# 1 Ein wenig Maßtheorie

**Wiederholung und Borel-/Lebesgue-messbare Mengen:**

$\Omega$ ist ein Wahrscheinlichkeitsraum mit Wahrscheinichkeitsmaß $\mathbb{P}$ und $\sigma$-Algebra $\mathcal{F}$:

- $\emptyset, \Omega \in \mathcal{F}$;
- $A_i \in \mathcal{F} \Rightarrow A_i^c \in \mathcal{F}, \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}, \bigcap_{i\in\mathbb{N}} A_i \in \mathcal{F}$.

Beispiele von $\sigma$-Algebren:

- $\mathcal{F} = P(\Omega)$, die Potenzmenge von $\Omega$, d.h. alle Teilmengen von $\Omega$. (Im Allgemeinen zu groß!)
- $\mathcal{F}$ sind alle abzählbaren und ko-abzählbaren Mengen.
- $\mathcal{F} = \mathcal{B}$, die Borel'schen Mengen (falls $\Omega$ eine Topologie hat).
- $\mathcal{F}$ sind die Lebesgue-messbaren Mengen (falls $\Omega \subset \mathbb{R}^n$).

**Definition 1** *In einem topologischen Raum ist die Borel-Algebra $\mathcal{B}$ die kleinste $\sigma$-Algebra, die alle offenen Mengen enthält. D.h. $\Omega$ hat eine Topologie und $\mathcal{B}$ wird von den offenen Mengen mittels abzählbarer Vereinigungen, abzählbarer Durchschnitte und Komplementen erzeugt.*

**Beispiel 2** $\Omega = [0,1]$ *mit Euklidischer Topologie.*
$\{a\}$ *ist eine Borel'sche Menge, weil* $\{a\} = \bigcap_{1\le n\in\mathbb{N}}(a-\frac{1}{n}, a+\frac{1}{n})$.
$[\frac{1}{3}, \frac{2}{3}]$ *ist eine Borel'sche Menge, weil* $[\frac{1}{3}, \frac{2}{3}] = [0,1] \setminus ([0,\frac{1}{3}) \cup (\frac{2}{3}, 1])$.
$\mathbb{Q} \cap \Omega$ *ist eine Borel'sche Menge, weil* $\mathbb{Q} \cap [0,1] = \bigcup_{1\le q\in\mathbb{N}} \bigcup_{0\le p\le q} \bigcap_{1\le n\in\mathbb{N}} (\frac{p}{q} - \frac{1}{n}, \frac{p}{q} + \frac{1}{n}) \cap [0,1]$.

Ein Maß $\mu$ (zB. das Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\Omega$) heißt:

- $\sigma$-additiv, falls $A_i \in \mathcal{F}$ paarweise disjunkt, dann $\mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i)$;
- translationsinvariant (auf $\Omega = \mathbb{R}^n$), falls für alle $v \in \mathbb{R}^n$ und $A \in \mathcal{F}$, $\mu(A+v) = \mu(A)$, wobei $A + v = \{a + v : a \in A\}$.

**Definition 3** *Eine Menge $N \subset \mathbb{R}^n$ heißt (Lebesgue-) Nullmenge falls*

$$\forall \varepsilon > 0 \ \exists \text{ eine Folge von Blöcken } (B_i)_{i\in\mathbb{N}} \text{ so dass } A \subset \bigcup_{i\in\mathbb{N}} B_i \text{ und } \sum_{i\in\mathbb{N}} \mathrm{Vol}(B_i) < \varepsilon.$$

*Ein Block is eine Menge der Form $B = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$ und dessen Volumen ist* $\mathrm{Vol}(B) = \prod_{i=1}^n (b_i - a_i)$.

Es ist klar, dass Teilmengen von Nullmengen wieder Nullmengen sind.

**Definition 4** *In $\mathbb{R}^n$ ist die $\sigma$-Algebra $\mathcal{F}_{Leb}$ von Lebesgue-messbaren Mengen die kleinste $\sigma$-Algebra, die alle offenen Mengen und Nullmengen enthält. D.h. $\mathcal{F}_{Leb}$ ist erzeugt von den offenen Mengen und Nullmengen mittels abzählbarer Vereinigungen, abzählbarer Durchschnitte und Komplementen.*

Charakterisierung: Das Lebesgue'sche Maß $\lambda$ auf $\mathbb{R}^n$ ist das eindeutige Maß mit den Eigenschaften:

- $\lambda$ ist defininiert auf $\mathcal{F}_{Leb}$ und $\sigma$-additiv;
- $\lambda$ ist translationsinvariant;
- $\lambda$ is normiert so dass $\lambda([0,1]^n) = 1$.

Hieraus folgt auch:

- $\lambda(N) = 0$ für jede Nullmenge $N$.

