int_a^b f^+(x)\, dx + \int_a^b (-f^-)(x)\, dx = \int_a^b f(x)\, dx$. Dabei wurde die Linearität des Riemannintegrals ausgenutzt. Wegen der Linearität des Lebegueintegrals folgt auch $\int f^+\chi_{[a,b]}\, d\ell - \int (-f^-)\chi_{[a,b]}\, d\ell = \int f\chi_{[a,b]}\, d\ell$. Aus den obigen Aussagen folgt schließlich die Behauptung.

todo: untenstehendes kann evtl. entfernt werden Es ist bekannt, dass das Riemannsche Integral der Limes der Darbouxschen Summen ist. Somit haben wir $\lim_{n\to\infty} \int f_n\, d\ell = \int_a^b f(x)\, dx$. Weiter ist klar, dass $f_n \le f_{n+1}$. Denn die Feinheit wird verkleinert, wenn die Intervalle vergrößert werden. Damit strebt f_n für $n \to \infty$ nach $f\chi_{[a,b]}$ und nach der Definition zweiter Stufe folgt, $\int f_n\, d\ell \xrightarrow{n\to\infty} \int f\chi_{[a,b]}\, d\ell$ ■

Bemerkung

1) Die Behauptung gilt auch dann, wenn nur vorausgesetzt wird, dass f auf $[a,b]$ riemann-integrierbar ist.

2) Der Satz 2.2.2 ist ausdehnbar auf die Fälle $a = -\infty, b = \infty$. Dabei ist die Forderung, dass f auf $[a,b]$ ℓ-integrierbar ist, notwendig. Das heißt nun aber $\int_a^b |f(x)|\, dx < +\infty$, d. h. $\int_a^b |f(x)|\, dx < +\infty$.

Satz 2.2.3 (Übertragungssatz)
Seien $[\Omega_1, \mathcal{F}_1, \mu]$ ein Maßraum, $[\Omega_2, \mathcal{F}_2]$ ein messbarer Raum und weiter $h\colon [\Omega_1, \mathcal{F}_1] \to [\Omega_2, \mathcal{F}_2], f\colon [\Omega_2, \mathcal{F}_2] \to [\mathbb{R}, \mathfrak{B}]$, wobei f bezüglich $\mu \circ h^{-1}$ integrierbar ist. Dann ist $f \circ$

$h\colon \Omega_1 \to \mathbb{R}$ bezüglich μ integrierbar und es gilt:

$$\int f \circ h \, d\mu = \int f \, d\mu \circ h^{-1} \tag{2.1}$$

Beweis:

0. Fall Sei $f = \chi_A$ mit $(A \in \mathcal{F}_2)$ und damit messbar. Wir betrachten die Verknüpfung $f \circ h(\omega) = \chi_A(h(\omega))$. Es ist klar, dass $\chi_A \circ h$ eine Indikatorfunktion auf Ω_1 sein muss, d. h. $\exists B \subseteq \Omega_1$ mit $\chi_A \circ h = \chi_B$ und $B = \{\omega \in \Omega_1 \colon \chi_A(h(\omega)) = 1\}$.
$\chi_B(\omega) = 1 \Leftrightarrow \chi_A(h(\omega)) = 1 \Leftrightarrow h(\omega) \in A \Leftrightarrow w \in h^{-1}(A) \Rightarrow B = h^{-1}(A) \in \mathcal{F}_1$.
Dabei ist $B = \{\omega \in \Omega_1 \colon \chi_A(h(\omega)) = 1\}$.

$$\Rightarrow \chi_A \circ h = \chi_{h^{-1}(A)}$$

$\int f \circ h \, d\mu = \int \chi_{h^{-1}(A)} \, d\mu = \mu(h^{-1}(A)) = \mu \circ h^{-1}(A) = \int \chi_A \, d\mu \circ h^{-1} = \int f \, d\mu \circ h^{-1}$

1. Fall Sei $f \geq 0$ und einfach, d. h. $f = \sum_{i=1}^{n} a_i \chi_{A_i}$ mit $A_i \in \mathcal{F}_2, a_i \geq 0$. Das Integral $\int f \, d\mu \circ h^{-1}$ entspricht wegen der Linearität der Summe $\sum_{i=1}^{n} a_i \int \chi_{A_i} \, d\mu \circ h^{-1} = \sum_{i=1}^{n} a_i \int \chi_{A_i} \circ h^{-1} \, d\mu = \int \sum_{i=1}^{n} a_i \chi_{A_i} \circ h^{-1} \, d\mu = \int f \circ h \, d\mu$

2. Fall Wir wählen $(f_n)_{n=1}^{\infty}$ mit $0 \leq f_n \leq f_{n+1}, f = \lim f_n$ und f_n einfach. Somit folgt $\int f_n d\mu \circ h^{-1} = \lim \int f_n d\mu \circ h^{-1} = \lim \int f_n \circ h d\mu$. Es ist klar, dass $g_n \colon = f_n \circ h$ die folgenden Bedingungen erfüllt: $0 \leq g_n \leq g_{n+1}, \lim g_n = f \circ h, g_n$ einfach.[§§]
$\Rightarrow \lim_{n \to \infty} \int g_n d\mu = \int f \circ h d\mu \Rightarrow \int f \circ h d\mu = \int f d\mu \circ h^{-1}$

Die Behauptung, dass $f \circ h$ μ-integrierbar ist, folgt sofort aus der $\mu \circ h^{-1}$-Integrierbarkeit von f und der Gleichung (Gleichung 2.1) im bewiesenen Fall nichtnegativer Funktionen. ■

Beispiel
Sei $\Omega_2 = \mathbb{R}, h\colon \Omega_1 \to \mathbb{R}, f(x) = x, f\colon \mathbb{R} \to \mathbb{R}$. Dann gilt $\int h(\omega)\mu(d\omega) = \int x \mu \circ h^{-1}(dx)$.

Definition
Seien f und g messbare Funktionen. f und g sind ***μ*-fast-überall-gleich**, wenn gilt:

$$\mu(\{\omega \in \Omega | f(\omega) \neq g(\omega)\}) = 0$$

Beispiel

1. Sei $\mu = \delta_{\omega_0} \Rightarrow f = g \; \mu -$ fast-überall $\Leftrightarrow f(\omega_0) = g(\omega_0)$ für $\omega_0 \in \Omega$. Denn sei $f(\omega) \neq g(\omega), \omega \neq \omega_0 \Rightarrow \mu(\{\omega \in \Omega | f(\omega) \neq g(\omega)\}) = \mu(\{\omega \in \Omega | f(\omega_0) \neq g(\omega_0)\}) \Rightarrow f =_{\delta_{\omega_0}} g \Leftrightarrow f(\omega_0) = g(\omega_0)$.

2. $[\Omega, \mathcal{F}, \mu] = [\mathbb{R}, \mathfrak{B}, \ell], f = \chi_{\mathbb{Q}}$ Da die rationalen Zahlen abzählbar sind, ist klar, dass $\ell(\mathbb{Q}) = 0$. Somit ist auch $\chi_{\mathbb{Q}} \equiv 0$, da $\ell(\{\omega \in \Omega | \chi_{\mathbb{Q}}(\omega) \neq 0\}) = \ell(\mathbb{Q})$. Allgemeiner gilt: $f =_\ell g \Leftrightarrow f\chi_{A^c} =_\ell g\chi_{A^c}$.

[§§] Wenn f_n nur endliche viele Werte annimmt, nimmt g_n ebenfalls nur endlich viele Werte an.

Bemerkung
Die Fast-Überall-Gleichheit ist eine Äquivalenzrelation. Die Klasse aller messbaren Funktionen zerfällt in von μ abhängigen Äquivalenzklassen.

Satz 2.2.4
Seien f, g μ-fast überall gleich ($f =_\mu g$). Weiter seien f, g μ-integrierbar. Dann gilt:

$$\int f d\mu = \int g d\mu$$

Beweis:
$f =_\mu g \Leftrightarrow f - g =_\mu 0 =_\mu h$ durch „Reduktion" der Allgemeinheit

- Wegen $\int (f-g)d\mu = \int f d\mu - \int g d\mu$ genügt es zu zeigen, dass $h = 0$ messbar und aus $h =_\mu 0$ folgt $\int h d\mu = 0$.
- Es genügt, den Fall $h \geq 0$ zu zeigen. Denn wir haben $h = h^+ + h^-.h^+, h^- \geq 0, \int h\,d\mu = \int h^+\,d\mu - \int(-h^-)\,d\mu$ und $0 = \mu(\{\omega \in \Omega | h(\omega) \neq =\}) \geq \mu(\{\omega \in \Omega | h^+(\omega) \neq 0\}) \geq 0$. Das gilt analog für h^-.
- Es reicht weiter, den Fall h einfach und größer 0 zu betrachten. Denn sei $\int h\,d\mu = \lim \int f_n\,d\mu$ mit $0 \leq f_n \leq f_{n+1}, \lim f_n = h$. Dann wissen wir nach dem obigen Punkt: $0 = \mu(\{\omega \in \Omega | h(\omega) \neq 0\}) = \mu(\{\omega \in \Omega | h(\omega) > 0\}) \geq \mu(\{\omega \in \Omega | f_n(\omega) > 0\}) \geq 0$.
- Das Problem kann weiter auf die Betrachtung von $h = \chi_A$ mit $A \in \mathfrak{P}$ reduziert werden. Denn sei h wie in der obigen Feststellung einfach. Dann gilt: $h = \sum_{j=1}^m a_j \chi_{A_j}$ für $a_j > 0$ und für alle $j = 1, \ldots, m$ ist $h \geq a_j \chi_{A_j}$. Weiter ist $0 = \mu(\{\omega \in \Omega | h(\omega) > 0\}) \geq \mu(\{\omega \in \Omega | \chi_{A_j}(\omega) > 0\})$. Wir haben also: $\int h\,d\mu = \sum a_j \int \chi_{A_j}\,d\mu$.
  - $\int \chi_{A_i} d\mu = \mu(A_i) = 0 \Rightarrow \int f d\mu = 0$
  - $f \geq a_i \chi_{A_i} \Rightarrow$ genügt zu zeigen: $\chi_A =_\mu 0 \Rightarrow \int \chi_A d\mu = 0$

  Dazu ist $\chi_A =_\mu 0 \Leftrightarrow \mu(\{\underbrace{\omega \in \Omega | \chi_A(\omega) \neq 0}_{A}\}) = 0 = \mu(A)$

  $\int \chi_A d\mu = \mu(A) = 0$ ∎

2.3. Absolute Stetigkeit von Maßen

Definition
Seien μ, ν Maße auf $[\Omega, \mathcal{F}]$ und $0 \leq f \colon [\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}]$. Dann heißt ν **absolut stetig** bezüglich μ mit Dichte f[¶¶], wenn gilt:

$$\nu(A) = \int f \chi_A d\mu \quad (A \in \mathcal{F})$$

[¶¶]Schreibweise: $\nu = \mu f$

Bemerkung

0. Schreibweisen: $\int_A f d\mu :=\int f\chi_A\, d\mu, \nu \ll \mu$. In der Literatur ist auch die folgende Definition gebräuchlich: Seien μ und ν Maße auf einer σ-Algebra $\mathcal{F}$. Dann heißt μ absolut stetig bezüglich ν, wenn für alle $A \in \mathcal{F}$ gilt:
$$\nu(A) = 0 \Rightarrow \mu(A) = 0$$

1. $\mu f(A) = \int f\chi_A\, d\mu$ definiert ein Maß auf $[\Omega, \mathcal{F}]$
   Beweis:
   - Fall: $A = \emptyset$: $\mu f(\emptyset) = \int f\chi_\emptyset d\mu = \int 0 d\mu = 0$
   - klar: $\mu f(A) = \int f\chi_A d\mu \geq 0$
   - Die (A_i) sind paarweise disjunkt: $A_i \in \mathcal{F} \Rightarrow \mu f(\bigcup A_i) = \int f\chi_{\bigcup A_i} d\mu = \int \sum f\chi_{A_i} d\mu = \sum \int f\chi_{A_i} d\mu = \sum \mu f(A_i)$
2. μf ist endlich. Es ist per Definition: $+\infty > \mu f(\Omega) = \int f\chi_\Omega d\mu = \int f d\mu \Leftrightarrow f$ ist μ-integrierbar.

μf ist Wahrscheinlichkeitsmaß. Per Definition ist $\mu f(\Omega) = 1 = \int f d\mu$. f heißt **Wahrscheinlichkeitsdichte** bezüglich μ.

Beispiel

Sei $\Omega = \mathbb{N}_0, \mathcal{F} = \mathfrak{P}(\mathbb{N}_0)$ und $\mu := \sum_{i=0}^{\infty} \delta_i$ das Zählmaß.
Behauptung: Jedes Maß ν auf $[\mathbb{N}_0, \mathfrak{P}(\mathbb{N}_0)]$ ist absolut stetig bezüglich μ ($\nu \ll \mu$) mit der Dichte $f(i) := \nu(\{i\}) \Rightarrow \nu(A) = \mu f(A)$.
Beweis: $\nu(A) = \sum_{i\in A} \nu(\{i\}) = \sum_{i\in\mathbb{N}_0} \nu(\{i\})\chi_A(i) = \sum_{i\in\mathbb{N}_0} f(i)\chi_A(i) = \int f(i)\chi_A(i)\mu(di)$

Bemerkung

Seien $f, g \geq 0$ und messbar. Dann ist $\mu f = \mu g$ genau dann, wenn $f =_\mu g$.
Beweis:*** $\mu f = \mu g \Leftrightarrow \int_A f d\mu = \int_A g d\mu \Leftrightarrow \int_A (f-g) d\mu = 0 = \int (f-g)\chi_A d\mu$

„$\Leftarrow$“ Sei $f =_\mu g \Rightarrow (f-g) =_\mu 0 \Rightarrow (f-g)\chi_A =_\mu 0 \Rightarrow \int (f-g)\chi_A d\mu = 0 \Rightarrow \mu f = \mu g$

„$\Rightarrow$“ Angenommen $\neg f =_\mu g$, d. h. $\mu(\{\omega \in \Omega | f(\omega) \neq g(\omega)\}) = 0 \Rightarrow \mu(\{\omega \in \Omega | f(\omega) > g(\omega)\}) > 0 \vee \mu(\{\omega \in \Omega | f(\omega) < g(\omega)\}) < 0$. o. B. d. A. gelte die erste Gleichung und sei $A_n := \{\omega \in \Omega | g(\omega) \geq f(\omega) + \frac{1}{n}\} \Rightarrow A_n \leq A_{n+1}$. Wegen der Stetigkeit von unten folgt: $\lim \mu(A_n) = \mu(\bigcup A_n) = \mu(\{\omega \in \Omega | f(\omega) < g(\omega)\}) > 0 \Rightarrow \exists n_0 \colon \mu(A_{n_0}) > 0 = \mu(\{g(\omega) \geq f(\omega) + \frac{1}{n_0}\}) = \mu(\{\omega \in \Omega | g(\omega) - f(\omega) \geq \frac{1}{n_0}\}) \Rightarrow \int (g-f)\chi_{A_{n_0}} d\mu \geq \int \frac{1}{n_0}\chi_{A_{n_0}} d\mu = \frac{1}{n_0}\mu(A_{n_0} > 0$. Somit folgt, dass *nicht* gilt: $\mu f = \mu g$

***Der Beweis erfolgt einschränkend für Mengen mit endlichem Maß.

Satz 2.3.1
Sei P ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$ mit stetig differenzierbarer Verteilungsfunktion $F(x) = P((-\infty, x))$. Dann gilt:

$$P = \ell F'$$

Beweis:
[†††] Wir definieren ein Maß $\nu := \ell F'$. F ist eine Verteilungsfunktion und monoton wachsend. Damit ist die Ableitung immer positiv. Zu prüfen ist nun, ob ν ein Wahrscheinlichkeitsmaß ist und ob dies mit P zusammen fällt.

1. Dazu $\nu(\mathbb{R}) = \int_{\mathbb{R}} F'(x)\ell(dx)$. Da F stetig ist und Satz 2.2.2 gilt: $\int_{-\infty}^{\infty} F'(x)dx = \lim_{a\to-\infty} \lim_{b\to\infty} \int_a^b F'(x)dx = \lim_{a\to-\infty} \lim_{b\to\infty}(F(b)-F(a))$. Wegen des Satz 1.5.3 Punkt 3 und 4 gilt $1 - 0 \cdot 1$. Weiterhin ist zu zeigen, dass $P = \nu$. Hier genügt es zu zeigen, dass beide die gleiche Verteilungsfunktion haben: Die Verteilungsfunktion F von P ist gleich der Verteilungsfunktion G von ν. Wir haben $G(x) = \nu((-\infty, x)) = \int_{(-\infty,x)} F'(x)\ell(dy)$. Wiederum nach Satz 2.2.2 ist $\int_{-\infty}^{x} F'(y)dy = \lim_{a\to-\infty}(F(x) - F(a)) = F(x)$ ■

Beispiel
Wir betrachten die **Exponentialverteilung**:

$$Ex_\lambda \to F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

$$F'(x) = \begin{cases} 0 & x \leq 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases} \qquad Ex_\lambda = \ell F'(x)$$

Satz 2.3.2
Seien μ ein Maß auf $[\Omega, \mathcal{F}]$, $f, g \colon [\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}]$, $f \geq 0$. Ist $f \cdot g$ bezüglich μ integrierbar, so ist g allein bezüglich μ mit Dichte f integrierbar und es gilt:

$$\int g d\mu f = \int f g d\mu$$

Beweis:
0. Fall $g = \chi_A$ $(A \in \mathcal{F})$: Es gilt zu zeigen, dass $\int g d\mu f = \int f g d\mu = \int \chi_A d\mu f = \mu f(A) = \int_A f d\mu = \int f \chi_A d\mu = \int f g d\mu$

1. Fall $g = \sum a_i \chi_{A_i}$
Die weiteren Schritte können wie bereits gehabt mittels der Linearität und Limes genzeigt werden. ■

[†††] Für die Gültigkeit der Behauptung reicht die Forderung, dass F' stückweise stetig ist. Der Beweis wird hierdurch jedoch sehr umfassend, sodass auf die schwächere Bedingung zurückgegriffen wird.