**Beispiel 5** *(Die Mittel-$\varepsilon$ Cantor-Menge):*
*- Fang mit dem Einheitsintervall $C_0 = [0,1]$ und $\varepsilon \in (0,1)$ an.*
*- $C_n$ wird aus $C_{n-1}$ konstruiert indem aus jeder Intervallkomponente $J$ von $C_{n-1}$ das offene mittlere Intervall der Länge $\varepsilon|J|$ entfernen wird.*
*- Die gefragte Cantor-Menge ist $C = \bigcap_{n\in\mathbb{N}} C_n$.*
*Dann hat $C_n$ genau $2^n$ Intervallkomponenten von Länge $\frac{(1-\varepsilon)^n}{2^n}$. Damit ist*

$$\lambda(C) = \lambda(\bigcap_{n\in\mathbb{N}} C_n) = \lim_{n\to\infty} 2^n \frac{(1-\varepsilon)^n}{2^n} = \lim_{n\to\infty} (1-\varepsilon)^n = 0.$$

*Also die Mittel-$\varepsilon$ Cantor-Menge ist eine überabzählbare Nullmenge.*

**Definition 6** *Ein Ereignis $A$ gilt als* **fast sicher** *(Abkürzung f.s. oder a.s für almost sure auf Englisch), falls $A^c$ eine Nullmenge ist. Wir können deswegen von fast sicherer Konvergenz sprechen:*

$$X_n \to X \quad f.s.$$

*falls $\lim_n X_n(\omega) = X(\omega)$ für alle $\omega \in \Omega$ bis auf eine Nullmenge.*

**Nicht alle Mengen sind Lebesgue-messbar.**

Zu solchen Mengen (Ereignissen) kann man also keine Wahrscheinlichkeit zuordnen:

**Beispiel 7** *(Vitali-Mengen auf $\Omega = [0,1]$)*

- *Äquivalenzrelation $x \sim y$ falls $x - y \in \mathbb{Q}$.*
- *Äquivalenzklassen $[x]$ sind also abzählbar und es gibt überabzählbar viele Äquivalenzklassen.*
- *Wähle aus jeder Äquivalenzklasse ein Element $v$. Diese formen die Menge $V$. Hier verwenden wir das* **Auswahlaxiom***!*
- *NB: $\#(V \cap [x]) = 1$ für jede Äquivalenzklasse $[x]$, und*

$$\bigcup_{q\in\mathbb{Q}\cap[0,1]} (V+q) \bmod 1 = [0,1].$$

  *aber die Mengen $(V+q) \bmod 1$ für $q \in \mathbb{Q}\cap[0,1]$ sind paarweise disjunkt.*

*Die Menge $V$ ist nicht messbar, d.h. $V \notin \mathcal{F}_{Leb}$. Nämlich gilt entweder $\lambda(V) = 0$, aber dann $\lambda(V + q \bmod 1) = 0$ wegen Translationsinvarianz und dann*

$$\lambda([0,1]) = \lambda(\bigcup_{q \in \mathbb{Q} \cap [0,1]} (V + q) \bmod 1) = \sum_{q \in \mathbb{Q} \cap [0,1]} \lambda(V + q \bmod 1) = 0,$$

*oder $\lambda(V) > 0$ und dann, wieder wegen Translationsinvarianz,*

$$\lambda([0,1]) = \lambda(\bigcup_{q \in \mathbb{Q} \cap [0,1]} (V + q) \bmod 1) = \sum_{q \in \mathbb{Q} \cap [0,1]} \lambda(V + q \bmod 1) = \infty.$$

*Beide sind natürlich nicht wahr. Die einzige Möglichkeit ist also dass $V$ nicht messbar ist.*

**Bemerkung 8** *Da alle Borel'schen Mengen Lebesgue-messbar sind, ist $Z$ auch keine Borel'sche Menge.*

**Produkträume und Produktmaße.**

Produkträume $\Omega = \prod_{i \in \mathbb{N}} \Omega_i$ werden häufig verwendet für Folgen von unendlich vielen Bernoulli-Experimenten. Zum Beispiel, $\Omega_i = \{K, Z\}$ entspricht den Münzwürfen und $\Omega_i = \{1,2,3,4,5,6\}$ den Würfelwürfen. Die $\sigma$-Algebra $\mathcal{F}$ auf $\Omega$ ist das unendliche Produkt $\prod_{i \in \mathbb{N}} \mathcal{F}_i$ wobei die $\mathcal{F}_i$ $\sigma$-Algebren aus $\Omega_i$ sind. Bei dem Münz- oder Würfelwurf mit endlichem $\Omega_i$ kann man für $\mathcal{F}_i$ einfach die Potenzmengen nehmen $\mathcal{F}_i = P(\Omega_i)$. Aber die Räume können zum Beispiel auch (Teilmengen von) $\mathbb{R}^n$ sein, mit $\mathbb{P} =$ Lebesgue-Maß.