Beispiel
Eine vielfache Anwendung stellt folgendes Beispiel dar:
$\mu = \ell, [\Omega, \mathcal{F}] = [\mathbb{R}, \mathfrak{B}], P = \ell f, f$ ist (stückweise) stetig, $g\colon [\mathbb{R}, \mathfrak{B}] \to [\mathbb{R}, \mathfrak{B}]$ bezüglich P integrierbar und stetig. Daraus folgt: $\int g dP = \int_{-\infty}^{\infty} f(x)g(x)dx$

2.4. Der Satz von Fubini

Definition
Sei $[\Omega, \mathcal{F}, \mu]$ ein Maßraum. Das Maß μ heißt ***σ*-endlich**, wenn eine messbare Zerlegung[‡‡‡] $(A_k)_{k=1}^{\infty}$ von Ω existiert mit $\mu(A_k) < +\infty \, (k = 1, 2, \ldots)$.

Bemerkung
Spezialfälle:

- Alle Wahrscheinlichkeitsmaße und alle endlichen Maße
- Lebesguesche Maße (jede Dimension); Wähle z. B. $A_k = [k, k+1)$ mit $(k \in \mathbb{Z}) \Rightarrow \bigcup A_k \mathbb{R}, A_k \cap A_j = \emptyset, \ell(A_k) = 1$
- Zählmaß auf abzählbaren Mengen Ω, Wähle $A_\omega = \{\omega\}$. Es ist klar, dass $\mu(A_\omega) = \sum_{\omega \in \Omega} \delta_\omega(A_\omega) = \sum_{\omega \in \Omega} \delta_\omega(\{\omega\}) = 1$

Im folgenden seien $\Omega_1, \mathcal{F}_1, \mu_1], \ldots, [\Omega_n, \mathcal{F}_n, \mu_n]$ Maßräume und $\mu_1, \ldots, \mu_n$ σ-endlich.

Satz 2.4.1 (Satz von Fubini)
Sei $f\colon \times_{i=1}^{n} [\Omega_i, \mathcal{F}_i] \to [\mathbb{R}, \mathfrak{B}]$ bezüglich $\times_{i=1}^{n} \mu_i$ integrierbar. Weiter sei $(i_1, \ldots, i_n)$ eine Permutation der Zahlen $\{1, \ldots, n\}$. Dann gilt:

$$\int f d \times_{i=1}^{n} \mu_i = \int \cdots \int f(\omega_1, \ldots, \omega_n) \mu_{i_1}(d\omega_{i_1}) \ldots \mu_{i_n}(d\omega_{i_n})$$

Beispiel
Wir wissen, $[\mathbb{R}^n, \mathfrak{B}_n, \ell^{(n)}] = [\mathbb{R}, \mathfrak{B}, \ell]^{n\times}$. Wenn $f(x_1, \ldots, x_n)$ stetig, ist sie in jeder Variablen stetig und $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |f(x_1, \ldots, x_n)| dx_1 \ldots dx_n < +\infty$. Es gilt:

$$\int f d\ell^{(n)} = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

Satz 2.4.2
Seien $[\Omega_1, \mathcal{F}_1, \mu_1], \ldots, [\Omega_n, \mathcal{F}_n, \mu_n]$ Maßräume mit σ-endlichen Maßen. Weiterhin seien $f_i\colon [\Omega_1, \mathcal{F}_i] \to [\mathbb{R}, \mathfrak{B}], f_i \geq 0, \int f_i d\mu_i = 1$[§§§]. Setzen $f(\omega_1, \ldots, \omega_n)\colon = \prod_{i=1}^{n} f_i(\omega_i)$. Dann gilt:

$$\mu f = \times_{i=1}^{n} \mu_i f_i \quad \text{wobei } \mu = \times_{i=1}^{n} \mu_i$$

[‡‡‡] d. h. $A_k \cap A_j = \emptyset \wedge \bigcup A_k = \Omega$
[§§§] d. h. f_i ist Wahrscheinlichkeitsdichte bezüglich μ_i bzw. $\mu_i f_i$ ist Wahrscheinlichkeitsmaß.

Beweis:

$$
\begin{aligned}
\mu f(\times_{j=1}^{n} A_j) &= \int \chi_{\times_{j=1}^{n} A_j} d(x\mu_i) f \\
&= \int \cdots \int \underbrace{\chi_{\times_{j=1}^{n} A_j}}_{\prod_{j=1}^{n} \chi_{A_j}(\omega_j)} (\omega_1, \dots, \omega_n) \mu_1(d\omega_1) \dots \mu_n(d\omega_n) \\
&= \int \cdots \int \prod_{j=1}^{n} \chi_{A_j}(\omega_j) \prod_{k=1}^{n} f_k(\omega_k) \mu_1(d\omega_1) \dots \mu_n(d\omega_n) \\
&= \int \cdots \int \prod_{j=1}^{n} (\chi_{A_j}(\omega_j) f_j(\omega_j)) \mu \dots \mu \\
&= \prod_{j=1}^{n} \int \chi_{A_j}(\omega_j) f_j(\omega_j) \mu_j(d\omega_j) = \prod_{j=1}^{n} \mu_j f_j(A_j)
\end{aligned}
$$

■

3. Wahrscheinlichkeitsräume

Das Anliegen dieses Kapitels ist es, die Grundtypen von Wahrscheinlichkeitsräumen zu vermitteln und die Bildung von Modellen zu trainieren. Weiterhin werden wichtige Typen von Verteilungen auf $[\mathbb{R}, \mathfrak{B}]$ abgeleitet.

3.1. Der klassische Wahrscheinlichkeitsraum

Dieser Wahrscheinlichkeitsraum war der erste, der in der Historie benutzt wurde. Wie schon in der Einleitung geschrieben, haben Spieler die Wahrscheinlichkeit zuerst gefühlsmäßig wahrgenommen und versucht, dies genauer zu errechnen. Dies geschah um im Anschluß eventuell eine Strategie zu entwickeln. Im 18. Jahrhundert kam dann ein französischer Mathematiker auf die Idee, den Zufall mit gleichen Wahrscheinlichkeiten zu modellieren. Dies ist die Wurzel des klassischen Wahrscheinlichkeitsraumes.

Bei der Betrachtung ergibt sich die folgende Situation: Gegeben sei ein Experiment oder Versuch mit n-möglichen Ergebnissen. Dabei bezeichnet das Wort „Experiment“ die prinzipielle Möglichkeit der Wiederholung. Weiterhin darf kein bestimmtes Ereignis gegenüber einem anderen bevorzugt werden. Abstrakt hat man dann, endliches $\Omega = \{\omega_1, \dots, \omega_n\}$ als die Klasse der möglichen Ereignisse. Weiterhin ist $\mathcal{F} = \mathfrak{P}(\Omega)$ das System aller möglichen Ereignisse und $A \in \mathcal{F}$ ist eine Aussage über ein Ereignis.

Beispiel

1. Münzwurf: $\Omega = \{W, Z\} = \{0, 1\}$
2. Würfel: $\Omega = \{1, 2, 3, 4, 5, 6\}$

Ein Ansatz für ein Wahrscheinlichkeitsmaß ist: $P(A) = \frac{\#A}{\#\Omega}$. Dieses Modell nimmt an, dass alle Ereignisse gleich verteilt sind: $P(\{\omega_i\}) = P(\{\omega_j\})$ für alle $i, j = 1, 2, \dots$. Das Wahrscheinlichkeitsmaß hier ist: $P(\Omega) = 1 = P(\bigcup_{i=1}^n \{\omega_i\}) = \sum_{i=1}^n P(\{\omega_k\}) \Rightarrow P(\{\omega_k\}) = \frac{1}{n} = \frac{1}{\#\Omega}$. Somit folgt dann $P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{\#\Omega} = \frac{\#A}{\#\Omega}$.

Definition
Ein Wahrscheinlichkeitsraum $[\Omega, \mathcal{F}, P]$ heißt **klassisch**, wenn gilt

- Ω ist endlich und nichtleer
- $\mathcal{F} = \mathfrak{P}(\Omega)$

- $P(A) = \frac{\#A}{\#\Omega}$ $(A \in \mathcal{F} \Leftrightarrow A \subseteq \Omega)$

Bemerkung
In klassischen Wahrscheinlichkeitsräumen wird P in der Regel wie folgt dargestellt: Sei $\mu = \sum_{\omega \in \Omega} \delta_\omega$ ein Zählmaß auf Ω. Dann kann man das **normierte Zählmaß** als $P(A) = \frac{1}{\mu(\Omega)}\mu$ darstellen. Allgemeiner gilt: Sei μ ein endliches Maß auf $[\Omega, \mathcal{F}]$ mit $\mu(\Omega) \neq 0$. Dann kann man $Q\colon = \frac{1}{\mu(\Omega)}\mu$ setzen.

Beispiel
1. Münzwurf: $\Omega = \{0, 1\} \Rightarrow P(\{\omega\}) = \frac{1}{2}$
2. Würfel: $\Omega = \{1, \dots, 6\} \Rightarrow P(\{\omega\}) = \frac{1}{6}$
3. „kompliziertere“ Modelle:
   - n-facher Münzwurf: „Menge der möglichen Protokolle“ $\Omega = \{[\omega_1, \dots, \omega_n] | \omega_i \in \{0, 1\}, i = 1, \dots, n\}$
   - n-mal Würfeln: $\Omega = \{1, \dots, 6\}^{n\times}$

Kombinatorische Formeln Dies stellt nur einen Einschub dar. In der Regel sollten diese Regeln dem Student bekannt sein.

Die kombinatorischen Formeln drücken aus, wieviele Möglichkeiten der Auswahl von k Elementen einer n-elementigen Menge existieren. Dabei kann man sich vorstellen, dass aus eine Urne mit insgesamt n numerierten Kugeln nach bestimmten Kriterien Kugeln gezogen werden.

1. Variante Auswahl *mit* Berücksichtigung der Reihenfolge und *mit* Zurücklegen
$A_{k,n}^{mWmR}\colon = \{[\omega_1, \dots, \omega_k] | \omega_i \in \{1, \dots, n\}, i = 1, \dots, k\} = \{1, \dots, n\}^k \Rightarrow \#A_{k,n}^{mWmR} = n^k$

2. Variante Auswahl *mit* Berücksichtigung der Reihenfolge und *ohne* Zurücklegen ($k \leq n$)
$A_{k,n}^{oWmR}\colon = \{[\omega_1, \dots, \omega_k] | \{\omega_1, \dots, \omega_k\} \in \{1, \dots, n\}\} = n(n-1)(n-2)\dots(n-k+1) \Rightarrow \#A_{k,n}^{oWmR} = \frac{n!}{(n-k)!}$
Für den Spezialfall $k = n$ folgt dann: $\#A = n!$. Dies entspricht den Permutationen von $\{1, \dots, n\}$.

3. Variante Auswahl *ohne* Berücksichtigung der Reihenfolge und *ohne* Zurücklegen
$A_{k,n}^{oWoR}\colon = \{\{\omega_1, \dots, \omega_k\} | \{\omega_1, \dots, \omega_k\} \subseteq \{1, \dots, n\}\}$
$\Rightarrow \#A_{k,n}^{oWmR} \frac{1}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k} = \#A_{k,n}^{oWoR}$

4. Variante Auswahl *ohne* Berücksichtigung der Reihenfolge und *mit* Zurücklegen
$A_{k,n}^{mWoR} = \binom{k+n}{n} = \frac{(k+n)!}{n!k!}$

Beispiel
n-facher Münzwurf: $\Omega = \{0,1\}^n = A_{n,2}^{mWmR} \Rightarrow \#\Omega = 2^n$. Wir betrachten das Ereignis k-mal wird Zahl geworfen. Dies entspricht $\{[\omega_1,\dots,\omega_n] \in \Omega | k = \sum_{i=1}^n \omega_i\} = A_{k,n}^{oWoR} = \binom{n}{k} \Rightarrow P(\text{„}A_k\text{“}) = \frac{\#A_k}{\#\Omega} = \frac{\binom{n}{k}}{2^n}$.

Das Ereignis B_k entspricht der Aussage „beim k-ten Wurf wird erstmalig Zahl geworfen“:

$$\begin{aligned} B_k &= \{[\omega_1,\dots,\omega_n] \in \Omega | \underbrace{\omega_1,\dots,\omega_{k-1}}_{=0}, \omega_k = 1\} \\ &= \{[0,\dots,0,1]\} \times \{[\omega_{k+1},\dots,\omega_n] | \omega_j \in \{0,1\}\} \\ P(\text{„}B_k\text{“}) &= \frac{\#B_k}{\#\Omega} = \frac{2^{n-k}}{2^n} = 2^{-k} \end{aligned}$$

Hypergeometrisches Modell

Zentrum der Betrachtung ist in der Regel eine Urne. Dabei steht dies für ein Behältnis, dass von außen nicht einsehbar ist und in dem alle Kugeln gleiche Oberflächeneigenschaften haben.*

In der Urne befinden sich N Kugeln. Darunter sind $M^\dagger$ rote und $N - M$ weiße Kugeln. Nun werden aus der Urne „auf gut Glück“ n Kugeln herausgegriffen. Wir betrachten das Ereignis A_k, welches für „genau k der n Kugeln sind rot“ steht. Als Hilfe nehmen wir an, dass die Kugeln nummeriert sind und zwar von 1 bis M für die roten und $M+1$ bis N für die weißen Kugeln. Mathematisch wird dies wie folgt ausgedrückt:

$$\begin{aligned} &\Omega \colon = \{\{\varepsilon_1,\dots,\varepsilon_n\} | \{\varepsilon_1,\dots,\varepsilon_n\} \subseteq \{1,\dots,N\}\} = A_{n,N}^{oRoW} \\ &\mathcal{F} = \mathfrak{P}(\Omega) \qquad P(A_k) \colon = \frac{\#A_k}{\#\Omega} = \frac{\#A_k}{\binom{N}{n}} \\ &A_k = \{\{\varepsilon_1,\dots,\varepsilon_n\} \subseteq \Omega | \#(\{\varepsilon_1,\dots,\varepsilon_n\} \cap \{1,\dots,M\}) = k \text{ und} \\ &\qquad \#(\{\varepsilon_1,\dots,\varepsilon_n\} \cap \{M+1,\dots,N\}) = n-k\} \end{aligned}$$

Somit ist $A_k = \emptyset$, wenn $k < 0 \vee k > M \vee k > n \vee n - k > N - M \Leftrightarrow k < n + M - N$.

*Untersuchungen während des Zweiten Weltkrieges ergaben, dass Menschen Kugeln anhand ihrer Oberflächen unbewusst wiedererkennen und dann Präferenzen ausbilden. Dies sollte bei den Experimenten immer ausgeschlossen werden.

†Wobei gilt: $M \leq N$.

Wir betrachten nun, $\max(0, n+M-N) \leq k \leq \min(M, n)$.

$$\#A_k = \#A^{oWoR}_{k,M} \cdot \#A^{oWoR}_{n-k,N-M} = \binom{M}{k}\binom{N-M}{n-k}$$

$$\#\Omega = \#A^{oWoR}_{n,N} = \binom{N}{n}$$

$$\Rightarrow P(A_k) = \begin{cases} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} & k \text{ mit } \max(0, n+M-N) \leq k \leq \min(M, n) \\ 0 & \text{sonst} \end{cases}$$

Aus der obigen Tatsache folgt nun:

$$\sum_{k=\max(0,n+M-N)}^{\min(M,n)} \binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n}$$

Zum Beweis dieser Aussage wissen wir, dass $A_k \cap A_j = \emptyset$ für $k \neq j$ und $\bigcup_k A_k = \Omega^\ddagger$ ist. Somit folgt, dass $P(\Omega) = 1 = \sum_k P(A_k) = \sum \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$.

Weiterhin kann man schlussfolgern:

$$H_{N,M,n} :\, = \sum_{k=\max(0,n+M-N)}^{\min(M,n)} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \delta_k$$

ist ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$.

3.2. Das Bernoullischema

Definition
Seien $n \geq 1$ mit $n \in \mathbb{N}$ und $0 \leq p \leq 1$. Der Wahrscheinlichkeitsraum $[\Omega, \mathcal{F}, P] = [\{0,1\}, \mathfrak{P}(\{0,1\}), p\delta_1 + (1-p)\delta_0]^{n\times}$ heißt (endliches) **Bernoullischema** zu n Versuchen mit der Erfolgswahrscheinlichkeit p.

Bemerkung

1. $[\Omega, \mathcal{F}, P] = [\{0,1\}^{n\times}, \mathfrak{P}(\{0,1\}^{n\times}), (p\delta_1 + (1-p)\delta_0)^{n\times}]$
2. $\Omega = \{[\omega_1, \ldots, \omega_n] | \omega_i \in \{0,1\}, i = 1, \ldots, n\} = \{0,1\}^{n\times}$
3. $\mathcal{F} = \mathfrak{P}(\Omega)$
4. $P(\{[\omega_1, \ldots, \omega_n]\}) = P(\times_{i=1}^n \{\omega_i\}) = (p\delta_1 + (1-p)\delta_0)^{n\times}(\times_{i=1}^n \{\omega_i\}) = \prod_{i=1}^n (p\delta_1 + (1-p)\delta_0)(\{\omega_i\}) = \prod_{i=1}^n (p\delta_1)(\{\omega_i\}) + (1-p)\delta_0(\{\omega_i\}) = \prod_{i=1}^n p^{\omega_i}(1-p)^{1-\omega_i} = p^{\sum \omega_i}(1-p)^{n-\sum \omega_i}$

‡Für k gilt $\max(0, n+M-N) \leq k \leq \min(M, n)$

Beispiel
Als Beispiel soll der n-fache Münzwurf mit einer Erfolgswahrscheinlichkeit von $p = 1/2$ nach dem klassichen Modell dienen. Wir betrachten die Ereignisse A_k und B_k. Dabei entspricht A_k der Aussage „k-mal Erfolg bei n Versuchen“ und B_k der Aussage „erstmalig beim k-ten Versuch Misserfolg“.

$$\begin{aligned}
A_k &= \{[\omega_1, \dots, \omega_n] \in \Omega | \sum \omega_i = k\} \\
P(A_k) &= \sum_{[\omega_1,\dots,\omega_n] \in A_k} p^{\sum \omega_i}(1-p)^{n-\sum \omega_i} = \sum_{\omega \in A_k} P(\{\omega\}) \\
&= p^k (1-p)^{n-k} \sum_{\omega \in A_k} 1 = \sum p^k (1-p)^{n-k} = p^k (1-p)^{n-k} \# A_k \\
&= \begin{cases} p^k (1-p)^{n-k} \binom{n}{k} & k = 0, \dots, n \\ 0 & \text{sonst} \end{cases}
\end{aligned}$$

Damit folgt, dass $\sum_{k=0}^{n} p^k (1-p)^{n-k} \binom{n}{k} = 1$. Denn die A_k sind disjunkt und ihre Vereinigung ergibt gerade Ω.