**Definition 9** *Sei $\mathbb{P}_i$ ein Wahrscheinlichkeitsmaß auf $(\Omega_i, \mathcal{F}_i)$ für jedes $i \in \mathbb{N}$. Das Produktmaß $\mathbb{P}$ auf $(\Omega, \mathcal{F})$ ist gegeben durch*

$$\mathbb{P}(A_0 \times A_1 \times \cdots) = \prod_{i \in \mathbb{N}} \mathbb{P}(A_i)$$

*wobei $A_i \in \mathcal{F}_i$. Also die Bernoulli-Experimenten sind hier unabhängig genommen.*

*Falls $\Omega_i$ immer die gleiche endliche (oder sogar abzählbare) Menge ist, können wir einen Wahrscheinlichkeitsvektor $p = (p_k)_{k \in \Omega_i}$ nehmen (d.h. $p_k \geq 0$ und $\sum_{k \in \Omega_i} p_k = 1$) und $\mathbb{P}_i$ definieren durch $\mathbb{P}(k) = p_k$ für $k \in \Omega_i$. In diesem Fall heißt das Produktmaß $\mathbb{P}$, das $(p_k)$-Bernoulli Maß.*

Wir schreiben kurz $\mathbb{P}(A_1 \times \cdots \times A_n)$ für $\mathbb{P}(A_0 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots)$. Nullmengen $N = N_0 \times N_1 \times \cdots$ sind also Mengen mit $\prod_{i \in \mathbb{N}} \mathbb{P}(N_i) = 0$, und das wird schon erreicht wenn eine der Mengen $N_i$ Maß null hat (also $\mathbb{P}_i(N_i) = 0$).

Das $(\frac{1}{2}, \frac{1}{2})$-Bernoulli Maß beschreibt unendlich viele Münzwürfe mit einer fairen Münze. In diesem Fall können wir natürlich auch $\Omega_i = \{0, 1\}$ statt $\{K, Z\}$ schreiben.

**Das starke Gesetz der großen Zahlen.**

**Frage:** Gehört die Menge

$$H := \left\{ (X_i)_{i\in\mathbb{N}} \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{2} \right\}$$

zu $\mathcal{F}$? In dem Fall, was ist $\mathbb{P}(H)$?

Antwort:

$$\begin{aligned} H &= \left\{ (X_i)_{i\in\mathbb{N}} \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{2} \right\} \\ &= \left\{ (X_i)_{i\in\mathbb{N}} \in \Omega : \forall\, \varepsilon > 0\ \exists\, N \in \mathbb{N}\ \forall n \geq N\ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{2} \right| < \varepsilon \right\} \\ &= \left\{ (X_i)_{i\in\mathbb{N}} \in \Omega : \forall\, 1 \leq k \in \mathbb{N}\ \exists\, N \in \mathbb{N}\ \forall n \geq N\ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{2} \right| < \frac{1}{k} \right\} \\ &= \bigcap_{1\leq k\in\mathbb{N}} \bigcup_{N\in\mathbb{N}} \bigcap_{N\leq n\in\mathbb{N}} H_{n,k}, \end{aligned}$$

wobei $H_{n,k} := \left\{ (X_i) \in \Omega : \left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{2}\right| < \frac{1}{k} \right\}$ endliche Vereinigungen von $n$-Zylindern[1], und deswegen offen, sind. Damit ist $H$ eine Borel'sche Menge (erzeugt von abzählbaren Anwendungen von $\cup$ und $\cap$), also $H \in \mathcal{F}$.

Jetzt berechnen wir $\mathbb{P}(H)$ mit Hilfe des Komplements

$$H^c = \bigcup_{1\leq k\in\mathbb{N}} \bigcap_{N\in\mathbb{N}} \bigcup_{N\leq n\in\mathbb{N}} H^c_{n,k}.$$

Wir schreiben $S_n = \sum_{i=1}^{n} X_i$. Die Münzwürfe $X_i$ haben $\mathbb{E}(X_i) = \frac{1}{2}$ und $\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$. Also $\mathbb{E}(\frac{S_n}{n}) = \frac{1}{2}$ und $\mathrm{Var}(\frac{S_n}{n}) = \frac{1}{4n}$. Mit Hilfe der Chebyshev'sche Ungleichung finden wir

$$\mathbb{P}(H^c_{n,k}) = \mathbb{P}\left( \left| \frac{S_n}{n} - \frac{1}{2} \right| \geq \frac{1}{k} \right) \leq \frac{k^2}{4n}.$$