Die **Binomialverteilung** zu n Versuchen mit der Erfolgswahrscheinlichkeit p ist wie folgt definiert:

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

Beispiel
Weiterführung des obigen Beispiels:
Es ist klar, dass $P(B_0) = 0$. Für $P(B_k)$ gilt:

$$\begin{aligned}
P(B_k) &= (p\delta_1 + (1-p)\delta_0)^{n\times}(\{[1, \dots, 1, 0]\} \times \{0,1\}^{n-k}) \\
&= (p\delta_1 + (1-p)\delta_0)^{k\times}(\{[1, \dots, 1, 0]\})(p\delta_1 + \underbrace{(1-p)\delta_0)^{n-k}(\{0,1\}^{n-k})}_{=1} \\
&= p^{k-1}(1-p)
\end{aligned}$$

Zur Erinnerung: Die Verteilung

$$G_p :\, = \sum_{k=1}^{\infty} p^k (1-p) \delta_k$$

heißt **geometrische Verteilung**.

Strategien beim wiederholten Glücksspiel Im Rahmen derartiger Spiele wird beabsichtigt, einen Gewinn K zu erzielen. Dabei kann man mit folgender Strategie vorgehen:

1. Schritt Setze K. Bei Erfolg wird der Gewinn K realisiert. Andernfalls gehe zum nächsten Schritt.

2. Schritt Setze $(1+1)K$. Bei Erfolg wird Gewinn K realisiert. Andernfalls gehe zum nächsten Schritt.

3. Schritt Setze $(1+2+1)K$. Bei Erfolg wird Gewinn K realisiert. Andernfalls gehe zum nächsten Schritt.

Im n Schritt wird also immer ein Betrag von $2^{n-1}K$ gesetzt und im Gewinnfall wird schliesslich K ausgeschüttet.

Im Falle $n = 10$ sind $512K$ einzusetzen. Sei nun $p = 1/2$. Dann ergibt sich eine Gewinnwahrscheinlichkeit von $(1-p)\sum_{k=0}^{9} p^k = (1-p)\frac{(1-p^9)}{(1-p)} = 1-p^9 \approx 0,999$

Beispiel

radioaktiver Zerfall Wir betrachten die Zeit t. Zu Beginn dieser Zeit existieren n Atome des radioaktiven Stoffes. Diese sind mit $1, \dots, n$ nummeriert. Das Ereignis $\omega_k = 1$ entspricht der Aussage, „Atom k ist in der Zeit t zerfallen“ und das Ereignis $\omega_k = 0$ entspricht der Aussage, „Atom k ist in der Zeit t nicht zerfallen“. Die Anzahl der zerfallenen Atome berechnet sich durch $\sum_{k=1}^{n} \omega_k$.

Genetik Hierbei wird die Teilung eines DNA-Stranges beobachtet. Solch ein Strang teilt und „kopiert“ sich. Dabei entsteht keine 1:1-Kopie, sondern an endlich vielen Stellen treten Fehler auf. Nach Ansicht der Neodarwinisten erfolgt auf diesem Wege die Evolution.

Insgesamt existieren n Verbindungen[§]. Das Ereignis $\omega_k = 1$ entspricht einem Fehler an der k-ten Stelle und das Ereignis $\omega_k = 0$ sagt aus, dass an der k-ten Stelle kein Fehler ist.

Will man nun bei beiden Beispielen versuchen, nach dem bisher gelernten die Wahrscheinlichkeit zu errechnen, wird man sehr schnell auf Probleme stoßen. So ist im Beispiel der Genetik $P(\text{„}k\text{-mal Erfolg“}) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{100000000}{25} p^{25} (1-p)^{\approx 100000000} \approx \infty \cdot 0$.

3.3. Der Poissonsche Grenzwertsatz

Wir betrachten die Binomialverteilung $B_{n,p}$ für sehr große n und kleine p, d. h.

$$B_{n,p}(\{k\}) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & 0 \le k \le n \\ 0 & \text{sonst} \end{cases}$$

oder

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

[§] Beim Mensch entspricht das n ungefähr 100 Millionen.

und weiter die Poissonverteilung zum Parameter λ:

$$\Pi_\lambda(\{k\}) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad (\lambda > 0)$$

Satz 3.3.1 (Poissonsche Grenzwertsatz)
Sei $(p_n)_{n=1}^{\infty}$ eine Folge mit $0 \leq p_n \leq 1$ für alle $n = 1, 2, \ldots$ und $0 < \lambda = \lim_{n\to\infty} np_n$. Dann gilt:

$$\lim_{n\to\infty} B_{n,p_n}(\{k\}) = \Pi_\lambda(\{k\}) \quad (k = 0, 1, \ldots)$$

Beweis:
in der Übung ■

Beispiel
Seien $n = 10^{11}, p = c \cdot 10^{-11} \Rightarrow np = c, \lambda = 1$. Es gilt die Annahme, $B_{n,p}(\{k\}) \approx \Pi_c(\{k\})$. Wir betrachten $k = 0, 1, 2, 3, 4, 5$ für $c = 1$:

$$B_{n,p}(\{p\}) \approx \Pi_1(\{5\}) = {}^1\!/_{5!}e^{-1} = {}^1\!/_{120\cdot e} \leq {}^1\!/_{250}$$

3.4. Ein kontinuierliches Analogon des klassischen Wahrscheinlichkeitsraumes

Im kontinuierlichen Modell ist $\Omega \subseteq \mathbb{R}^n, 0 \leq \ell^{(n)}(\Omega) < +\infty$:

$$\Rightarrow Gl_\Omega :\, = \frac{\ell^{(n)}}{\ell^{(n)}(\Omega)}$$

4. Bedingte Wahrscheinlichkeit und Unabhängigkeit

4.1. Bedingte Wahrscheinlichkeit

Wenn man einer Lottoziehung zuschaut und feststellt, dass man nacheinander die gezogenen Zahlen auch getippt hat, verändern sich die Wahrscheinlichkeiten kontinuierlich. Im folgenden soll dies erklärt werden.

Im folgenden gilt immer, dass $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum ist.

Definition
Seien $A, B \in \mathcal{F}$ mit $P(B) > 0$. Dann heißt

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

bedingte Wahrscheinlichkeit von A unter der Bedingung B.

Beispiel
Als Beispiel dient hier wieder der Würfel:

$$\Omega = \{1,2,3,4,5,6\}, P(\{k\}) = 1/6, k \in \Omega, A = \{6\}, B = \{4,5,6\}$$
$$P(A|B) = \frac{P(\{6\} \cap \{4,5,6\})}{P(\{4,5,6\})} = \frac{P(\{6\})}{P(\{4,5,6\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

Satz 4.1.1 (Elementare Eigenschaften)
Sei $B \in \mathcal{F}, P(B) > 0$. Dann gelten:

(a) $0 \leq P(A|B) \leq 1$

(b) $P(B|B) = P(\Omega|B) = 1$

(c) $P(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$ für alle $(A_i)_{i=1}^{\infty} \in \mathcal{F}$ mit $A_i \cap A_j \neq \emptyset$ $(i \neq j)$

Beweis:
(a) Nach der Definition ist klar, dass $P(A|B) \geq 0$ gilt. Weiterhin ist $A \cap B \subseteq B$. Nach der Monotonie von P folgt damit, dass $P(A \cap B) \leq P(B)$. Durch Umstellen erhält man nun $\frac{P(A \cap B)}{P(B)} \leq 1$. Es ist klar, dass $P(\emptyset|B) = \frac{P(\emptyset \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0$ ist.

(b) $P(\Omega|B) = \frac{P(B)}{P(B)} = 1 = P(B|B)$

(c) Seien die (A_i) aus $\mathcal{F}$ paarweise disjunkt. Dann sind auch $(A_i \cap B)$ paarweise disjunkt und es folgt, $P(\bigcup_{i=1}^{\infty}(A_i \cap B)) = \sum_{i=1}^{\infty} P(A_i \cap B)$. Somit hat man damit:

$$\frac{P(\bigcup_{i=1}^{\infty} A_i \cap B)}{P(B)} = \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)}$$

■

Bemerkung
Durch $A \to P(A|B)$ ist ein Wahrscheinlichkeitsmaß auf $[\Omega, \mathcal{F}]$ gegeben. Man bezeichnet $P(\cdot|B)$ als **bedingte Verteilung** unter der Bedingung B.

Satz 4.1.2 (Formel bzw. Satz über die totale Wahrscheinlichkeit)
Sei $(B_i)_{i\in I}$ eine abzählbar messbare Zerlegung von Ω, d. h. $B_i \in \mathcal{F}, i \in I$ abzählbar, $B_i \cap B_j = \emptyset$ für $i \neq j, \bigcup B_i = \Omega$. Dann gilt:

$$P(A) = \sum_{i\in I} P(A|B_i)P(B_i) \qquad A \in \mathcal{F}, P(B_i) > 0$$

Beweis:

$$\sum_{i\in I} P(A|B_i)P(B_i) = \sum_{i\in I} \frac{P(A \cap B_i)}{P(B_i)} P(B_i) = \sum_{i\in I} P(A \cap B_i)$$
$$P\left(\bigcup_{i\in I}(A \cap B_i)\right) = P\left(A \cap \bigcup_{i\in I} B_i\right) = P(A \cap \Omega) = P(A)$$

■

Bemerkung
Die zusätzliche Forderung $P(B_i) > 0$ ist nicht notwendig, da $P(A|B_i) \cdot P(B_i) = 0$. In dem Zusammenhang ist die Forderung jedoch sinnvoll.

Satz 4.1.3 (Satz von Bayes)
Sei $(B_i)_{i\in I}$ eine messbare Zerlegung von Ω und $A \in \mathcal{F}$ mit $P(A) > 0$. Dann gilt:

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{\sum_{j\in I} P(A|B_j) \cdot P(B_j)} \qquad i \in I$$

Für den Spezialfall $A, B \in \mathcal{F}$ und $P(A) > 0$ gilt:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$

Beweis:
Aus Satz 4.1.2 folgt $P(A) = \sum_{j\in I} P(A|B_j)P(B_j)$. Weiter folgt aus der Definition der bedingten Wahrscheinlichkeit: $P(A|B_i)P(B_i) = \frac{P(A\cap B_i)}{P(B_i)} \cdot P(B_i) = P(A\cap B_i)$. Aus beiden schließlich ergibt sich:

$$\frac{P(A|B_i)\cdot P(B_i)}{\sum_{j\in I} P(A|B_j)P(B_j)} = \frac{P(A\cap B_i)}{P(A)} = P(B_i|A)$$

■

Beispiel
In einem BSE-Schnelltest werden 100 Millionen Kühe getestet. Davon sind 100 Kühe an BSE erkrankt und bei nahezu allen ist der Test auch positiv. Aber bei 0,001 % der gesunden Kühe schlägt der Test auch an. Wir betrachten folgende Ereignisse:

- $A =$ „Test positiv“ $\Rightarrow A^c =$ „Test negativ“
- $B =$ „Kuh hat BSE“ $\Rightarrow B^c =$ „Kuh hat kein BSE“
- „Testzuverlässigkeit“: $P(A|B) \approx 1 - 10^{-3} = 99,9\,\%$ Dies entspricht der Wahrscheinlichkeit, dass der Test bei einer BSE-Kuh auch BSE anzeigt.

Informationsübertragungsmodell Seien $S_1 = S_2 = S$ ein endliches Alphabet. S_1 bezeichnet das Alphabet des Senders, S_2 das des Empfängers. Es ist $\mathcal{F} = \mathfrak{P}(\Omega)$ und $\Omega = S_1\times S_2 = S\times S = \{[s_1,s_2]|s_1,s_2\in S\}$. A_s entspricht dem Ereignis, dass S empfangen und B_s das S gesendet wird. Dabei ist $s\in S$. Weiter ist $A_{s_2} = S\times\{s_2\}$, $B_{s_1} = \{s_1\}\times S$, $A_{s_2}\cap B_{s_1} = \{[s_1,s_2]\}$ und $q(s_1,s_2)\colon = P(A_{s_2}|B_{s_1})$. Die Wahrscheinlichkeit p_{s_1} bezeichnet die a-priori-Wahrscheinlichkeit $P(B_{s_1})$. Damit folgt, $P(A_{s_2}) = \sum_{s_1\in S} P(A_{s_2}|B_{s_1})P(B_{s_1}) = \sum_{s_1\in S} p_{s_1} q(s_1,s_2)$. Ideal ist $q(s_1,s_2) = \delta_{s_1,s_2}$.

4.2. Lebensdauerverteilungen

Im folgenden gilt immer: $\Omega = \mathbb{R}_+ = (0,+\infty), \mathcal{F} = \mathfrak{B}\cap\mathbb{R}_+ = \{A\cap\mathbb{R}_+|A\in\mathfrak{B}\}$

Definition
Ein Wahrscheinlichkeitsmaß Q auf $[\Omega,\mathcal{F}]$ heißt **Lebensdauerverteilung**.

- Sei $Q = \sum_{i=1}^{\infty} q_i\delta_i$ (Q ist bezüglich des Zählmaßes absolut stetig.). Dann heißt

  $$r_Q(i)\colon = \frac{q_i}{\sum_{j\geq i}^{\infty} q_i} \qquad q_i \geq 0, \sum q_i = 1, i = 1,2,\ldots$$

  Sterberate von Q zum Zeitpunkt i.

- Sei $Q = \ell f$ mit $f\geq 0, \int f d\ell = 1$ (Q ist absolut stetig mit Dichte f.). Dann heißt

  $$r_Q(x)\colon = \frac{f(x)}{\int_{[x,\infty)} f(y)\ell(dy)} \qquad x\in\mathbb{R}_+$$

  Sterberate vom Q zur Zeit $x\in\mathbb{R}_+$.

Satz 4.2.1

(a) Sei $Q = \sum_{i=1}^{\infty} q_i \delta_i$. Dann gilt

$$r_Q(i) = Q(\{i\} | \{i, i+1, \dots\}) \qquad i = 1, 2, \dots$$

(b) Sei $Q = \ell f$ und f stetig. Dann gilt

$$r_Q(x) = \lim_{\Delta \to 0} {}^1\!/\!_\Delta Q([x, x+\Delta) | [x, +\infty)) \qquad x \in \mathbb{R}_+$$

Beweis:

(a) $Q(\{i\} | \{i, i+1, \dots\}) = \frac{Q(\{i\} \cap \{i, i+1, \dots\})}{Q(\{i, i+1, \dots\})} = \frac{Q(\{i\})}{Q(\{i, i+1, \dots\})} = \frac{Q(\{i\})}{\sum_{j \geq i} Q(\{j\})} = \frac{q_i}{\sum_{j \geq i} q_j} = r_Q(i)$

(b) $Q([x, x+\Delta) | [x, +\infty)) = \frac{Q([x, x+\Delta) \cap [x, +\infty))}{Q([x, +\infty))} = \frac{Q([x, x+\Delta))}{Q(x, +\infty)} = \frac{\int_{x, x+\Delta} f(y) \ell(dy)}{\int_{x, +\infty} f(y) \ell(dy)}$ Nunmehr werden die Zähler separat betrachtet und mit der Verwendung, dass f stetig ist, folgt: ${}^1\!/\!_\Delta \int_{[x, x+\Delta)} f(y) \ell(dy) \cdot {}^1\!/\!_\Delta \int_x^{x+\Delta} f(y)(dy) = {}^1\!/\!_\Delta \left(\int_0^{x+\Delta} f(y) dy \cdot \int_0^x f(y) dy \right) \xrightarrow{\Delta \to 0} f(x)$. Somit folgt die Behauptung. ∎

Beispiel (Interpretation zu Satz 4.2.1)

Fall a) Der Wert $\{i\}$ bedeutet, dass das Sterbealter i ist und $\{i, i+1, \dots\}$ heißt, dass mindestens das Alter i erreicht wird. $\Rightarrow r_Q(\{i\}) = P($„Sterben im Alter i“|„Alter i wurde erreicht“$)$

Fall b) $[x, x+\Delta)$ heißt, Sterben in einem Alter zwischen x und $x+\Delta$ und $[x, +\infty)$ bedeutet, dass mindestens das Alter x erreicht wird.

Beispiel

1. Sei $Q = G_p = (1-p) \sum_{i=1}^{\infty} p^{i-1} \delta_i \Rightarrow r_Q(i) = 1 - p$ für $i = 1, 2, \dots$.
2. Sei $Q = \text{Ex}_\lambda = \ell f_\lambda$ mit $f_\lambda = \begin{cases} 0 & x \leq 0 \\ \lambda e^{-\lambda x} & x > 0 \end{cases}$. Dabei ist $\lambda > 0$ gegeben. Sei $F(x) = \text{Ex}_\lambda((-\infty, x)) = \text{Ex}([0, x)) = \int_{[0,x)} f_\lambda(y) \ell(dy) = \int_0^x f(x)\, dy$. F ist Verteilungsfunktion zu $Q \to \int_{[x, +\infty)} f(y) \ell(dy) = 1 - Fx) \Rightarrow r_Q(x) = \frac{f(x)}{1 - F(x)}$. Es ist $Q = \text{Ex}_\lambda \to F(x) = 1 - e^{-\lambda x} \Rightarrow r_Q(x) = \frac{\lambda e^{-\lambda x}}{e^{-\lambda x}} = \lambda$. Also hängt die Sterberate nicht vom augenblicklichen Alter ab.

Typisches Verhalten der Sterberate r_Q

1. technische/mechanische Systeme (Reibung, allg. Abnutzung)
2. Elektronische Bauteile
3. Biologische Systeme (höher organisiert, insbesondere Mensch)

4.3. Unabhängigkeit von Ereignissen

Im folgenden gilt immer, dass $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum ist.