Leider ergibt die Abschätzung $\mathbb{P}(\bigcup_{n\geq N} H^c_{n,k}) \leq \sum_{n\geq N} \frac{k}{4n} = \infty$ nichts nützliches. Deswegen verwenden wir den folgenden Trick: wir schauen es uns erst nur für $n$ in der Teilfolge der Quadrate an.

$$\mathbb{P}\left( \underbrace{\bigcup_{n\geq N} H^c_{n^2,k}}_{\tilde{H}^c_{N,k}} \right) \leq \sum_{n\geq N} \frac{k^2}{4n^2} \leq \frac{k^2}{4(N-1)}.$$

[1]Die Menge $\{\omega_1\} \times \cdots \times \{\omega_n\} \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots$ heißt eine Zylindermenge der Länge $n$, also $n$-Zylinder. Sie sind offen in der Produkttopologie auf $\Omega$, und erzeugen die ganze Produkttopologie

Die Mengen $\bigcup_{n\geq N} H^c_{n^2,k}$ sind geschachtelt: $\bigcup_{n\geq N} H^c_{n^2,k} \supset \bigcup_{n\geq N+1} H^c_{n^2,k}$. Also

$$\mathbb{P}\left(\bigcap_{1\leq N\in\mathbb{N}} \bigcup_{n\geq N} H^c_{n^2,k}\right) = \lim_{N\to\infty} \mathbb{P}(\bigcup_{n\geq N} H^c_{n^2,k}) \leq \lim_{N\to\infty} \frac{k^2}{4N} = 0.$$

Damit ist auch das Maß

$$\mathbb{P}(\bigcup_{1\leq k\in\mathbb{N}} \bigcap_{1\leq N\in\mathbb{N}} \bigcup_{n\geq N} H^c_{n^2,k}) \leq \sum_{k=1}^{\infty} 0 = 0.$$

Also, entlang der Teilfolge der Quadrate geht $\frac{S_n}{n}$ fast sicher gegen $\frac{1}{2}$. Es bleiben die Zahlen zwischen Quadraten übrich. Sei $m \in \mathbb{N}$ beliebig. Dann gibt es ein $n \in \mathbb{N}$ und $0 \leq d \leq 2n$, so dass $m = n^2 + d$.

$$\begin{aligned}\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \frac{1}{2}\right| &= \left|\frac{n^2}{n^2+d}\left(\frac{1}{n^2}\sum_{i=1}^{n^2} X_i - \frac{1}{2}\right) - \frac{d}{2(n^2+d)} + \frac{1}{n^2+d}\sum_{i=n^2+1}^{n^2+d} X_i\right| \\ &\leq \left|\frac{n^2}{n^2+d}\left(\frac{1}{n^2}\sum_{i=1}^{n^2} X_i - \frac{1}{2}\right)\right| + \frac{1}{n} + \frac{2}{n}. \qquad (1)\end{aligned}$$

Im Grenzwert $m = n^2 + d \to \infty$ geht $\frac{n^2}{n^2+d}$ gegen 1 und $\frac{3}{n} \to 0$. Also, falls $\frac{1}{n^2}\sum_{i=1}^{n^2} X_i - \frac{1}{2}$ gegen 0 konvergiert als $n \to \infty$, dann konvergiert auch $\frac{1}{m}\sum_{i=1}^{m} -\frac{1}{2}$ gegen 0 als $m \to \infty$. Aber wir wissen schon, dass $\frac{1}{n^2}\sum_{i=1}^{n^2} X_i \to \frac{1}{2}$ fast sicher. Damit gilt $\frac{1}{m}\sum_{i=1}^{m} X_i \to \frac{1}{2}$ fast sicher und $\mathbb{P}(H) = 1$.

Dieses f.s. Ergebnis heißt das **starke Gesetz der großen Zahlen** und gilt allgemein für den Mittelwert von unabhängigen identisch verteilten Zufallsvariablen mit endlicher Varianz:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}(X_i) \quad \text{f.s.}$$

**Die Borel-Cantelli Lemmas.**

**Lemma 10** *(Borel-Cantelli) Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(A_n)_{n\in\mathbb{N}}$, $A_n \in \mathcal{F}$, eine Folge Ereignisse.*

*(a) Falls $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$, dann passiert f.s. $A_n$ nur für höchstens endlich viele $n$, also*

$$\mathbb{P}(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} A_n) = 0.$$

*(b) Falls $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ und die $A_n$ sind paarweise unabhängig, dann passiert f.s. $A_n$ für unendlich viele $n$, also*

$$\mathbb{P}(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} A_n) = 1.$$

**Beweis.** (a) Weil $\mathbb{P}$ $\sigma$-additiv ist, gilt $\mathbb{P}(\bigcup_{n\geq N} A_n) \leq \sum_{n\geq N} \mathbb{P}(A_n) < \infty$. Auch gilt $\bigcup_{n\geq N} A_n \supset \bigcup_{n\geq N+1} A_n$ für jedes $N$. Also

$$\mathbb{P}(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} A_n) = \lim_{N\to\infty} \mathbb{P}(\bigcup_{n\geq N} A_n) \leq \lim_{N\to\infty} \sum_{n\geq N} \mathbb{P}(A_n) = 0,$$

weil der "Schwanz" einer konvergenten Summe gegen Null konvergiert.