Definition
Seien $A, B \in \mathcal{F}$. A und B heißen **unabhängig**, wenn gilt

$$P(A \cap B) = P(A) \cdot P(B)$$

Bemerkung

1. Sei $P(B) > 0$. Dann gilt A, B unabhängig $\Leftrightarrow P(A|B) = P(A)$. Denn $P(A) = P(A|B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A)P(B) = P(A \cap B)$
2. Sei $P(B) = 0 \Rightarrow A, B$ unabhängig für beliebige $A \in \mathcal{F}$. Denn $P(B) = 0 \Rightarrow P(A)P(B) = 0; A \cap B \subseteq B \Rightarrow 0 = P(A \cap B) \leq P(B) = 0$
3. $P(B) = 1 \Rightarrow A, B$ unabhängig. Zum Beweis hierfür siehe auch den folgenden Satz: $P(B) = 1 \Leftrightarrow P(B^c) = 0 \Rightarrow A, B^c$ unabhängig $\Rightarrow A, (B^c)^c$ unabhängig.

Satz 4.3.1
A, B sind genau dann unabhängig, wenn A, B^c unabhängig.

Beweis:
Rein mengentheoretisch ist klar, dass gilt $A = (A \cap B) \cup (A \cap B^c)$ und $(A \cap B) \cap (A \cap B^c) = \emptyset$.

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap B^c) \\ &= P(A)P(B) + P(A \cap B^c) \\ P(A \cap B^c) &= P(A) - P(A)P(B) = P(A)\underbrace{(1 - P(B))}_{P(B^c)} = P(A)P(B^c) \end{aligned}$$ ■

Definition
Sei $\mathfrak{A} \subseteq \mathcal{F}$. Dann heißt $\mathfrak{A}$ **vollständig unabhängig**, wenn $\forall \{A_1, A_2, \ldots, A_n\} \subseteq \mathfrak{A}$ gilt

$$P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P(A_k)$$

Weiter heißt $\mathfrak{A}$ **paarweise unabhängig**, wenn $\forall \{A, B\} \subseteq \mathfrak{A}$ gilt

$$P(A \cap B) = P(A)P(B)$$

Bemerkung

1. Aus der Tatsache, dass $\mathfrak{A}$ (vollständig) unabhängig ist, folgt, dass $\mathfrak{A}$ auch paarweise unabhängig ist. Die Umkehrung hingegen ist falsch.
2. Wenn $\mathfrak{A} = \{A, B\} \Rightarrow \mathfrak{A}$ vollständig unabhängig $\Leftrightarrow \mathfrak{A}$ paarweise unabhängig $\Leftrightarrow A, B$ unabhängig.

3. Sei $\mathfrak{A} = \{A_1, A_2, A_3\}$. Aus $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$ folgt im allgemeinen nicht, dass $\mathfrak{A}$ vollständig unabhängig ist.

4. $\mathfrak{A}$ ist genau dann vollständig unabhängig, wenn $\forall\{A_1, \ldots, A_n\} \subseteq \mathfrak{A}$ gilt: $A_1, \bigcap_{k=2}^n A_k$ unabhängig. Beweisidee: Die betrachtete Unabhängigkeit heißt: $P(A_1 \cap \bigcap_{k=2}^n A_k) = \prod_{j=1}^n P(A_j) = P(A_1)P(\bigcap_{k=2}^n A_k)$. Die andere Richtung wird durch vollständige Induktion gezeigt: $k = 2 \xrightarrow{n=1} P(A_1 \cap A_2) = P(A_1)P(A_2)$.

5. $\mathfrak{A}$ ist genau dann vollständig unabhängig, wenn $\forall\{A_1, \ldots, A_n\} \subseteq \mathfrak{A}$ gilt, dass $\{A_1, \ldots, A_n\}$ vollständig unabhängig. Dabei werden keine beliebigen Systeme betrachtet, sondern wir beschränken uns auf abzählbare Mengen

Beispiel
Sei $A \in \mathcal{F}, 0 < P(A) < 1$. Setzen $A_1 = A_2 = A, A_3 = \emptyset \Rightarrow P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) = 0$. Aber $P(A_1 \cap A_2) = P(A) \neq (P(A))^2 = P(A_1) \cdot P(A_2)$

Satz 4.3.2
Wenn $\mathfrak{A}$ vollständig unabhängig ist, dann ist $\tilde{\mathfrak{A}} := \{A^c | A \in \mathfrak{A}\}$ vollständig unabhängig.

Beweis:
Nach der obigen Bemerkung 5 genügt es, für den Fall $\mathfrak{A} = \{A_1, \ldots, A_n\}$ die Unabhängigkeit zu zeigen. Dies ist jedoch nach der Bemerkung 4 und Satz 4.3.1 klar. ■

Satz 4.3.3
Sei $(A_i)_{i=1}^\infty$ vollständig unabhängig und $\sum_{i=1}^\infty P(A_i) = +\infty$. Dann gilt:

$$P\left(\bigcap_{n=1}^\infty \bigcup_{i \geq n} A_i\right) = 1$$

Interpretation: $\omega \in \bigcap_{n=1}^\infty \bigcup_{i \geq n} A_i \Leftrightarrow \omega \in \bigcup_{i \geq n} A_i$ für $n = 1, 2, \ldots \Leftrightarrow \#\{i | \omega \in A_i\} = +\infty \Rightarrow \bigcap_{n=1}^\infty \bigcup_{i \geq n} A_i$ entspricht dass unendlich viele Ereignisse A_i eintreten.

Beweis:
Es genügt zu zeigen, dass $P((\bigcap_{n=1}^\infty \bigcup_{i \geq n} A_i)^c) = 0$. Es ist klar, dass $P((\bigcap_{n=1}^\infty \bigcup_{i \geq n} A_i)^c) = P(\bigcup_{n=1}^\infty \bigcap_{i \geq n} A_i^c) \leq \sum_{n=1}^\infty P(\bigcap_{i \geq n} A_i^c)$. Damit genügt zu zeigen, dass $P(\bigcap_{i \geq n} A_i^c) = 0 = P(\bigcap_{k=n}^\infty \bigcap_{i=n}^k A_i^c) \Rightarrow P(\bigcap_{i \geq n} A_i^c) = P(\bigcap_{k=n}^\infty \bigcap_{i=n}^k A_i^c)$. Wegen der Stetigkeit von oben ist $\lim_{k \to \infty} P(\bigcap_{i=n}^k A_i^c)$. Da A_i unabhängig ist, ist auch A_i^c unabhängig. Daraus folgt, dass $\lim_{k \to \infty} \prod_{i=n}^k P(A_i^c) = \lim_{k \to \infty} \prod_{i=n}^k (1 - P(A_i)) = 0$. Somit genügt es noch zeigen, dass $\ln(\prod_{i=n}^k (1 - P(A_i)) = \sum_{i=n}^\infty \ln(1 - P(A_i)) = -\infty = \sum_{i=n}^\infty - \ln(1 - P(A_i)) = +\infty$. Weiterhin wissen wir, dass $\sum P(A_i) = +\infty$. Somit muss nur noch gezeigt werden, dass $-\ln(1 - P(A_i)) \geq P(A_i)$. Wir haben allgemein somit die Ungleichung $-\ln(1 - x) \geq x$ oder $e^{-\ln(1-x)} \geq e^x$. Durch die Ableitung erhält man: $\frac{1}{1-x} \geq 1$ für $0 \leq x \leq 1$. ■

Beispiel
Man stelle sich ein unendliches Lottospiel vor. Dann entspricht A_i der Aussage, dass beim i-ten Spiel 6 Richtige gezogen werden und $\bigcap_{n=1}^\infty \bigcup_{i \geq n} A_i$ entspricht unendlich oft 6 Richtigen. Es ist klar, dass die A_i unabhängig sind und dass $P(A_i) = c > 0$.

Bemerkung
Ohne die Forderung nach Unabhängigkeit gilt die Aussage in Satz 4.3.3 im Allgemeinen nicht. Denn wähle beispielsweise $A_i = A$ und sei $P(A) \notin \{0, 1\}$. Dann ist die Summe aller $A_i = +\infty$, aber $P(\bigcup_{n=1}^{\infty} \bigcap_{i \geq n}) = P(A)$ ist ungleich eins.

Satz 4.3.4
Sei $(A_i)_{i=1}^{\infty}$ eine Folge aus $\mathcal{F}$ mit $\sum_{i \geq 1} P(A_i) < +\infty$. Dann gilt

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{i \geq n} A_i\right) = 0$$

Hierbei wurde die Unabhängigkeit *nicht* vorausgesetzt.

Beweis:
Trivialerweise ist: $\bigcap_{n \geq 1} \bigcup_{i \geq n} A_i \subseteq \bigcup_{i \geq k} A_i$ und $0 \leq P(\bigcap \bigcup A_i) \leq P(\bigcup_{i \geq k} A_i)$ für $k = 1, 2, \ldots$. Es genügt zu zeigen, dass $\lim_{k \to \infty} P(\bigcup_{i \geq k} A_i) = 0$ ist. Es ist, $P(\bigcup_{i \geq k} A_i) \leq \sum_{i \geq k} P(A_i) \xrightarrow{k \to \infty} 0$ und es reicht zu zeigen, dass $\lim_{k \to \infty} \sum_{i \geq k} P(A_i) = 0$. Das folgt aber aus $\sum_{i \geq k} P(A_i) < +\infty$. ∎

Bemerkung
Aus den obigen beiden Sätzen kann folgende Folgerung gemacht werden: Sei $(A_i)_{i=1}^{\infty}$ vollständig unabhängig. Dann gelten: a) $P(\bigcap_{n=1}^{\infty} \bigcup_{i \geq n} A_i) \in \{0, 1\}$ (0, 1-Gesetz) b) $P(\bigcap_{n \geq 1} \bigcup_{i \geq n} A_i) \Leftrightarrow P(A_i) = +\infty$ (Lemma von Borel-Cantelli)

5. Zufallsgrößen

5.1. Zufallsvariablen

Definition
Seien $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum und $[X, \mathfrak{X}]$ ein messbarer Raum.

- Eine messbare Abbildung $\xi\colon [\Omega, \mathcal{F}] \to [X, \mathfrak{X}]$ heißt **Zufallsvariable** über $[\Omega, \mathcal{F}, P]$ mit Werten in $[X, \mathfrak{X}]$.

- $P_\xi\colon = P \circ \xi^{-1}$ heißt **Verteilungsgesetz** von ξ.

- Ist $[X, \mathfrak{X}] = [\mathbb{R}, \mathfrak{B}]$, so heißt ξ **Zufallsgröße**.

Um sich dies etwas genauer vorzustellen, kann man annehmen, dass man ein zufälliges System hat. Dies ist ein Vorgang, der bei mehreren Versuchen wechselwirkungsfrei andere Ergebnisse bringt. Als Realisierung erhält man $\omega \in \Omega$. Daraus folgt die Abbildung $\xi(\omega)$. Das P muss die Verteilung P_ξ haben. Das Anliegen der Statistik ist die Bestimmung von P_ξ.

Beispiel (Bernoulli-Schema)

$$\begin{aligned}
[\Omega, \mathcal{F}, P] &= [\{0,1\}, \mathfrak{P}(\{0,1\}), p\delta_1 + (1-p)\delta_0]^{n\times} \\
&= [\{0,1\}^n, \mathfrak{P}(\{0,1\}^n), (p\delta_1 + (1-p)\delta_0)nx] \\
\Omega &= \{[\varepsilon_1, \dots, \varepsilon_n] | \varepsilon_i \in \{0,1\}, i = 1, \dots, n\} \\
\xi_k &\colon \Omega \to \mathbb{R}, \xi_k([\omega_1, \dots, \omega_n])\colon = \omega_k
\end{aligned}$$

Nun fixieren wir ein i und betrachten das Tupel $\xi_i([\varepsilon_1,\dots,\varepsilon_n]) = \xi_i$. Es ist klar, dass ξ_i eine Zufallsgröße ist.

$$\begin{aligned}
P_{\xi_i}(A) &= P\circ\xi_i^{-1}(A) = P(\xi_i^{-1}(A)) \quad A\in\mathfrak{B}\\
\xi_i^{-1} &= \{[\varepsilon_1,\dots,\varepsilon_n]|\xi_i([\varepsilon_1,\dots,\varepsilon_n])\in A\} = \{0,1\}^{(i-1)}\times A\times\{0,1\}^{(n-k)}\\
P(\xi_i^{-1}(A)) &= P(\{0,1\}^{(i-1)\times}\times A\times\{0,1\}^{(n-k)\times})\\
&= (p\delta_1+(1-p)\delta_0)^{n\times}(\{0,1\}^{(i-1)\times}\times A\times\{0,1\}^{(n-1)\times})\\
&= \underbrace{((p\delta_1+(1-p)\delta_0)^{(i-1)\times}(\{0,1\})^{(i-1)\times})}_{=1}(p\delta_1+(1-p)\delta_0)A\\
&\quad\ \underbrace{(p\delta_1+(1-p)\delta_0)(\{0,1\})^{(n-1)\times})}_{=1}\\
&= (p\delta_1+(1-p)\delta_0)(A) = B_{1,p}(A)\\
&\Rightarrow P_{\xi_i} = B_{1,p} = p\delta_1+(1-p)\delta_0
\end{aligned}$$

Im allgemeinen Fall gilt: Wenn $P = \times_{k=1}^n P_k$ und ξ_i die Projektion in $\times_{k=1}^n P_k$ ist, dann ist $P_{\xi_i} = P_i$

Wir setzen $\xi\colon = \sum_{k=1}^n \xi_k$. Die Frage ist nun, was P_ξ ist. Anzunehmen ist, dass sich die Binomialverteilung ergibt. Dazu genügt es, statt $P_\xi(A)$ nur $P_\xi(\{j\})$ zu betrachten

$$\begin{aligned}
P_\xi(\{j\}) &= P(\xi^{-1}(\{j\})) = P(\{[\omega_1,\dots,\omega_n]\colon \xi([\omega_1,\dots,\omega_n]) = j\})\\
&= P(\{[\omega_1,\dots,\omega_n]\colon \sum_{k=1}^n \underbrace{\xi_k([\omega_1,\dots,\omega_n])}_{=\omega_k} = j\})\\
&= P(\{[\omega_1,\dots,\omega_n]\colon \sum_{k=1}^n \omega_k = j\}) = \binom{n}{j}p^j(1-p)^{n-j} = B_{n,p}(\{j\})\\
&\Rightarrow P_\xi = B_{n,p}
\end{aligned}$$

Satz 5.1.1
Sei Q ein Wahrscheinlichkeitsmaß auf $[\mathbb{R},\mathfrak{B}]$ mit stetiger und streng monoton wachsender Verteilungsfunktion F_Q. Dann ist $\xi = F_Q^{-1}$ eine **Zufallsgröße** über dem Wahrscheinlichkeitsraum $[(0,1),\mathfrak{B}\cap(0,1),Gl_{(0,1)}]$ mit $P_\xi = Q$.

Beweis:
Die strenge Monotonie und Stetigkeit bedingen, dass die Umkehrfunktion wieder stetig und damit messbar ist. Weiter ist bekannt, dass das Wahrscheinlichkeitsmaß auf $[\mathbb{R},\mathfrak{B}]$ durch seine Verteilungsfunktion eindeutig bestimmt ist. Damit genügt es zu zeigen, dass $F_{P_\xi} = F_Q$.

Es gilt, $F_Q(x)\in[0,1]$ mit $x\in\mathbb{R}$. Aber F_Q kann nie die Werte 0 oder 1 wegen der strengen Monotonie annehmen $\Rightarrow F_Q(x)\in(0,1)$. Somit existiert F_Q^{-1} auf $(0,1)$ und F_Q^{-1}

ist stetig $\Rightarrow F_Q^{-1}$ messbar. Also ist $\xi = F_Q^{-1}$ eine Zufallsgröße. (Genauere Betrachtungen zu oben.)

Dazu ist nach Definition $F_{P_\xi}(x) = P_\xi((-\infty, x)) = P(\xi^{-1}((-\infty, x)))$. Weiterhin ist:

$$\begin{aligned}
P_\xi((-\infty, x)) &= P(\xi^{-1}((-\infty, x))) = P(\{y \in (0,1) | \xi(y) \in (-\infty, x)\}) \\
&= P(\{y \in (0,1) | \xi(y) < x\}) \\
&= P(\{y \in (0,1) | F_Q^{-1}(y) < x\}) = P(\{y \in (0,1) | y < F_Q(x)\}) \\
&= P((0, F_Q(x)))
\end{aligned}$$

Wegen $P = Gl_{(0,1)}$ gilt $Gl_{(0,1)}((0, F_Q(x))) = \frac{\ell((0,F_Q(x)))}{\ell((0,1))} = \frac{F_Q(x)}{1} = F_Q(x) \Rightarrow F_{P_\xi}(x) = F_Q(x)$ ■

Bemerkung
Schreibweise: $F_\xi := F_{P_\xi}$ und $P(\xi^{-1}(A)) =: P(„\xi \in A“) := P(\{\omega \in \Omega | \xi(\omega) \in A\}) \Rightarrow F_\xi(x) = P(„\xi \in (-\infty, x)“) = P(„\xi < x“)$

Es reicht die Forderung, F_Q ist streng monoton wachsend auf (a, b) mit $Q((a, b)) = 1$. Dabei ist $a = \infty, b = -\infty$ zulässig. Ohne Stetigkeit gilt die Aussage nicht.

Beispiel
Betrachten die Exponentialverteilung $Q = \mathrm{Ex}_\lambda \rightarrow Q((0, +\infty)) = 1, F_Q(x) = 1 - e^{-\lambda x}$ für $x \in (0, +\infty)$. Weitere Betrachtung in der Übung.

5.2. Unabhängige Zufallsgrößen

Im folgenden Abschnitt gilt immer, dass $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum und eine Familie von Zufallsvariablen $(\xi_i)_{i \in I}$ mit Werten in $[X_i, \mathfrak{X}_i]$ ist.

Spezialfall: $[X_i, \mathfrak{X}_i] = [\mathbb{R}, \mathfrak{B}]$.

Definition
Die Familie $(\xi_i)_{i \in I}$ heißt **unabhängig**, wenn für alle Folgen $B_i \in \mathfrak{X}_i\, (i \in I)$ die Familie $\{\xi_i^{-1}(B_i) | i \in I\} \subseteq \mathcal{F}$ (vollständig) unabhängig ist.

Satz 5.2.1
Wegen der Bemerkung h nach der Definition von unabhängigen Ereignissystemen ist klar: $(\xi_i)_{i \in I}$ ist genau dann unabhängig, wenn für alle $n \geq 2$ gilt, dass $\{\xi_1, \ldots, \xi_n\} \subseteq \{\xi_i | i \in I\}$ unabhängig ist.