(b) Die Wahrscheinlichkeit, dass kein Ereignis $A_n$ passiert für $n \geq N$ ist, wegen Unabhängigkeit,

$$\mathbb{P}(\bigcap_{n\geq N} A_n^c) = \prod_{n\geq N} \mathbb{P}(A_n^c) = \prod_{n\geq N} (\, 1 - \mathbb{P}(A_n)\,).$$

Die Ungleichung $1 - x \leq e^{-x}$ ergibt

$$\mathbb{P}(\bigcap_{n\geq N} A_n^c) = \prod_{n\geq N} (\, 1 - \mathbb{P}(A_n)\,) \leq \prod_{n\geq N} e^{-\mathbb{P}(A_n)} = \exp(-\sum_{n\geq N} \mathbb{P}(A_n)) = 0,$$

weil $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n)$ gegen $\infty$ divergiert. Die abzählbare Vereinigung von Nullmengen ist wieder eine Nullmenge, also $\mathbb{P}(\bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} A_n^c) = 0$. Die Gegenwahrscheinlichkeit ist also

$$\mathbb{P}\left(\bigcap_{N\in\mathbb{N}} \bigcup_{n\geq N} A_n\right) = 1 - \mathbb{P}\left(\bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} A_n^c\right) = 1.$$

□

Das erste Borel-Cantelli Lemma ist nötig um die allgemeine Version des starken Gesetzes der großen Zahlen zu beweisen.

**Theorem 11** *(Das starke Gesetz der großen Zahlen). Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen, so dass $\mathbb{E}(X_i) = \mu$ und $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \quad f.s.$$

**Beweis.** Der Beweis ist genau so wie bei dem Münzwurf, bis auf eine Sache. In dem Schritt (1) hatten wir verwendet, dass die $X_i$ beschränkt sind, was wir hier nicht voraussetzen. Also, statt der Abschätzung $\sum_{i=n^2+1}^{n^2+d} X_i \leq d$, werden wir zeigen, dass f.s.

$$\left|\sum_{i=n^2+1}^{n^2+d} X_i\right| \leq n^{3/2} \qquad \text{für nur endlich viele } n.$$

Die Chebyscev'sche Ungleichung ergibt

$$\mathbb{P}\left(\left|\sum_{i=n^2+1}^{n^2+d} X_i - \mu d\right| \geq n^{3/2}\right) \leq \frac{d}{n^3}\,\mathrm{Var}(X_i) \leq \frac{2n}{n^3}\,\mathrm{Var}(X_i) = \frac{1}{2n^2},$$

und das ist summierbar über $n$. Das Borel-Cantelli Lemma (a) sagt, dass f.s. das Ereignis

$$\left|\sum_{i=n^2+1}^{n^2+d} X_i - \mu d\right| \geq n^{3/2}$$

nur endlich oft auftreten kann. Deswegen gibt es f.s. ein $n_0$, so dass für all $n \geq n_0$:

$$\left|\sum_{i=n^2+1}^{n^2+d} X_i\right| = \left|\left(\sum_{i=n^2+1}^{n^2+d} X_i - \mu d\right) + \frac{d}{2}\right| \leq n^{3/2} + \mu d.$$

Damit können wir in (1) die Abschätzung verbessern zu

$$\left|\frac{1}{n^2+d}\sum_{i=n^2+1}^{n^2+d} X_i\right| \leq \frac{n^{3/2}+\mu d}{n^2+d} \to 0$$

als $n \to \infty$ (weil $0 \leq d \leq 2n$), und der obere Beweis kann vollendet werden. $\square$

**Mehr $\sigma$-Algebren und Kolmogorovs Null-Eins-Gesetz.**

Seien $\mathcal{F}_i$ die Borel'schen $\sigma$-Algebren der Räume $X_i$ in einem Produktraum $\Omega = \prod_{i\in\mathbb{N}} \Omega_i$, wobei jedes $\Omega_i$ von einer Wahrscheinlichkeitsfunktion $\mathbb{P}_i$ und Borel $\sigma$-Algebra $\mathcal{F}_i$ versehen ist. Obwohl $\mathcal{F}_i$ eigentlich Teilmengen von $\Omega_i$ enthält fassen wir das doch allgemeiner auf, als eingebettet in $\Omega$. Das heißt: $A \in \Omega_i$ wird $\Omega_1 \times \Omega_{i-1} \times A \times \Omega_{i+1} \times \Omega_{i+2} \times \cdots$ gleichgestellt.

- $\mathcal{B} = \mathcal{F}_1^\infty = \sigma(\bigcup_{i\in\mathbb{N}} \mathcal{F}_i)$ ist die Borel'sche $\sigma$-Algebra.
- $\mathcal{F}_1^n = \sigma(\bigcup_{i=1}^n \mathcal{F}_i)$ ist die Algebra von allen $n$-Zylindern (eine endliche Algebra).
- $\mathcal{F}_{n+1}^\infty = \sigma(\bigcup_{i=n+1}^\infty \mathcal{F}_i)$.
- $\mathcal{T} = \sigma(\bigcap_{n\in\mathbb{N}} \mathcal{F}_{n+1}^\infty)$ ist die terminale $\sigma$-Algebra (oder tail algebra auf Englisch). Es ist die $\sigma$-Algebra von Borel'sche Mengen, die nicht von den ersten endlich vielen Koordinaten abhängen.

Weil die Koordinaten $X_i$ unabhänging sind, sind die $\sigma$-Algebren $\mathcal{F}_1^n$ und $\mathcal{F}_{n+1}^\infty$ es auch. Das heißt, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ für alle $A \in \mathcal{F}_1^n$ und $B \in \mathcal{F}_{n+1}^\infty$.

**Theorem 12** *(Kolmogorovs Null-Eins-Gesetz) Für jede Menge $A \in \mathcal{T}$ gilt $\mathbb{P}(A) = 0$ oder 1.*

Das heißt, dass zum Beispiel die folgenden "Grenzwert"-Mengen nur Maß Null oder Eins haben können.
$\{(X_i)_{i\in\mathbb{N}} : X_i = 1 \text{ unendlich oft}\}$;
$\{(X_i)_{i\in\mathbb{N}} : \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = p\}$;
$\{(X_i)_{i\in\mathbb{N}} : \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i > q\}$.

**Beweis.** (Skizze) Für jedes $n \in \mathbb{N}$ gilt, dass $\mathcal{F}_1^n$ und $\mathcal{F}_{n+1}^\infty$ unabhängig sind. Deswegen sind auch $\mathcal{F}_1^n$ und $\mathcal{T} = \bigcap_n \mathcal{F}_{n+1}^\infty$ unabhängig. Also ist die Borel-$\sigma$-Algebra $\mathcal{B} = \sigma(\bigcup_n \mathcal{F}_1^n)$ unabhängig von $\mathcal{T}$. Aber $\mathcal{T} \subset \mathcal{B}$, also $\mathcal{T}$ ist unabhängig von sich selbst! Daher gilt $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ für alle $A, B \in \mathcal{T}$. Aber mit $A = B$ ergibt dies $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, und die einzigen Zahlen $x$ die $x = x^2$ erfühlen, sind $x = 0$ und $x = 1$. Also $\mathbb{P}(A) = 0$ oder 1. $\square$

## 2 Moment-erzeugende und charakteristische Funktionen

Das Thema dieses Abschnitts ist ein gewisses Hilfsmittel aus der Analysis. Gegeben ein Zufallsvariable $X$ (mit Dichte $f_X(x)$ falls kontinuierlich) definieren wir:

| **Moment-erzeugende Funktion** | **Charakteristiche Funktion** |
| --- | --- |
| $m_X(t) := \mathbb{E}(e^{tX})$ | $\varphi_X(t) := \mathbb{E}(e^{itX}) \qquad (i = \sqrt{-1})$ |
| $= \begin{cases} \sum_{k \in \mathbb{Z}} e^{tk}\mathbb{P}(X = k) & \text{diskret} \\ \int_{-\infty}^\infty e^{tx} f_X(x)\, dx & \text{kontinuierlich} \end{cases}$ | $= \begin{cases} \sum_{k \in \mathbb{Z}} e^{itk}\mathbb{P}(X = k) & \text{diskret} \\ \int_{-\infty}^\infty e^{itx} f_X(x)\, dx & \text{kontinuierlich} \end{cases}$ |
| ist die Laplace-transformierte von $X$ | ist die Fourier-transformierte von $X$ |
| konvergiert für alle $t < \log R$ | konvergiert für alle $t \in \mathbb{R}$ |
| $\frac{d^n}{dx^n}\ m_X(t)\Big\vert_{t=0} = \mathbb{E}(X^n)$ | $\frac{d^n}{dx^n}\ \varphi_X(t)\Big\vert_{t=0} = i^n\, \mathbb{E}(X^n)$ |

Hier bedeutet $R$ der Konvergenzradius von der Potenzreihe $\sum_k \mathbb{P}(Y = k)y^k$, und ähnlich für das Integral im kontinuierlichen Fall.