Andere Formulierung: Die $(\xi_i)_{i \in I}$ sind genau dann unabhängig, wenn für jedes $n \geq 2$ und $i_1, \ldots, i_n \subseteq I$ die Folge $\xi_{i_1}, \ldots, \xi_{i_n}$ unabhängig ist.

Satz 5.2.2
Sei $(A_i)_{i \in I} \subseteq \mathcal{F}$. Die Aussage $(A_i)_{i \in I}$ ist genau dann (vollständig) unabhängig, wenn die Familie der Zufallsgrößen $(\chi_{A_i})_{i \in I}$ unabhängig ist.

Beweis:
Da die A_i in $\mathcal{F}$ liegen, folgt wegen des Satz 2.1.9, dass auch die χ_{A_i} messbar und somit Zufallsgrößen sind.

„$\Leftarrow$“ Es sei vorausgesetzt, dass die $(\chi_{A_i})_{i\in I}$ unabhängig sind. Daraus folgt nach Definition, dass die $(\chi_{A_i}^{-1}(B_i))_{i\in I}$ für alle $B_i \in \mathfrak{B}$ vollständig unabhängig sind. Wir wählen speziell $B_i = \{1\}$ und es folgt, dass $\chi_{A_i}(\{1\}) = A_i$. Dies liefert, dass $(A_i)_{i\in I}$ (vollständig) unabhängig sind.

„$\Rightarrow$“ Wir setzen voraus, dass die $(A_i)_{i\in I}$ (vollständig) unabhängig sind und wissen, dass $\chi_{A_i}^{-1}(B_i) \in \{\emptyset, \Omega, A_i, A_i^c\}$. Die leere und die gesamte Menge müssen nicht betrachtet werden, da beide unabhängig sind. Somit genügt es zu zeigen, dass die $(c_i)_{i\in I}$ in den Fällen $c_i = A_i$ oder $c_i = A_i^c$ unabhängig sind. Dem Satz 4.3.2 folgend gilt hierfür die Behauptung. ∎

Bemerkung
Wegen dem Satz 5.2.1 werden folgend nur endliche Folgen $\xi_1, \dots, \xi_n$ von Zufallsvariablen betrachtet.

- Schreibweise: $\overline{\xi}(\omega) \colon = [\xi_1(\omega), \dots, \xi_n(\omega)] \Rightarrow \overline{\xi}$ ist Zufallsvariable mit Werten in $[\mathbb{R}^n, \mathfrak{B}^n]$.
  $\overline{\xi}$ heißt n-dimensionaler **zufälliger Vektor**.
- Frage: Ist $P_{\overline{\xi}}$ durch $P_{\xi_1}, \dots, P_{\xi_n}$ eindeutig bestimmt? Im allgemeinen nicht. Dies gilt nur im Spezialfall, dass $\xi_1, \dots, \xi_n$ unabhängig sind.

Satz 5.2.3
$(\xi_i)_{i=1}^n$ ist genau dann unabhängig, wenn $P_{\overline{\xi}} = \times_{i=1}^n P_{\xi_i}$

Beweis:
Der Beweis zeigt, dass der Satz auch im allgemeinen Fall von Zufallsvariablen $\xi_1, \dots, \xi_n$ gilt.

„$\Rightarrow$“ Wir haben die Unabhängigkeit von $\xi_1, \dots, \xi_n$ gegeben und es ist zu zeigen, dass $P_{\overline{\xi}}(\times_{i=1}^n B_i) = \prod_{i=1}^n P_{\xi_i}(B_i)$ mit $B_1, \dots, B_n \in \mathfrak{B}$. Dazu wissen wir:

$$\begin{aligned}
P_{\overline{\xi}}(\times_{i=1}^n B_i) &= P(\overline{\xi_1^{-1}}(\times_{i=1}^n B_i)) = P(\{\omega \in \Omega | \overline{\xi}(\omega) \in \times_{i=1}^n B_i\}) \\
&= P(„[\xi_1, \dots, \xi_n] \in \times_{i=1}^n B_i“) = P(„\xi_i \in B_i“) = P(\bigcap_{i=i}^n \{\omega \in \Omega | \xi_i(\omega) \in B_i\}) \\
&= P\left(\bigcap_{i=1}^n \xi_i^{-1}(B_i)\right) = \prod_{i=1}^n P(\xi_i^{-1}(B_i)) \\
&= \prod_{i=1}^n P_{\xi_i}(B_i)
\end{aligned}$$

„$\Leftarrow$“ Wir haben die Aussage, $P_{\overline{\xi}}(\times_{i=1}^n B_i) = \prod_{i=1}^n P_{\xi_i}(B_i)$ und es ist zu zeigen, dass $(\xi_i^{-1}(B_i))_{i=1}^n$ vollständig unabhängig sind. Das heißt, für $\{i_1, \ldots, i_k\} \subseteq \{i_1, \ldots, i_n\}$ und $B_{ij} \in \mathfrak{B}$ muss $P(\bigcap_{j=1}^k \xi_{ij}^{-1}(B_{ij})) = \prod_{j=1}^k P(\xi_{ij}^{-1}(B_{ij}))$ gelten. Wir setzen, $A_j := \mathbb{R}$ mit $j \notin \{i_1, \ldots, i_k\}$ und $A_{ij} := B_{ij}$ mit $j = 1, \ldots, k$.

$$\Rightarrow P_{\overline{\xi}}(\times_{i=1}^n A_i) = \prod_{i=1}^n P_{\xi_i}(A_i) \tag{5.1}$$

Wegen der obigen Feststellungen ist klar:

$$P_{\overline{\xi}}(\times_{i=1}^n A_i) = P\left(\bigcap_{i=1}^n \xi_i^{-1}(A_i)\right) = P\left(\bigcap_{j=1}^k \xi_i^{-1}(B_{ij})\right) \tag{5.2}$$

Falls $A_i = \mathbb{R}$, dann ist $\xi_i^{-1}(A_i) = \Omega$ und kann weggelassen werden.

$$P_{\xi_i}(A_i) = P(\xi_i^{-1}(A_i)) = \begin{cases} 1 & i \notin \{i_1, \ldots, i_k\} \\ P(\xi_i^{-1}(B_i)) & i \in \{i_1, \ldots, i_k\} \end{cases}$$

$$\Rightarrow \prod_{i=1}^k P_{\xi_i}(A_i) = \prod_{j=1}^k P(\xi_{ij}^{-1}(B_{ij})) = \prod_{j=1}^k P_{\xi_{ij}}(B_{ij}) \tag{5.3}$$

Aus den Gleichungen (Gleichung 5.1), (Gleichung 5.2) und (Gleichung 5.3) folgt die Behauptung. ■

Bemerkung
Als Folgerung lässt sich festhalten, dass die $\xi_1, \ldots, \xi_n$ genau dann unabhängig sind, wenn gilt $P(\bigcap_{i=1}^n \xi_i^{-1}(B_i)) = \prod_{i=1}^n P(\xi_i^{-1}(B_i))$ mit $B_1, \ldots, B_n \in \mathfrak{B}$.

Beispiel (Bernoullischema)
Sei $[\Omega, \mathcal{F}, P] = [\{0,1\}, \mathfrak{P}(\{0,1\}), B_{1,p}]^{n\times} = [\{0,1\}, \mathfrak{P}(\{0,1\}), p\delta_1 + (1-p)\delta_0]^{n\times} = [\{0,1\}^{n\times}, \mathfrak{P}(\{0,1\})^{n\times}, [p\delta_1 + (1-p)\delta_0]^{n\times}]$. Wir betrachten $\xi_k([\varepsilon_1, \ldots, \varepsilon_n]) = \varepsilon_k$ mit $\varepsilon_1, \ldots, \varepsilon_n \in \{0,1\}$. Hieraus folgt, dass $P_{\xi_k} = B_{1,p}$ und es ist klar, $\overline{\xi}([\varepsilon_1, \ldots, \varepsilon_n]) = [\varepsilon_1, \ldots, \varepsilon_n]$ (identische Abbildung). Somit folgt, $P_{\overline{\xi}} = (p\delta_1 + (1-p)\delta_0)^{n\times} = \times_{i=1}^n B_{1,p} = \times_{i=1}^n P_{\xi_i}$ und in Verbindung mit Satz 5.2.3 ergibt sich, dass die $\xi_1, \ldots, \xi_n$ unabhängig sind.

5.3. Summen unabhängiger Zufallsgrößen

Definition
Seien P_1, P_2 Wahrscheinlichkeitsmaße auf $[\mathbb{R}, \mathfrak{B}]$.

$$P_1 * P_2 := \int P_1(\underbrace{B - x}_{\{y-x|y\in B\}})P_2(dx) \qquad (B \in \mathfrak{B})$$

heißt **Faltung** von P_1 und P_2.

Bemerkung
Die Faltung $P_1 * P_2$ ist ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$. Denn die Eigenschaften lassen sich wie folgt nachweisen:

1) Da das Integral nie negativ werden kann, ist $P_1 * P_2(B) \geq 0$

2) $P_1 * P_2(\mathbb{R}) = \int P_1(\mathbb{R} - x)P_2(dx) = \int P_1(\mathbb{R})P_2(dx) = \int 1P_2(dx) = P_2(\mathbb{R}) = 1$
$P_1 * P_2(\emptyset) = \int P_1(\emptyset - x)P_2(dx) = \int P_1(\emptyset)P_2(dx) = \int 0P_2(dx) = 0$

3) Seien die $(B_i)_{i=1}^{\infty}$ paarweise disjunkt und $B_i \in \mathfrak{B}$. Dann sind auch die $(B_i - x)_{i=1}^{\infty}$ paarweise disjunkt und $\bigcup_{i=1}^{\infty}(B_i - x) = (\bigcup_{i=1}^{\infty} B_i) - x$. Somit folgt, $P_1 * P_2(\bigcup_{i=1}^{\infty} B_i) = \int P_1(\bigcup_{i=1}^{\infty}(B_1 - x))P_2(dx) = \int \sum_{i=1}^{\infty} P_1(B_i - x)P_2(dx) = \sum_{i=1}^{\infty} \int P_1(B_1 - x)P_2(dx) = \sum_{i=1}^{\infty} P_1 * P_2(B_i)$.

Satz 5.3.1
Seien ξ_1, ξ_2 unabhängige Zufallsgrößen. Dann gilt

$$P_{\xi_1+\xi_2} = P_{\xi_1} * P_{\xi_2}$$

Beweis:
Wir setzen $h(x_1, x_2) = x_1 + x_2$. Die x_1, x_2 sind aus $\mathbb{R}$ und somit ist die Abbildung h eine Abbildung von $\mathbb{R}^2$ nach $\mathbb{R}$ ($h \colon \mathbb{R}^2 \to \mathbb{R}$), die stetig und damit messbar ist.

$$\begin{aligned}
\xi_1 + \xi_2 = h([\xi_1, \xi_2]) = h([\overline{\xi}]) = h \circ \overline{\xi} \Rightarrow (\xi_1 + \xi_2)^{-1}(B) = \overline{\xi^{-1}}(h^{-1}(B))& \\
\Rightarrow P_{\xi_1+\xi_2}(B) = P \circ (\xi_1 + \xi_2)^{-1} = P((\xi_1 + \xi_2)^{-1}(B)) = P(\overline{\xi^{-1}}(h^{-1}(B))) = P \circ \overline{\xi}^{-1} \circ h^{-1}& \\
= P_{\overline{\xi}}(h^{-1}(B)) = P_{\overline{\xi}} \circ h^{-1}(B) = P_{\xi_1} \times P_{\xi_2}(h^{-1}(B)) = P_{\overline{\xi}}(\{[x_1, x_2] | x_1 + x_2 \in B\})& \\
= \int \chi_{\{[x_1,x_2]|x_1+x_2 \in B\}}(y_1, y_2) P_{\overline{\xi}}(d[y_1, y_2]) = \int \chi_{B-y_2}(y_1) P_{\overline{\xi}}([y_1, y_2]) = (P_{\xi_1} \times P_{\xi_2}) \circ h^{-1}& \\
= \iint \chi_{B-y_2}(y_1) P_{\xi_1}(dy_1) P_{\xi_2}(dy_2) = \int P_{\xi_1}(B - y_2) P_{\xi_2}(dy_2)& \\
= P_{\xi_1} * P_{\xi_2}(B)&
\end{aligned}$$

■

Bemerkung
$*$ ist kommutativ und assoziativ (siehe auch Satz 5.3.1). Somit ist für $P_1, \ldots, P_n$:

$$*_{i=1}^{n} P_i = (((P_1 * P_2) * P_3) * \cdots * P_n) \tag{5.4}$$

Satz 5.3.2 (Verallgemeinerung von Satz 5.3.1)
Seien $\xi_1, \ldots, \xi_n$ unabhängige Zufallsgrößen:

$$\Rightarrow P_{\sum_{i=1}^{n} \xi_i} = *_{i=1}^{n} P_i$$

Beweis:
Der Beweis erfolgt durch vollständige Induktion. Für $n = 2$ ist die Behauptung nach Satz 5.3.1 richtig und wir nehmen an, dass die Gleichung (Gleichung 5.4) für gewisse n gilt. Dann ist zu zeigen, dass, wenn $\xi_1, \ldots, \xi_i$ unabhängige Zufallsgrößen sind, folgt, $*_{i=1}^{n} P_i = (((P_1 * P_2) * P_3) * \cdots * P_n)$. Seien $\xi_1, \ldots, \xi_n$ unabhängig und $[\xi_1, \ldots, \xi_n], \xi_{n+1}$ unabhängig sowie $\sum_{i=1}^{n} \xi_i, \xi_{n+1}$ ebenfalls unabhängig*. Dann folgt, $P_{\sum_{i=1}^{n} \xi_i} = *_{i=1}^{n} P_{\xi_i}$. Unter Anwendung von Satz 5.3.1 ergibt sich, $P_{\sum_{i=1}^{n} \xi_i + \xi_{i+1}} = *_{i=1}^{n} P_{\xi_i} * P_{\xi_{n+1}} \Rightarrow P_{\sum_{i=1}^{n+1}} = *_{i=1}^{n+1} P_{\xi_i}$ ∎

Satz 5.3.3
Seien ξ_1, ξ_2 unabhängige Zufallsgrößen. Dann gilt:

$$F_{\xi_1+\xi_2}(y) = \int F_{\xi_1}(y-x) P_{\xi_2}(dx)$$

Beweis:
Wir wenden den Satz 5.3.1 mit $B = (-\infty, y)$ an und haben:

$$\begin{aligned} F_{\xi_1+\xi_2} &= P_{\xi_1+\xi_2}((-\infty, y)) = P_{\xi_1} * P_{\xi_2}((-\infty, y)) \\ &= \int P_{\xi_1}((-\infty, y) - x) P_{\xi_2}(dx) = \int P_{\xi_1}((-\infty, y - x)) P_{\xi_2}(dx) \\ &= \int F_{\xi_1}(y - x) P_{\xi_2}(dx) \end{aligned}$$ ∎

Satz 5.3.4
Seien ξ_1, ξ_2 unabhängige Zufallsgrößen mit $P_{\xi_i} = \ell f_i$ $(i = 1, 2)$ und f_1, f_2 stetig. Dann gilt:

$$P_{\xi_1+\xi_2} = \ell(f_1 * f_2)$$

wobei $f_1 * f_2(y) \colon = \int_{-\infty}^{\infty} f_1(y-x) f_2(x) \ell(dx)$.

Beweis:
Der Beweis erfolgt weniger mathematisch streng. Da der Aufwand hierfür unangemessen hoch ist. Für streng mathematische Rechnungen sei der Leser an weitere Lektüre verwiesen.

Wegen Satz 5.3.2 ist $F_{\xi_1+\xi_2}(x) = \int F_1(x-y) P_{\xi_2}(dy)$ und wegen $P_{\xi_2} = \ell f_2$ ergibt sich $\int F_1(x-y) f_2(y) \ell(dy)$†. Man geht nun davon aus, dass die Funktion differenzierbar ist. Alle Fälle, wo die Funktion nicht überall differenzierbar ist, werden nicht mit abgedeckt.

Somit folgt, dass $F'_{\xi_1+\xi_2}(x) = \int F'_1(x-y) f_2(y) \ell(dy)$ die Dichte von $P_{\xi_1+\xi_2}$ ist (nach Definition). Dies ist aber $\int f_1(x-y) f_2(y) \ell(dy) = f_1 * f_2(x)$ ∎

*Der Beweis hierfür findet sich in der gängigen Lektüre zum Thema.
†Der Sachverhalt wurde früher bewiesen.

Satz 5.3.5
Seien ξ_1, ξ_2 unabhängige Zufallsgrößen mit $P_{\xi_i}(\mathbb{Z}) = 1$ $(i = 1, 2)$, d. h. $P_{\xi_1} = \sum_{k\in\mathbb{Z}} a_k\delta_k$, $P_{\xi_2} = \sum_{k\in\mathbb{Z}} b_k\delta_k$, $a_k, b_k \geq 0$, $\sum a_k = \sum b_k = 1$. Dann gilt:

$$\begin{aligned} P_{\xi_1+\xi_2}(\mathbb{Z}) &= 1 \\ P_{\xi_1+\xi_2}(\{n\}) &= \sum_{k\in\mathbb{Z}} P_{\xi_1}(\{n-k\}) P_{\xi_2}(\{k\}) \end{aligned}$$

d. h. $P_{\xi_1+\xi_2} = \sum_{k\in\mathbb{Z}} c_k\delta_k$ mit $c_k = \sum a_{n-k}b_k$.

Beweis:
Nach Satz 5.3.1 wissen wir, dass $P_{\xi_1+\xi_2}(B) = \int P_{\xi_1}(B-x)P_{\xi_2}(dx)$ ist und wählen $B = \{n\}$. Somit erhalten wir nun, $P_{\xi_1+\xi_2}(\{n\}) = \int P_{\xi_1}(\{n-x\})P_{\xi_2}(dx) = \sum_{k\in\mathbb{Z}} P_{\xi_1}(\{n-k\})P_{\xi_2}(\{k\})$. ∎

Beispiel
Seien ξ_1, ξ_2 unabhängig und $P_{\xi_i} = B_{n_i,p}$ mit $i = 1, 2$. Zur Bestimmung von $P_{\xi_1+\xi_2}$ wäre es jetzt möglich, Satz 5.3.5 anzuwenden:

$$\begin{aligned} P_{\xi_1+\xi_2}(\{n\}) &= \sum_{k\in\mathbb{Z}} P_{\xi_1}(\{n-k\})P_{\xi_2}(\{k\}) = \sum_{k\in\mathbb{Z}} B_{n_1,p}(\{n-k\})B_{n_2,p}(\{k\}) \\ &= \sum_{0\leq k\leq n_2} \binom{n_1}{n-k} p^{n-k}(1-p)^{n_1-(n-k)} \binom{n_2}{k} p^k (1-p)^{n_2-k} \\ &= \begin{cases} \binom{n_1+n_2}{n} p^n (1-p)^{n_1+n_2-n} & 0 \leq n \leq n_1+n_2 \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

$$= B_{n_1+n_2,p}(\{n\})$$

Diese Rechnung ist relativ kompliziert und es existieren weitere Möglichkeiten zur Bestimmung. Wir betrachten das Bernoullischema zu $n_1 + n_2, p$ und haben durch den Wahrscheinlichkeitsraum $[\{0,1\}, \mathfrak{P}(\{0,1\}), B_{1,p}]^{n_1+n_2}$ ein Bernoullischema gegeben. Dabei stellen $\hat{\xi_1} := \sum_{i=1}^{n_1} \pi_i$ und $\hat{\xi_2} := \sum_{i=n_i+1}^{n_1+n_2} \pi_i$ die Anzahl der Erfolge im Bernoullischema dar. Wir wissen, dass $P_{\hat{\xi_1}} = B_{n_1,p}, P_{\hat{\xi_2}} = B_{n_2,p}, \hat{\xi_1}, \hat{\xi_2}$ unabhängig sind. Daraus folgt: $P_{\xi_1+\xi_2} = P_{\hat{\xi_1}+\hat{\xi_2}}$ und es ist klar, dass $\hat{\xi_1} + \hat{\xi_2} = \sum_{i=1}^{n_1+n_2} \pi_i \Rightarrow P_{\hat{\xi_1}+\hat{\xi_2}} = B_{n_1+n_2,p}$

5.4. Der Erwartungswert

Im folgenden sind immer $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum und wir betrachten die Zufallsgrößen ξ, η, ζ.

Definition
Sei ξ eine Zufallsgröße, so dass $E\xi := \int \xi(\omega)P(d\omega)$ existiert. Dann heißt $E\xi$ **Erwartungswert** von ξ.

Bemerkung

- In älteren Büchern wird $E\xi$ auch manchmal als $M\xi$ bezeichnet.
- Ist ξ bezüglich P integrierbar, so ist $E\xi$ endlich. Man sagt, ξ hat einen endlichen Erwartungswert.
- $E\xi$ existiert, wenn $\xi \geq 0$ oder ξ P-integrierbar ist. Der letztere stellt einen wichtigen Fall dar.
- $E\xi$ ist ein fundamentales, aber grobes Kennzeichen der Verteilung P_ξ. Die Definition lässt sich allein mit P_ξ aufschreiben.

Satz 5.4.1
Sei ξ eine Zufallsgröße mit endlichen Erwartungswert[‡]. Dann gilt:

$$E\xi = \int x P_\xi(dx) = \left(\int \iota dP_\xi\right)$$

Beweis:
Zum Beweis kommt der Übertragungssatz (Satz 2.2.3) zur Anwendung. Wir haben $[\Omega_1, \mathcal{F}_1, \mu] := [\Omega, \mathcal{F}, P], [\Omega_2, \mathcal{F}_2] = [\mathbb{R}, \mathfrak{B}], h\colon [\Omega_1, \mathcal{F}_1] \to [\Omega_2, \mathcal{F}_2], f\colon [\Omega_2, \mathcal{F}_2] \to [\mathbb{R}, \mathfrak{B}]$. Speziell sind $h = \xi$ und $f = \iota$. Somit folgt, $\int f \circ h d\mu = \int f d\mu \underbrace{\circ h^{-1}}_{P \circ \xi^{-1} = P_\xi}$ und in unserem Fall: $\int \xi dP = \int \iota dP_\xi$ ∎

Satz 5.4.2
Sei $P_\xi(\mathbb{Z}) = 1$[§] und ξ eine Zufallsgröße mit $P_\xi = \sum_{k\in\mathbb{Z}} a_k \delta_k$. Falls $E\xi$ absolut konvergiert, gilt:

$$E\xi = \sum_{k\in\mathbb{Z}} k P\xi(\{k\}) = \sum_{k\in\mathbb{Z}} k a_k$$

Beweis:
Aus Satz 5.4.1 folgt, $E\xi = \int x P_\xi(dx) = \sum_{k\in\mathbb{Z}} k a_k$. ∎

Satz 5.4.3
Sei $P_\xi = \ell f$ mit f stetig. Falls der Erwartungswert $E\xi$ existiert, gilt:

$$E\xi = \int_{-\infty}^{\infty} x f(x) dx$$

Beweis:
Aus Satz 5.4.1 folgt, $E\xi = \int x P_\xi(dx) = \int x f(x) \ell dx = \int_{-\infty}^{\infty} x f(x) dx$. ∎

[‡] Die Aussage gilt auch für unendlichen Erwartungswert. Doch der hierzu zu führende Beweis ist deutlich aufwendiger.

[§] diskrete Verteilung, ξ nimmt nur ganze Zahlen an.

Satz 5.4.4
Sei ξ eine nichtnegative Zufallsgröße, d. h. $P(„\xi \geq 0“) = 1 \Leftrightarrow P_\xi([0, +\infty)) = 1$. Dann gilt:

$$E\xi = \int\limits_{[0,+\infty)} (1 - F_\xi(x))\ell dx$$

Beweis:
Nach Satz 5.4.1 ist $E\xi = \int_{[0,x]} 1\ell(dy) = \int_{[0,+\infty)} xP_\xi(dx) = \int_{[0,+\infty)}(\int_{[0,x)} 1\ell(dy))P_\xi(dx) =$
$\int \chi_{[0,+\infty)}(x) \int \chi_{[0,x)}(y)\ell(dy)P_\xi(dx) = \iint \chi_{\{[z_1,z_2]\in\mathbb{R}^2|0\leq z_2<z_1\}}(x, y)\ell(dy)P_\xi(dx) = \iint \chi_{\{[z_1,z_2]\in\mathbb{R}^2|0\leq z_2<z_1}$
$\iint \chi_{[0,\infty)}(y)\chi_{[y,\infty)}(x)P_\xi(dx)\ell(dy) = \int \chi_{[0,\infty)}(y)(\int \chi_{[y,\infty)}(x)P_\xi(dx))\ell(dy) = \int \chi_{[0,\infty)}(y)P_\xi([y, \infty))\ell(dy)$
$\int \chi_{[0,\infty)}(y)(1 - F_\xi(y))\ell(dy) = \int_{[0,\infty)}(1 - F_\xi(y))\ell(dy)$. ■

Definition
Sei ξ eine Zufallsgröße, so dass ein endlicher $E\xi^k$ existiert. Dann heißt $m_k(\xi) = E\xi^k$ k-tes **Moment** der Zufallsgröße ξ.

Momentenproblem Ist P_ξ durch die Folge $(m_k(\xi))_{k=1}^n$ eindeutig bestimmt? Diese Frage ist negativ zu beantworten. Nach dem Übertragungssatz ist klar, dass $E\xi^k = \int x^k P_\xi(dx)$ gilt. Aus der Kenntnis von P_ξ folgt somit die Kenntnis des Moments. Aber im allgemeinen ist $E\xi^k \not\Rightarrow P_\xi$. Eine positive Antwort auf die obige Frage ist nur unter gewissen Beschränktheitsforderungen möglich.

Satz 5.4.5 (Elementare Eigenschaften des Erwartungswerts)
(E1) Seien $E\xi_1, E\xi_2$ endlich. Dann folgt: $E(a_1\xi_1 + a_2\xi_2) = a_1E\xi_1 + a_2E\xi_2$ mit $a_1, a_2 \in \mathbb{R}$.

(E2) Sei $\xi = \chi_A, A \in \mathcal{F} \Rightarrow E\xi = P(A)$

(E3) Sei $P(„\xi = c“) = 1 \Leftrightarrow P_\xi = \delta_c \Rightarrow E\xi = c$

(E4) Sei $P(„a \leq \xi \leq b“) = 1 \Rightarrow a \leq E\xi \leq b$

Beweis:
(E1) $E(a_1\xi_1 + a_2\xi_2) = \int(a_1\xi_1 + a_2\xi_2)dP = a_1 \int \xi_1 dP + a_2 \int \xi_2 dP = a_1E\xi_1 + a_2E\xi_2$

(E2) $E\chi_A = \int \chi_A dP = P(A)$

(E3) $P(„\xi = c“) = 1 \Rightarrow P_\xi = \delta_c$
Nach Satz 5.4.1 folgt nun $E\xi = \int xP_\xi(dx) = \int x\delta_c(dx)$ und letzteres ist ein Integral bei diskreten Maßen. Damit ergibt sich nach Satz 5.4.2, dass die letzte Aussage gleich $\sum_{k\in\mathbb{R}} 1 \cdot \delta_c(\{k\}) = c$ ist. Im allgemeinen Fall gilt $\int f(x)\delta_c(dx) = f(x)$. help: muss das nicht eigentlich f(c) sein? Noch allgemeiner formuliert: $\int f(x) \sum_k a_k\delta_{\omega_k}(dx) = \sum_k a_k f(\omega_k)$.

(E4) $P(„a \leq \xi \leq b“) = 1 \Rightarrow P_\xi([a, b]) = 1 \Rightarrow E\xi = \int xP_\xi(dx) = \int_{[a,b]} xP_\xi(dx) \geq \int_{[a,b]} aP_\xi(dx) = aP([a, b]) = a$ ■

Beispiel

1. Bernoulli-Verteilung: $P_\xi = B_{1,p} = p\delta_1 + (1-p)\delta_0 \Rightarrow E\xi = 1p + 0(1-p) = p$
2. $P_\xi = B_{n,p} = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k}\delta_k \Rightarrow E\xi = \sum_{l=0}^n k\binom{n}{k}p^k(1-p)^{n-k} = np$
   Ein anderer Weg, um dies zu ermitteln: Sei $[\Omega, \mathcal{F}, P] = [\{0,1\}, \mathfrak{P}(\{0,1\}), B_{1,p}]^{n\times}$ und ξ_k die Projektion auf die k-te Ebene in Ω. Es ist bekannt, dass $P_{\xi_k} = B_{1,p}$ und $P_{\sum \xi_k} = B_{n,p} = P_\xi$. Damit folgt, $E\xi = E\sum_{k=1}^n \xi_k = \sum_{k=1}^n E\xi_k = \sum_{k=1}^n p = np$.
3. Cauchy-Verteilung: $P_\xi = \ell f$ mit $f(x) = \frac{1}{\pi(1+x^2)}$ und $x \in \mathbb{R}$. Es ist klar, dass $\int_0^{+\infty} xf(x)dx = \infty$ gilt.

   Ein anderer Weg ist folgender: Sei $[\Omega, \mathcal{F}, P] = [\{0,1\}, \mathfrak{P}(\{0,1\}), B_{1,p}]^{n\times}$ und ξ_k die Projektion auf die k-te Komponente in Ω. Es ist bekannt, dass $P_{\xi_k} = B_{1,p}$ und $P_{\sum \xi_k} = B_{n,p} = P_\xi$. Somit folgt, $E\xi = E\sum_{k=1}^n \xi_k = \sum_{k=1}^n E\xi_k = \sum_{k=1}^n p = np$.
4. Sei $P\xi = \ell f$ mit $f(x) = \frac{1}{\pi(1+x)^2}$ für $x \in \mathbb{R}$. Es ist klar, dass $\int_0^{+\infty} xf(x)dx = +\infty$ gilt. Formal wäre $E\xi = +\infty$↯. Damit folgt, dass $E\xi$ nicht existiert.
5. $P_\xi = \prod_\lambda \Rightarrow E\xi = \lambda$
6. $P_\xi = \mathrm{Ex}_\lambda \Rightarrow E\xi = 1/\lambda$

Satz 5.4.6
Seien ξ, η unabhängige Zufallsgrößen mit endlichem Erwartungswert. Dann hat die Zufallsgröße $\xi \cdot \eta$ einen endlichen Erwartungswert und es gilt:

$$E(\xi \cdot \eta) = E\xi \cdot E\eta$$

Beweis:
Spezialisierung des Übertragungssatzes:

$$\begin{aligned} [\Omega_1, \mathcal{F}_1, \mu] &:= [\Omega, \mathcal{F}, P] & [\Omega_2, \mathcal{F}_2]&\colon = [\mathbb{R}^2, \mathfrak{B}_2] \\ h\colon \Omega_1 &\to \Omega_2 & h &= [\xi, \eta] \\ f\colon \Omega_2 &\to \mathbb{R} & f(x,y) &= xy \end{aligned}$$

$$\begin{aligned} \Rightarrow \int f \circ h d\mu &= \int f d\mu \circ h^{-1} \\ \int \xi \cdot \eta dP &= \int xy dP_{[\xi,\eta]} d([x,y]) \\ \Rightarrow f \circ h &= \xi \cdot \eta, \mu \circ h^{-1} = P \circ h^{-1} = P_{[\xi,\eta]} = P_\xi \times P_\eta \\ \Rightarrow E(\xi \cdot \eta) &= \int xy dP_{[\xi,\eta]} d([x,y]) = \int xy(P_\xi \times P_\eta)(d[x,y]) \\ &= \iint xy P_\xi(dx) P_\eta(dy) = \int x P_\xi(dx) \cdot \int y P_\eta(dy) \\ &= E\xi \cdot E\eta \end{aligned}$$

■

Bemerkung
Ohne die Forderung der Unabhängigkeit von ξ und η ist die Behauptung im obigen Satz im allgemeinen nicht richtig.

5.5. Die Varianz

Sei $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum. Wir betrachten die Zufallsgrößen ξ, η.

Definition
Sei ξ eine Zufallsgröße mit *endlichem* Erwartungswert.

$$D\xi :\, = E(\xi - E\xi)^2 = \int (\xi - E\xi)^2 \, dP$$

heißt **Varianz** von ξ. Weiterhin heißt $\sqrt{D\xi}$ **Standardabweichung** oder **Streuung** von ξ.

Satz 5.5.1 (Elementare Eigenschaften der Varianz)
a) $D\xi \geq 0$

b) $D\xi = 0 \Leftrightarrow P_\xi = \delta_{E\xi}$. d. h. $P(„\xi = c“) = 1$, also Determinismus.

c) $D\xi$ endlich $\Leftrightarrow E\xi^2$ endlich. In diesem Fall gilt $D\xi = E\xi^2 - (E\xi)^2$

d) $D(c_1\xi + c_2) = c_1^2 D\xi$ mit $c_1, c_2 \in \mathbb{R}$ (Varianz einer linearen Transformation)

e) $D\xi = \int (x - E\xi)^2 P_\xi(dx)$

Beweis:
a) klar, weil $(\xi - E\xi)^2 \geq 0 \Rightarrow \int (\xi - E\xi)^2 dP \geq 0$

b) „$\Leftarrow$“ $P_\xi = \delta_{E\xi} \Rightarrow D\xi = \int (x - E\xi)^2 P_\xi(dx) = (E\xi - E\xi)^2 = 0$

„$\Rightarrow$“ Sei $\eta \geq 0, E\eta = 0 \Rightarrow P(„\eta = 0“) = 1$.

c) $D\xi = E(\xi - E\xi)^2 = E(\xi)^2 - 2e\xi E\xi + (E\xi)^2) = E\xi^2 - 2E\xi E\xi + (E\xi)^2 = E\xi^2 - 2(E\xi)^2 + (E\xi)^2 = E\xi^2 - (E\xi)^2$

d) $D(c_1\xi + c_2) = E(c_1\xi + c_2 - E(c_1\xi + c_2))^2 = E(c_1\xi - c_1 E\xi)^2 = E(c_1(\xi - E\xi))^2 = c_1^2 E(\xi - E\xi)^2 = c_1^2 D\xi$

e) $D\xi = E(\xi - E\xi)^2 = \int (\xi - E\xi)^2 dP$. Nach dem Übertragungssatz folgt mit $h = \xi, f(x) = (x - E\xi)^2, \mu = P$ sowohl $f \circ h = (\xi - E\xi)^2$ als auch $\mu \circ h^{-1} = P_\xi$. Weiter folgt somit $E(\xi - E\xi)^2 = \int f(x) P_\xi(dx) = \int (x - E\xi)^2 P_\xi(dx)$. Daraus ergibt sich die Behauptung. ∎

Bemerkung (Folgerungen)
1. Aus den Punkten a) und c) ergibt sich: $E\xi^2 \geq (E\xi)^2$

2. Aus c) ergibt sich: $D\xi = 0 \Leftrightarrow E\xi^2 = (E\xi)^2 \Leftrightarrow P_\xi = \delta_{E\xi}$

Beispiel (für $E\xi \cdot \eta \neq E\xi \cdot E\eta$)
Sei ξ eine Zufallsgröße mit $D\xi > 0$, d. h. ξ ist keine Konstante, und $\eta :\, = \xi$. Dann folgt, $E\xi \cdot \eta = E\xi^2$. Da weiter gilt, $D\xi = E\xi^2 - (E\xi)^2$ und $D\xi \neq 0$, folgt, dass $E\xi^2 \neq E\xi \cdot E\xi$. Angenommen, es gelte, $E\xi \cdot \eta = E\xi \cdot E \cdot \eta$. Dann wäre $E\xi^2 = (E\xi)^2 \Rightarrow D\xi = E\xi^2 - (E\xi)^2 = 0$ ↯

Satz 5.5.2
Seien ξ, η unabhängige Zufallsgrößen mit endlicher Varianz. Dann gilt:

$$D(\xi + \eta) = D\xi + D\eta$$

Beweis:
Es ist klar, dass aus ξ, η unabhängig auch $\xi - E\xi, \eta - E\eta$ unabhängig folgt.

$$\begin{aligned} D(\xi+\eta) &= E(\xi+\eta-E(\xi+\eta))^2 = E((\xi-E\xi)+(\eta-E\eta))^2 \\ &= E\left((\xi-E\xi)^2 + 2(\xi-E\xi)(\eta-E\eta) + (\eta-E\eta)^2\right) \\ &= E(\xi-E\xi)^2 + E(\eta-E\eta)^2 + 2\underbrace{E(\xi-E\xi)}_{0=} \cdot \underbrace{E(\eta-E\eta)}_{=0} = D\xi + D\eta \end{aligned}$$ ■

Bemerkung
Ohne die Forderung der Unabhängigkeit ist die Behauptung im allgemeinen falsch.