**Bemerkung 13** *Die Eigenschaft* $\frac{d^n}{dx^n} m_X(t)\Big\vert_{t=0} = \mathbb{E}(X^n)$ *erklärt den Namen "moment-erzuegend", aber* 100% *richtig ist er nicht, denn das n-te Moment* $M_n(X) = \mathbb{E}(|X^n|)$ *hat Betragsstriche.*

**Bemerkung 14** *In der Analysis gibt es* Umkehrfunktionen *für die Laplacetranformation und für die Foureier-Tranformation. Mit deren Hilfe kann man zum Beispiel aus* $\varphi_X(t)$ *die Dichte* $f_X(x)$ *zurückfinden. Auf dieser Tatsache stützt der Stetigkeitssatz von Lévy, den man für den Beweis des zentralen Grenzwertsatzes verwenden kann. Ebenso, wenn alle Erwartungen* $\mathbb{E}(X^n)$ *endlich sind, kann man die Dichte* $f_X(x)$ *zurückfinden, aber in vielen Fällen sind nur endlich viele Momente endlich.*

**Beispiel 15** *Sei* $X \simeq Geo(q)$*, also* $\mathbb{P}(X = k) = q^{k-1}(1 - q)$ *für* $k = \{1, 2, 3, \dots\}$ *und* $q \in (0, 1)$*. Dann ist die moment-erzeugende Funktion*

$$m_X(t) \;=\; \sum_{k=1}^\infty e^{tk} q^{k-1}(1 - q) = (1 - q)e^t \sum_{k=1}^\infty \left(e^t q\right)^{k-1} = \frac{(1 - q)e^t}{1 - qe^t},$$

*und das konvergiert für all* $t < \log \frac{1}{q}$*. Bei* $t = \log \frac{1}{q}$ *haben wir ein Problem weil der Nenner gleich Null ist.*

*Die charakteristische Funktion ist*

$$\varphi_X(t) \quad = \quad \sum_{k=1}^{\infty} e^{itk} q^{k-1}(1-q) = (1-q)e^{it} \sum_{k=1}^{\infty} \left(e^{it}q\right)^{k-1} = \frac{(1-q)e^{it}}{1-qe^{it}},$$

*und diese konvergiert für alle $t \in \mathbb{R}$, denn $1 - qe^{it} \neq 0$ für alle $t \in \mathbb{R}$.*

*Wir können nachrechnen: $m_X(0) = 1$ und mit der Quotientenregel:*

$$m'_X(0) = \frac{d}{dt} \frac{(1-q)e^{it}}{1-qe^{it}}\Big|_{t=0} = \frac{(1-q)e^{it}}{(1-qe^{it})^2}\Big|_{t=0} = \frac{1}{1-q},$$

*und*

$$m''_X(0) = \frac{d^2}{dt^2} \frac{(1-q)e^{it}}{1-qe^{it}}\Big|_{t=0} = \frac{d}{dt} \frac{(1-q)e^{it}}{(1-qe^{it})^2}\Big|_{t=0} = \frac{(1-q)(1+2q-qe^t)e^{it}}{(1-qe^{it})^2}\Big|_{t=0} = \frac{1+q}{(1-q)^2}.$$

*Damit finden wir die Varianz* $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{1+q}{(1-q)^2} - \frac{1}{(1-q)^2} = \frac{q}{(1-q)^2}$.

**Beispiel 16** *Sei $X \simeq \mathcal{N}(\mu, \sigma^2)$, also $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Wir rechnen nach:*

$$tx - \frac{(x-\mu)^2}{2\sigma^2} = \frac{1}{2\sigma^2}\left(2t\sigma^2 x - (x-\mu)^2\right) = \frac{1}{2\sigma^2}\left(x - (t\sigma^2+\mu)\right)^2 + \mu t + \frac{t^2\sigma^2}{2}.$$

*Damit ist die moment-erzeugende Funktion:*

$$m_X(t) \quad = \quad \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = e^{\mu t + \frac{t^2\sigma^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(x-(t\sigma^2+\mu)\right)^2}{2\sigma^2}}\, dx = e^{\mu t + \frac{t^2\sigma^2}{2}}$$

*weil das letzte Integral über die Dichte von $Y \simeq \mathcal{N}(\mu + t\sigma^2, \sigma^2)$ ist. Also in diesem Fall existiert $m_X(t)$ für alle $t \in \mathbb{R}$.*