Weiterhin kann man per vollständiger Induktion zeigen, dass für $\xi_1, \ldots, \xi_n$ unabhängige Zufallsgrößen mit endlichen Erwartungswerten folgt,

$$D\left(\sum_{k=1}^{n} \xi_k\right) = \sum_{k=1}^{n} D\xi_k$$

Satz 5.5.3
a) Sei $P_\xi = \sum_{k\in\mathbb{Z}} a_k \delta_k$ und $E\xi$ endlich. Dann gilt:

$$D\xi = \sum_{k\in\mathbb{Z}} (k - E\xi)^2 a_k = \sum_{k\in\mathbb{Z}} (k - \sum_{j\in\mathbb{Z}} j a_j)^2 a_k = \sum_{k\in\mathbb{Z}} k^2 a_k - (\sum_{k\in\mathbb{Z}} k a_k)^2$$

b) Sei $P_\xi = \ell f$ mit f stetig und $E\xi$ endlich. Dann ist:

$$D\xi = \int_{-\infty}^{+\infty} (x - E\xi)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(\int_{-\infty}^{+\infty} x f(x)dx\right)^2$$

Beispiel
1. $P_\xi = B_{1,p} = p\delta_1 + (1-p)\delta_0 \Rightarrow D\xi = (0-p)^2(1-p) + (1-p)^2 p = p^2(1-p) + p - 2p^2 + p^3 = p - p^2 = p(1-p)$

2. $P_\xi = B_{n,p}$
   Seien $\xi_1, \ldots, \xi_n$ unabhängig und $P_{\xi_k} = B_{1,p}$ für $k = 1, \ldots, n$. Dann gilt $P_{\sum \xi_k} = B_{n,p} \Rightarrow P_\xi = P_{\sum \xi_k} \Rightarrow D\xi = D\sum_{k=1}^n \xi_k = \sum_{k=1}^n D\xi_k = np(1-p)$. Andererseits gilt auch: $D\xi = \sum_{k=0}^n (k-np)^2 p^k (1-p)^{n-k}\binom{n}{k} = np(1-p)$

Interpretation der Varianz Sei $(H, \|\cdot\|)$ ein **Hilbertraum**. Dabei ist H ein linearer Raum und $\|\cdot\|$ die Norm. Weiter seien $h \in H, G \subseteq H$.

Wir betrachten den Spezialfall:

$$H\colon = \left\{ f[\Omega, \mathcal{F}] \to [\mathbb{R}, \mathfrak{B}] \middle| \|f\| = \sqrt{\int f^2 dP} < +\infty \right\}$$

Dabei ist f eine Zufallsgröße und Ef ist endlich, denn es gilt, $\infty > Ef^2 \geq (Ef)^2$. Damit ist auch Df endlich und existiert. Weiter sei:

$$G\colon = \{a\colon \Omega \to \mathbb{R} | \text{konstant}\}$$

und ξ eine Zufallsgröße mit $E\xi^2 < +\infty$. Es ist klar, dass $\|\xi - a\| = \sqrt{E(\xi-a)^2}$

Satz 5.5.4
Sei ξ eine Zufallsgröße mit $E\xi^2 < \infty$. Dann gilt:

$$D\xi = \inf_a E(\xi - a)^2 = \min_a E(\xi - a)^2$$

Dabei ist $\sqrt{D\xi}$ der Abstand einer Zufallsgröße zur Menge der Konstanten mit $\sqrt{D\xi} = \inf_a \sqrt{E(\xi-a)^2} = \inf_a \|\xi - a\|$. Wenn der Abstand 0 ist, ist somit ξ konstant. Für einen kleinen Abstand bedeutet dies, dass der Zufall gering ist.

Beweis:

$$\begin{aligned}
E(\xi - a)^2 &= E((\xi - E\xi) + (E\xi - a))^2 \\
&= E((\xi - E\xi)^2 + E(\xi - E\xi)(E\xi - a) + (E\xi - a)^2) \\
&= D\xi + E(E\xi - a)\underbrace{E(\xi - E\xi)}_{=0} + (E\xi - a)^2 \\
&= D\xi + (E\xi - a)^2 \\
\Rightarrow \min_a E(\xi - a)^2 &= \min_a D\xi + (E\xi - a)^2 \Leftrightarrow a = E\xi \\
\Rightarrow D\xi &= \min_a E(\xi - a)^2
\end{aligned}$$

■

Bemerkung
Sei ξ eine Zufallsgröße mit $E\xi, E\xi^2 < +\infty$ und endlich.

- $\xi - E\xi$ heißt **Zentrierung** von ξ. Die Zufallsgröße η heißt zentriert, wenn $E\eta = 0$.
- Sei $D\xi > 0$. Dann heißt $\frac{\xi}{\sqrt{D\xi}}$ **Normierung** von ξ. Die Zufallsgröße η heißt normiert, wenn $D\eta = 1$.
- Wir betrachten Normierung und Zentrierung: $\eta\colon = \frac{\xi - E\xi}{\sqrt{D\xi}}$. Dann ist:

$$E\eta = \frac{E(\xi - E\xi)}{\sqrt{D\xi}} = 0$$
$$D\eta = D\frac{\xi - E\xi}{\sqrt{D\xi}} = \frac{1}{D\xi}D(\xi - E\xi) = \frac{1}{D\xi}D\xi = 1$$

5.6. Die Tschebyscheffsche Ungleichung

Sei $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum. Wir betrachten Zufallsgrößen über diesen Wahrscheinlichkeitsraum.

Satz 5.6.1 (Markoffsche Ungleichung)
Sei $\xi \geq 0$ eine nichtnegative Zufallsgröße[¶] und $a \geq 0$ eine reelle Zahl. Dann gilt:

$$E\xi \geq aP(„\xi \geq a“) = aP_\xi([a, +\infty]) = a(1 - F_\xi(a))$$

Beweis:
Sei $\xi \geq 0$. Dann ist $\xi(\omega) \geq \xi(\omega)\chi_{\{\hat\omega|\xi(\hat\omega)\geq a\}}(\omega) \geq a\chi_{\{\hat\omega|\xi(\hat\omega)\geq a\}}(\omega)$. Da das Integral monoton ist, folgt $E\xi = \int \xi(\omega)P(d\omega) \geq \int a\chi_{\{\hat\omega|\xi(\hat\omega)\geq a\}}dP = a\int \chi_{\{\hat\omega|\xi(\hat\omega)\geq a\}}dP = aP(\{\hat\omega|\xi(\hat\xi) \geq a\}) = aP(„\xi \geq a“)$. ■

Satz 5.6.2 (Tschebyscheffsche Ungleichung)
Sei η eine Zufallsgröße mit endlichem Erwartungswert. Dann existiert $D\eta$ und es gilt:

$$P(„|\eta - E\eta| \geq \varepsilon“) \leq \frac{D\eta}{\varepsilon^2} \qquad \varepsilon > 0$$

Beweis:
Zum Beweis wird der Satz 5.6.1 für folgenden Spezialfall angewendet. Wir setzen $\xi := (\eta - E\eta)^2 \geq 0$ und $a\colon = \varepsilon^2$. Dann ist klar, dass $E\xi = D\eta$ gilt. Dann ergibt sich: $D\eta = E\xi \geq aP(„\xi \geq a“) = \varepsilon^2 P(„(\eta - E\eta)^2 \geq \varepsilon^2“) = \varepsilon^2 P(„|\eta - E\eta| \geq \varepsilon“)$. ■

Satz 5.6.3
Seien $\xi_1, \ldots, \xi_n$ eine Folge unabhängiger Zufallsgrößen mit $E\xi_i = a, D\xi_i = \sigma^2$ für $i = 1, \ldots, n$. Dann gilt:

$$P\left(„|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a| \geq \varepsilon“\right) \leq \frac{\sigma^2}{n\varepsilon^2} \qquad \varepsilon > 0$$

[¶]Damit existiert $\int \xi d\mu$ nach der Definition zweiter Stufe immer.

Beweis:
Für den Beweis wird Satz 5.6.2 angewendet und wir setzen: $\eta\colon = \frac{1}{n}\sum_{i=1}^{n}\xi_i$

$$\Rightarrow E\eta = E\left({}^1\!/\!_n\sum_i \xi_i\right) = {}^1\!/\!_n\sum_i \underbrace{E\xi_i}_{=a} = a\,{}^1\!/\!_n n = a$$

$$D\eta = D\left({}^1\!/\!_n\sum_i \xi_i\right) = {}^1\!/\!_{n^2}\sum_i \underbrace{D\xi_i}_{=\sigma^2} = {}^1\!/\!_{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$\Rightarrow P\left(„|{}^1\!/\!_n\sum \xi_i - a| \geq \varepsilon“\right) = P(„|\eta - E\eta| \geq \varepsilon“) \leq= \frac{D\eta}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$ ■

Anwendung von Satz 5.6.3 auf Wette

- $n = 250$ Münzwürfe mit $\xi_1, \ldots, \xi_{250}$, Bernoullischema mit $n = 250, p = {}^1\!/\!_2$.
- $\xi_i = 1$ entspricht der Aussage, dass der i-te Wurf Wappen zeigt.
- Damit sind die $\xi_1, \ldots, \xi_n$ unabhängig
- $P_{\xi_i} = B_{1,p} = {}^1\!/\!_2(\delta_1 + \delta_0)$ mit $p = {}^1\!/\!_2 \Rightarrow E\xi_i = {}^1\!/\!_2, D\xi_i = {}^1\!/\!_2(1 - {}^1\!/\!_2) = {}^1\!/\!_4$
- $|\sum_{i=1}^{250}\xi_i - 125| < 25 =: A$ entspricht der Aussage, dass ich gewinne.

Aus Satz 5.6.3 folgt nun: $P(A) = P(„|{}^1\!/\!_n\sum_{i=1}^{n}\xi_i - a| \geq \varepsilon“) < \frac{\sigma^2}{n\varepsilon^2} = 1 - P(„|{}^1\!/\!_2\sum \xi_i - a| \geq \varepsilon“) = 1 - {}^1\!/\!_{10} = 0,9$

(reale) Anwendung — Meßtheorie

Viele Messgeräte arbeiten nicht exakt. Die Ursachen liegen in systematischen Fehlern und durch zufällige Einflüsse bedingte Fehler. Die systematischen Fehler können durch Eichung minimiert oder abgeschafft werden.

Die ideale Messapparatur liefert der Wert a und der reale Vorgang die zufällige Größe ξ. Da das Gerät zwar korrekt misst, aber auch anfällig für äußere Einflüsse ist, ergibt sich ein nicht systematischer Fehler, d. h. $E\xi = a$. Vom Hersteller erhält man $D\xi = \sigma^2$. Die übliche Verfahrensweise ist nun, dass wiederholte Messungen durchgeführt werden und man so viele $\xi_1, \ldots, \xi_n$ bekommt. Für diese wird angenommen, dass sie unabhängig und identisch verteilt (iid‖) sind und es gilt:

$$D\xi_i = \sigma^2 \qquad\qquad E\xi_i = a$$

Als Verfahren wird das arithmetische Mittel als „Schätzwert“ für a gebildet:

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i = \eta_n$$

‖vom englischen independent and identical distributed

Folgende Ansprüche werden formuliert:

1. Die Schätzung soll „hinreichend“ exakt sein, d. h. für ein vorgegebenes $\varepsilon > 0$ wird gefordert: $|\eta_n - a| < \varepsilon$.

2. Vorgabe eines Sicherheitsniveaus $s < 1$ mit

$$P(„|\eta_n - a| < \varepsilon“) \geq s \tag{5.5}$$

Damit stellt sich das Problem, wie n gewählt werden muss, so dass die Gleichung (Gleichung 5.5) gilt. Eine mögliche Antwort ist:

$$n \geq \frac{\sigma^2}{(1-s)\varepsilon^2} \tag{5.6}$$

Bemerkung

1. Wählen „kleinstes“ zulässiges n

2. Kompromiss bei der Wahl von s, ε gewöhnlich erforderlich. Denn aus großem s und kleinem ε folgt ein großes n.

3. Bei Informationen über den Typ P_ξ kann man bessere Grundlagen für die Ermittlung des n als Satz 5.6.3 erhalten. (siehe Vorlesung Stochastik).

Beweis:
Es folgt ein Beweis dafür, dass aus Gleichung (Gleichung 5.6) die Gleichung (Gleichung 5.5) folgt. Aus dem Satz 5.6.3 folgt, $P(„|\eta_n - a| < \varepsilon“) = 1 - P(„|\underbrace{\eta_n}_{1/n\sum\xi_i} - a| \geq \varepsilon“)$. Nunmehr sich hier wieder Satz 5.6.3 anwenden und die obige Gleichung ist größergleich $1 - \frac{\sigma^2}{n\varepsilon^2} \geq s$. d. h. Gleichung (Gleichung 5.5) wird realisiert, wenn $s \leq 1 - \frac{\sigma^2}{n\varepsilon^2} \Leftrightarrow n \geq \frac{\sigma^2}{(1-s)\varepsilon^2}$. Letzteres entspricht gerade Gleichung (Gleichung 5.6). ■

5.7. Konvergenzarten und das Gesetz der großen Zahlen

Sei $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum und wir betrachten Zufallsgrößen oder -variablen über diesem Wahrscheinlichkeitsraum.

Definition
Seien ξ eine Zufallsgröße und $(\xi_n)_{n\geq 1}$ eine Folge von Zufallsgrößen. Die Folge (ξ_n) konvergiert mit Wahrscheinlichkeit gegen ξ, wenn für alle $\varepsilon > 0$ gilt:

$$P(„|\xi_n - \xi| \geq \varepsilon“) \xrightarrow{n\to\infty} 0 \Leftrightarrow P(„|\xi_n - \xi| < \varepsilon“) \xrightarrow{n\to\infty} 1$$

Satz 5.7.1 (Schwaches Gesetz der großen Zahlen nach Tschebyscheff)
Sei $(\xi_i)_{i\geq 1}$ eine Folge unabhängiger identisch verteilter Zufallsgrößen mit $E\xi_i = a, D\xi_i = \sigma^2$. Dann gilt:

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \xrightarrow{P} a$$

Beweis:
Aus Satz 5.6.3 folgt: help: Beweis unklar ■

Bezug zu relativen Häufigkeiten (Spezialfall) Wir betrachten die Folgen von Zufallsvariablen $(\eta_i)_{i\geq 1}$ mit Werten in $[X, \mathfrak{X}]$, wählen ein $A \in \mathfrak{X}$ und betrachten $P_{n_1}(A) = P_{n_i}(A)$. Dann wird gefordert, dass die $(\eta_i)_{i\geq 1}$ iid sind. Nun betrachten wir die $\xi_i := \xi_i^A := \chi_A(\eta_i) = \chi_A \circ \eta_i$ für $i = 1, 2, \ldots$. Diese sind messbar, da $A \in \mathfrak{X}$ und die Summe über alle ξ_i entspricht der Anzahl, wie oft A bei $1, \ldots, n$ realisiert wird, d. h. $1/n \sum_{i=1}^{n} \xi_i$ ist die relative Häufigkeit des Eintretens von A. Es ist klar, dass (ξ_i) iid sind und aus Satz 5.7.1 ergibt sich $1/n \sum \xi_i \xrightarrow{P} P_{\eta_1}(A)$. Dies ist eine rein qualitative Aussage und macht keine Angabe über die Konvergenzgeschwindigkeit. Durch die Anwendung von Satz 5.6.3 kann nun folgender Schluss gezogen werden: $P(„lvert 1/n \sum \xi_i - P_{\eta_1}(A)| \geq \varepsilon“) \leq \frac{P_{\eta_1}(A)(1-P_{\eta_1}(A))}{n\varepsilon^2} \leq \frac{1}{n\varepsilon^2} \Rightarrow \sup_A P(„|1/n \sum \xi_i - P_{\eta_1}(A)| \geq \varepsilon“) \xrightarrow{n\to\infty} 0$

andere Konvergenzarten Wir betrachten die schwache Konvergenz. Dazu sei $(Q_n)_{n=1}^{\infty}$ und Q ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}, \mathfrak{B}]$. Man sagt, (Q_n) konvergiert schwach gegen Q, wenn gilt:

$$\int f dQ_n \xrightarrow{n\to\infty} \int f dQ$$

Dabei ist f eine Abbildung von $\mathbb{R}$ nach $\mathbb{R}$, die stetig und beschränkt ist. Die Schreibweise dafür, dass (Q_n) schwach gegen Q konvergiert, ist $Q_n \Rightarrow Q$.

Satz 5.7.2

$$Q_n \Rightarrow Q \Leftrightarrow F_{Q_n}(x) \to F_Q(x)$$

für alle Stetigkeitspunkte von F_Q.

Satz 5.7.3

$$\eta_n \xrightarrow{P} \Rightarrow P_{\eta_n} \Rightarrow P_\eta$$

Bemerkung
Die Umkehrung ist im Allgemeinen falsch. Denn sei (η_n) iid, $P_\eta = P_{\eta_1}, \eta, \eta_n$ unabhängig. Dann folgt, dass $P_{\eta_n} = P_\eta \Rightarrow P_{\eta_1} \Rightarrow P_n$. Aber $P(„|\eta_n - \eta| \geq \varepsilon“)$ ist konstant und ungleich 0. Außer für den Fall, dass P_η, P_{η_n} Diracmaße sind.

Satz 5.7.4
Sei $P_\eta = \delta_a$. Dann folgt:

$$\eta_n \xrightarrow{P} \eta \Leftrightarrow P_{\eta_n} \Rightarrow P_\eta$$

Definition
Sei $(\eta)_{n=1}^{\infty}, \eta$ eine zufällige Größe. Dann konvergiert $(\eta_n)_{n=1}^{\infty}$ mit Wahrscheinlichkeit 1 (fast sicher) gegen η, wenn gilt**:

$$P(\text{„}\lim \eta_n = \eta\text{“}) = 1 = P(\{\omega \in \Omega\colon \eta(\omega) = \lim_{n\to\infty} \eta_n(\omega)\})$$

Satz 5.7.5

$$\eta_n \xrightarrow{P=1} \eta \Rightarrow \eta_n \xrightarrow{P} \eta$$

Die Umkehrung dieses Satzes ist im Allgemeinen falsch.

Bemerkung
Andere Formulierung:
Wenn zu jeder Folge eine Teilfolge existiert, so dass man immer Konvergenz in Wahrscheinlichkeit hat, hat man auch fast sichere Konvergenz.

Satz 5.7.6 (Starkes Gesetz der großen Zahlen nach Kolmogoroff)
Sei $(\xi_k)_{k=1}^{n}$ eine Folge unabhängiger identisch verteilter Zufallsgrößen mit endlichem Erwartungswert a. Dann gilt:

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k \xrightarrow{P=1} a$$

Bemerkung
Betrachten Zufallsgrößen mit $E|\xi|^p < +\infty$ und führen hier eine Konvergenz ein. Diese Räume werden als L_p-Räume bezeichnet. Es gilt dann $L_p(\Omega, \mathcal{F}, P)$ und man hat die Spezialfälle:

1. $p = 1 \to \rho(\eta, \xi)\colon = E|\xi - \eta|$
2. $p = 2 \to \langle f, g\rangle\colon = \int \overline{f} g dP$. Der Raum L_2 ist der **Hilbertraum** mit $\|f\|^2 = \langle f, f\rangle$.

Satz 5.7.7 (Spezialfall des Satz von Weierstraß)
Sei $f\colon [0,1] \to \mathbb{R}$ stetig. Weiter sei $b_n(x) := \sum_{k=0}^{n} f(\frac{k}{n})\binom{n}{k} x^k (1-x)^{n-k}$ ein Bernoullipolynom mit $x \in [0,1]$. Dann gilt:

$$\lim_{n\to\infty} b_n(x) = f(x)$$

**Schreibweise: $\eta_n \xrightarrow{P=1} \eta$

Beweis:
Der Beweis erfolgt durch die Anwendung des schwachen Gesetzes der großen Zahlen (Satz 5.7.1). Seien $\xi_1, \ldots, \xi_n$ unabhängige, identische verteilte Zufallsgrößen mit $P_{\xi_i} = B_{1,x} \Rightarrow E\xi_k = x, D\xi_k = x(1-x)$. Aus dem obigen Gesetz folgt ${}^1\!/\!{}_n \sum_{k=1}^n \xi_k \xrightarrow{P} x$. Dies ist gleichwertig zu $P_{1/n \sum_{k=1}^n \xi_k} \Rightarrow \delta_x \Leftrightarrow \int f(y) P_{1/n \sum_{k=1}^n \xi_k}(dy) \Rightarrow \int f(y)\delta_x(dy) = f(x)$. Nach dem Übertragungssatz ist $\int f(y) P_{1/n \sum_{k=1}^n \xi_k}(dy) = \int f(y/n) P_{\sum_{k=1}^n \xi_k}(dy) = \sum_{k=1}^n f(k/n) a_k = \sum_{k=1}^n f(k/n) \binom{n}{k} x^k (1-x)^{n-k}$. ■

A. Übungsaufgaben

1. Sei Ω eine nichtleere Menge und es gelte $A_i \subseteq \Omega, i = 1, 2, \ldots,$ Wir definieren:

$$\limsup_{n\to\infty} A_n \colon = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

$$\liminf_{n\to\infty} A_n \colon = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

Zeigen Sie:

a) Berechnen Sie $\limsup_{n\to\infty} A_n$ und $\liminf_{n\to\infty} A_n$ für die Mengen

$$A_n \colon = (1/n, 1 + 1/n]$$

b) Beweisen Sie:

$$\liminf_{n\to\infty} A_n = \{\omega \in \Omega \colon \omega \in A_n \forall n \in \mathbb{N} \text{ bis auf endlich viele}\}$$

c) Geben Sie eine zu (b) analoge Beschreibung von $\limsup_{n\to\infty} A_n$.

d) Zeigen Sie:

$$\liminf_{n\to\infty} A_n^c = \left(\limsup_{n\to\infty} A_n\right)^c$$

„$\Leftarrow$" $x \in (\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k)^c \Rightarrow x \notin \bigcup_{n=1}^{\infty}(\bigcup_{k=n}^{\infty} A_k) \Rightarrow \exists n_0 \colon x \notin \bigcap_{k=n_0}^{\infty} A_k \Rightarrow x \notin A_k$ für $k = 1, 2, \ldots \Rightarrow x \in A_k^c$ für $k = n_0, n_0 + 1, \ldots \Rightarrow x \in \bigcap_{k=n_0}^{\infty} A_k^c \Rightarrow x \in \bigcup_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k^c = \liminf_{n\to\infty} A_n^c$.

„$\Rightarrow$" selbst

2. Sei Ω eine nichtleere Menge und $\mathfrak{E} \in \mathfrak{P}(\Omega)$. Die von $\mathfrak{E}$ erzeugte σ-Algebra $\sigma(\mathfrak{E})$ ist definiert durch

$$\sigma(\mathfrak{E}) = \bigcap \{\mathcal{F} \colon \mathcal{F} \supseteq \mathfrak{E}, \mathcal{F} \text{ ist } \sigma\text{-Algebra über } \Omega\}$$

Es sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ und $\mathfrak{E} = \{\{1, 2, 3\}, \{1, 3\}\}$. Ermitteln Sie $\sigma(\mathfrak{E})$.

i	1	2	3	4	5	6	7	8
A_i	$\emptyset$	$\{2\}$	$\{1,3\}$	$\{1,2,3\}$	$\{4,5,6\}$	$\{2,4,5,6\}$	$\{1,3,4,5,6\}$	Ω

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8
A_1	1	2	3	4	5	6	7	8
A_2	x	2	4	4	6	6	8	8
A_3	x	x	3	4	7	8	7	8
A_4	x	x	x	4	8	8	8	8
A_5	x	x	x	x	5	6	7	8
A_6	x	x	x	x	x	6	8	8
A_7	x	x	x	x	x	x	7	8
A_8	x	x	x	x	x	x	x	8

3. Sei $[\Omega, \mathcal{F}, \mu]$ ein Maßraum mit $\mu(\Omega) < \infty$. Zeigen Sie für beliebige meßbare Mengen $A, B \in \mathcal{F}$ gilt:

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

Wir können $A \cup B$ auch als $A \cup (B \setminus A)$ aufschreiben und erhalten so zwei disjunkte Mengen. Also folgt $\mu(A \cup B) = \mu(A) + \mu(B \setminus A)$. Also müssen wir noch zeigen, dass $\mu(B \setminus A) = \mu(B) - \mu(A \cap B)$ gilt. Hierzu ist wichtig zu wissen, dass es sich um ein endliches Maß handelt. Wir können die Menge $B \setminus A$ zu $B \setminus (A \cap B)$ umformen und können einen Satz aus der Vorlesung anwenden: $\mu(A \cap B) = \mu(B \setminus (A \cap B)) = \mu(B) - \mu(B \cap A)$.

4. Wir betrachten die messbaren Räume $[\Omega_k, \mathcal{F}_k]$ mit $k = 1, 2, \ldots$. Dabei sei Ω_k endlich oder abzählbar unendlich und es gelte $\mathcal{F}_k = \mathfrak{P}(\Omega_k)$.

   a) Zeigen Sie

$$\otimes_{k=1}^{n} \mathcal{F}_k = \mathfrak{P}(\times_{k=1}^{n} \Omega_k)$$

   Hierfür müssen wir zeigen, dass $\otimes_{k=1}^{n} \mathcal{F}_k \subseteq \mathfrak{P}(\Omega_k)$ sowie $\mathfrak{P}(\Omega_k) \subseteq \otimes_{k=1}^{n} \mathcal{F}_k$ gilt.

   Es ist $\times_{k=1}^{n} \Omega_k = \{[\omega_1, \ldots, \omega_n] \colon \omega_k \in \Omega_k\}$. Weiterhin wissen wir, dass $\otimes_{k=1}^{n} \mathcal{F}_k = \sigma(\underbrace{\{\times_{k=1}^{n} A_k, A_k \in \mathcal{F}_k\}}_{=\mathfrak{E}})$ und es gilt, $\sigma(\mathfrak{E}) = \bigcap\{\mathcal{F} \text{ ist } \sigma\text{-Algebra}, \mathfrak{E} \subseteq \mathcal{F}\}$. Somit folgt $\{\times_{k=1}^{n} A_k, A_k \in \mathcal{F}_k\} \subseteq \mathfrak{P}(\times_{k=1}^{n} \Omega_k) \Rightarrow \sigma(\{\times_{k=1}^{n} A_k, A_k \in \mathcal{F}_k\}) \subseteq \mathfrak{P}(\times_{k=1}^{n} \Omega_k)$.

5. Zeigen Sie, dass durch nachfolgende Mengenfunktionen Wahrscheinlichkeitsmaße auf $[\mathbb{R}, \mathfrak{B}]$ gegeben sind und skizzieren Sie jeweils deren Verteilungsfunktion:

   a) Das Maß heißt **Binomialverteilung** mit den Parametern n und p:

$$B_{np} \colon = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k, \qquad p \in [0,1], n \in \mathbb{N}$$

   Wir betrachten zunächst die allgemeine Formulierung:
   Sei $(\mu_i)_{i \in I}$ eine abzählbare Familie von Maßen auf $[\mathbb{R}, \mathfrak{B}]$ und $(\alpha_i)_{i \in I}$ eine

Familie nichtnegativer reeller Zahlen. Falls gilt, $\sum_{i\in I}\alpha_i = 1$ und $\mu(\mathbb{R}) = 1$, dann ist $\mu = \sum_{i\in I}\alpha_i\mu_i(\mathbb{R})$ ein Wahrscheinlichkeitsmaß. In unserem Beispiel ist $I = \{0,1,2,\dots,n\} \subseteq \mathbb{N}_0$. Weiter ist $\alpha_i = \binom{n}{i}p^i(1-p)^{n-i}$. Für $0 < p < 1$ ist $\sum_{i\in I}\alpha_i = \sum_{i=0}^n \binom{n}{i}p^i(1-p)^{n-i} = (p+(1-p))^n = 1^n = 1$. Es ist klar, dass alle $\alpha_i > 0$.

Für die Verteilungsfunktion gilt: $F_{np}(x) = B_{np}((-\infty,x)) = \sum_{i=0}^n \binom{n}{i}p^i(1-p)^{n-i}\delta_i((-\infty,x))$. Dabei bildet F von $\mathbb{R}$ nach $[0,1]$ ab ($F\colon \mathbb{R} \to [0,1]$).

1. Fall Falls $-\infty < x \leq 0$, dann ist $F_{np}(x) = 0$.

2. Fall Falls $0 < x \leq 1$, dann ist $F_{np}(x) = \alpha_0$.

3. Fall Falls $1 < x \leq 2$, dann ist $F_{np}(x) = \alpha_0 + \alpha_1$.

4. Fall Falls $n < x < \infty$, dann ist $F_{np}(x) = \sum_{i=0}^n \alpha_i = 1$.

Allgemein gilt:

$$P((-\infty,x)) = F_{np}(x) = \sum_{i=0}^{\lceil x\rceil -1}\alpha_i$$

b) Diese Verteilung heißt **Poissonverteilung**:

$$\pi_\lambda\colon = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}\delta_k, \qquad \lambda > 0$$

c)

$$G_q\colon = \sum_{k=1}^{\infty}(1-q)q^{k-1}\delta_k, \qquad 0 \leq q \leq 1$$

6. Es sei $A \in \mathfrak{B}_n$ mit $0 < \ell^{(n)}(A) < +\infty$ gegeben. Zeigen Sie, dass durch:

$$\mathrm{Gl}_A(B)\colon = \frac{\ell^{(n)}(A\cap B)}{\ell^{(n)}(A)}, \qquad (B \in \mathfrak{B}_n)$$

ein Wahrscheinlichkeitsmaß auf $[\mathbb{R}^n, \mathfrak{B}_n]$ gegeben ist.

Es ist $\mathrm{Gl}_A(\emptyset) = \frac{\ell^{(n)}(A\cap\emptyset)}{\ell^{(n)}(A)} = \frac{\ell^{(n)}(\emptyset)}{\ell^{(n)}(A)} = 0$. Weiter gilt $\mathrm{Gl}_A(\mathbb{R}^n) = \frac{\ell^{(n)}(A\cap\mathbb{R}^n)}{\ell^{(n)}(A)} = \frac{\ell^{(n)}(A)}{\ell^{(n)}(A)} = 1$.

7. $[\Omega, \mathcal{F}, P] \xrightarrow{\xi_i} [\mathbb{R}, \mathfrak{B}, P_{\xi_i}], P_{\xi_i} = Ex_{\lambda_i}, \lambda_i > 0$. Weiter ist $P_{\xi_i}\colon F_i(x) = (1-e^{-\lambda_i x})\chi_{(0,\infty)}(x) = \begin{cases} 0 & x \leq 0 \\ 1-e^{-\lambda_i x} & x > 0 \end{cases}$. Gesucht ist die Verteilungsfunktion der Zufallsgröße $\xi = \min\{\xi_1, \ldots, \xi_n\}$.
Wie ist $F(x)$ definiert? Es gilt, $F(x) = P_\xi((-\infty, x)) = P(\min\{\xi_1, \ldots, \xi_n\} < x) = 1 - P(\min\{\xi_1, \ldots, \xi_n\} \geq x) = 1 - P(\xi_1 \geq x, \ldots, \xi_n \geq x) = 1 - P(\{\omega \in \Omega\colon \xi_1(\omega) \geq x, \ldots, \xi_n(\omega) \geq x\} = 1 - P(\bigcap_{i=1}^n \{\omega \in \Omega\colon \xi_i \geq x\})$ mit $x \in \mathbb{R}$. Dabei sind die Ereignisse unabhängig und man kann das schreiben als $1 - \prod_{i=1}^n P(\{\omega \in \Omega\colon \xi_i \geq x\}) = 1 - \prod_{i=1}^n P(\xi_i \geq x) = 1 - \prod_{i=1}^n (1 - F_i(x))$. Die weitere Betrachtungen erfordern eine Fallunterscheidung. Im Fall $x \leq 0$ ist $F(x) = 0$. Für $x > 0$ gilt $F(x) = 1 - \prod_{i=1}^n (1 - (1 - e^{-\lambda_i x})) = 1 - \prod_{i=1}^n e^{-\lambda_i x} = 1 - e^{-\lambda x}$ für $\lambda = \sum_{i=1}^n \lambda_i$. Also folgt, $F(x) = (1 - e^{-\lambda x})\chi_{(0,\infty)}(x)$. Somit ist $P_\xi = \mathrm{Ex}_\lambda$.)

8. Es sind $n \geq 1$ unabhängige Zufallsgrößen $\xi_1, \ldots, \xi_n$ über einem Wahrscheinlichkeitsraum $[\Omega, \mathcal{F}, P]$ mit $P_{\xi_i} = \mathrm{Ex}_\lambda, \lambda > 0$ gegeben. Weiterhin ist $\eta = \xi_1 + \cdots + \xi_n$. Zeigen Sie: P_η ist absolut stetig bezüglich des LEBESGUE-Maßes ℓ auf $[\mathbb{R}, \mathfrak{B}]$ mit der Dichte

$$f_n(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{(n-1)!}\lambda^n e^{-\lambda x} x^{n-1} & x \geq 0 \end{cases}$$

Das Wahrscheinlichkeitsmaß P_η heißt ERLANG-Verteilung mit n Freiheitsgraden und dem Parameter $\lambda > 0$. Speziell ist f_1 die Dichte der Exponentialverteilung.
Wir wissen: $\eta = \xi_1 + \cdots + \xi_n, P_\eta = \ell f_n$ und müssen eine Induktion über n durchführen: Sei $n = 2, \eta = \xi_1 + \xi_2$. Dabei sind die ξ_1, ξ_2 unabhägig und gleich verteilt. Nach der Vorlesung ist $P_\eta = P_{\xi_1} * P_{\xi_2}$. Wir wissen weiter, $f_{\xi_1} = f_{\xi_2} = f_1$. Die Dichte ist $(f_{\xi_1} * f_{\xi_2})(x) = \int_0^y \underbrace{f_{\xi_1}(y-x)}_{>0 \text{ für } y-x>0} \underbrace{f_{\xi_2}(x)}_{>0} \ell(dx) = \lambda^2 \int_0^y e^{-\lambda(y-x)} e^{-\lambda x}\, dx = \lambda^2 e^{-\lambda y} \int_0^y dx = \lambda^2 e^{-\lambda y} y$. Denn $e^{-\lambda x} e^{-\lambda(-x)} = 1$. Also folgt, dass $f_{\xi_1} * f_{\xi_2} = f_2$. Es gelte nun die Behauptung für $n-1$. Nun ist $P_{\eta_n} = *_{i=1}^n P_{\xi_i} = P_{\eta_{n-1}} * P_{\xi_n}$. Es gilt $(f_{n-1} * f_n)(y) = \int_0^y f_{n-1}(y-x) f_n(x) \ell(dx) = \frac{1}{(n-2)!}\lambda^{n-1+1} \int_{(0,y)} e^{-\lambda(y-x)}(y-x)^{n-2} e^{-\lambda x}\, \ell(dx) = \frac{1}{(n-2)!}\lambda^n e^{-\lambda y} \int_0^y (y-x)^{n-2}\, dx$ Nun substituieren wir $t = y - x, dt = -dx$ und schreiben das Integral $\frac{1}{(n-2)!}\lambda^n e^{-\lambda y}(-\int_y^0 t^{n-2}\, dt) = \frac{1}{(n-2)!}\lambda^n e^{-\lambda y} \int_0^y t^{n-2}\, dt = \frac{1}{(n-2)!(n-1)}\lambda^n e^{-\lambda y} t^{n-1}|_0^y = \frac{1}{(n-1)!}\lambda^n e^{-\lambda y} y^{n-1}$. Somit folgt die Behauptung.

9. Sei $[\Omega, \mathcal{F}, P]$ ein Wahrscheinlichkeitsraum. Berechnen Sie das erste Moment, das zweite Moment und die Varianz folgender Zufallsgrößen: $\xi_2\colon \Omega \to \mathbb{N}_0$ mit $P_{\xi_2} = G_q, 0 < q < 1$.
$m_2 = (1-q)\sum_{k=1}^\infty k^2 q^{k-1} = 2m_1 \frac{1}{1-q}$

Index