*Die charakteristische Funktion bekommen wir indem wir $t$ durch $it$ ersetzten:*

$$\varphi_X(t) \quad = \quad e^{i\mu t - \frac{t^2\sigma^2}{2}}.$$

*Wir können nachrechnen: $m_X(0) = 1$ und mit der Kettenregel:*

$$m'_X(0) = \frac{d}{dt} e^{\mu t + \frac{t^2\sigma^2}{2}}\Big|_{t=0} = (\mu + t\sigma^2)e^{\mu t + \frac{t^2\sigma^2}{2}}\Big|_{t=0} = \mu,$$

*und mit der Ketten- und Produktregel:*

$$\begin{aligned} m''_X(0) \quad &= \quad \frac{d^2}{dt^2} e^{\mu t + \frac{t^2\sigma^2}{2}}\Big|_{t=0} = \frac{d}{dt}(\mu + t\sigma^2)e^{\mu t + \frac{t^2\sigma^2}{2}}\Big|_{t=0} \\ &= \quad \sigma^2 e^{\mu t + \frac{t^2\sigma^2}{2}} + (\mu + t\sigma^2)^2 e^{\mu t + \frac{t^2\sigma^2}{2}}\Big|_{t=0} = \sigma^2 + \mu^2. \end{aligned}$$

*Damit finden wir die Varianz* $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$.

Die moment-erzeugenden und charakterischen Funktionen haben die folgenden nützlichen Eigenschaften gemeinsam:

**Proposition 17** *(a) Es gilt*

$$m_{aX+b}(t) = e^{bt} m_X(at) \qquad \text{und} \qquad \varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$$

*(b) Falls $X_1, \ldots, X_n$ unabhängige Zufallsvariablen sind, so gilt*

$$m_{X_1+\cdots+X_n}(t) = \prod_{j=1}^{n} m_{X_j}(t) \qquad \text{und} \qquad \varphi_{X_1+\cdots+X_n}(t) = \prod_{j=1}^{n} \varphi_{X_j}(t).$$

**Beweis.** Eigenschaft (a) folgt direkt aus der Definition: $m_{aX+b}(t) = \int e^{(ax+b)t} f_X(x)\,dx = e^{bt}\int e^{ax\,at} f_X(x)\,dx = e^{bt} m_X(at)$, und ähnlich für $\varphi_X(t)$ und falls $X$ diskret ist.

Die Unabhängigkeit in Eigenschaft (b) bedeutet das die Dichte des Zufallsvektors $\vec{X} = (X_1, \ldots, X_n)$ gleich $f_{\vec{X}} = f_{X_1} \cdots f_{X_n}$ ist. Deswegen

$$m_{X_1+\cdots+X_n}(t) = \int_{\mathbb{R}^n} e^{(X_1+\ldots X_n)t} f_{\vec{X}}\,d\vec{x} = \int_{\mathbb{R}^n} e^{X_1 t}\cdots e^{X_n t}\, f_{X_1}\cdots f_{X_n}\,dx_1\ldots dx_n = \prod_{j=1}^{n} m_{X_j}(t)$$

und ähnlich für $\varphi_{X_1+\cdots+X_n}(t)$ und falls die $X_j$ diskret sind. □

**Beispiel 18** *Seien die Zufallsvariablen $X_j, j = 1, \ldots, r$, unabhängig und exponentiell-verteilt mit Parameter $\lambda$. Dann sind die moment-erzeugenden Funktionen*

$$m_{X_j}(t) = \int_0^\infty e^{tx}\lambda e^{-\lambda x}\,dx = \lambda\int_0^\infty e^{(\lambda-t)x}\,dx = \frac{\lambda}{\lambda-t}, \qquad j = 1, \ldots, r,$$

*und sie konvergieren für $t < \lambda$. Dann sollte*

$$m_{X_1+\cdots+X_r}(t) = \prod_{j=1}^{r} m_{X_j}(t) = \prod_{j=1}^{r}\frac{\lambda}{\lambda-t} = \left(\frac{\lambda}{\lambda-t}\right)^r$$

*logisch die moment-erzeugende Funktion einer Gamma$(\lambda, r)$-verteilten Zufallsvariable $Y$ sein, und tatsächlich,*

$$\begin{aligned} m_Y(t) &= \int_0^\infty e^{ty}\,\frac{\lambda^r}{\Gamma(r)} y^{r-1}e^{-\lambda y}\,dy = \frac{\lambda^r}{\Gamma(r)}\int_0^\infty y^{r-1}e^{(\lambda-t)y}\,dy \\ &=_{u=(\lambda-t)y} \frac{\lambda^r}{(\lambda-t)^r}\,\frac{1}{\Gamma(r)}\int_0^\infty u^{r-1}e^{-u}\,du = \left(\frac{\lambda}{\lambda-t}\right)^r. \end{aligned}$$