Stochastik 1

Mitschrift von www.kuertz.name

Hinweis: Dies ist **kein offizielles Script**, sondern nur eine private Mitschrift. Die Mitschriften sind teweilse **unvollständig, falsch oder inaktuell**, da sie aus dem Zeitraum 2001–2005 stammen. Falls jemand einen Fehler entdeckt, so freue ich mich dennoch über einen kurzen Hinweis per E-Mail – vielen Dank!

Klaas Ole Kürtz (klaasole@kuertz.net)

Inhaltsverzeichnis

1 Zufallsexperimente

1.1 Definitionen

Eine Simulation, die zufällige Ergebnisse liefert, wird als *Zufallsexperiment* bezeichnet. Es besteht aus einem *Ergebnisraum* Ω, der sämtliche möglichen Ergebnisse $w \in \Omega$ etnhält. Als *Ereignisse* A, denen Wahrscheinlichkeiten zugeordnet werden sollen, werden Teilmengen $A \subseteq \Omega$ betrachtet. Den Ereignissen A wird eine Zahl $P(A) \in [0,1]$ zugeordnet, die als *Wahrscheinlichkeit von* A bezeichnet wird und die Wahrscheinlichkeit angibt, dass das Ergebnis des Zufallsexperiments in A liegt. Für Ereignisse A, B ist

- $A \cup B$: Ereignis, daß das Ergebnis in A oder B liegt
- $A \cap B$: Ereignis, daß das Ergebnis in A und B liegt
- Für $A_1, A_2, \ldots$, bedeutet $w \in \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k = \limsup A_n$, dass für alle n ein $k \geq n$ existiert so, dass gilt $w \in A_k$. Also entspricht $\limsup A_n$ das Ereignis, dass für unendlich viele k_i das Ergebnis $w \in A_{k_i}$ liegt.

1.2 Würfeln

Bei einem Würfel ist $\Omega = \{1, 2, 3, 4, 5, 6\}$. Das Ereignis, dass eine gerade Zahl gewürfelt wird, ist $A = \{2, 4, 6\}$. Es gilt $P(\{i\}) = \frac{1}{6}$, daher $P(A) = \frac{|A|}{6} = \frac{1}{2}$.

1.3 Kartenverteilung

$\Omega = \{\, (B_1, B_2, B_3, B_4) \mid B_i \subset \{1, 2, \ldots, 52\}, |B_i| = 13, B_i \cap B_j = \emptyset \text{ für } i \neq j\}$.
Es gilt: $P(\{w\}) = \frac{1}{53\,644\,737\,765\,488\,792\,839\,237\,440\,000}$

1.4 Lottospiel

Beim Lotto werden 6 Zahlen aus 49 Zahlen ohne Zurücklegen gezogen, dies entspricht dem Ergebnisraum $\Omega = \{\, (\omega_1, \ldots, \omega_6) \mid 1 \leq \omega_1 < \omega_2 < \ldots < \omega_6 \leq 49\}$. Dabei ist

$$|\Omega| = \frac{49 \cdot 48 \cdot \ldots \cdot 44}{6!} = \binom{49}{6} = 13\,983\,816$$

Da jede Ziehung gleich wahrscheinlich ist, liegt ein Laplace-Experiment vor, Festlegung: $P(\{\omega\}) = \frac{1}{|\Omega|}$ für alle $\omega \in \Omega$, zudem ist

$$P(A) = \sum_{w \in A} P(\{\omega\}) = \frac{|A|}{|\Omega|} \text{ für } A \subseteq \Omega$$

Wir geben einen festen Top $(b_1, \ldots, b_6)$ ab.

> **Frage:** Wie groß ist die Wahrscheinlichkeit, einen Dreier zu erzielen?

Dies entspricht dem Ereignis

$$A = \{ (\omega_1, \ldots, \omega_6) \mid |\{\omega_1, \ldots, \omega_6\} \cap \{b_1, \ldots, b_6\}| = 3\}$$

Es gilt: $|A| = \binom{6}{3} \cdot \binom{43}{3}$, denn es gibt $\binom{6}{3}$ Möglichkeiten, 3 Zahlen aus den vorgegebenen zu ziehen und $\binom{43}{3}$ Möglichkeiten, aus den nicht vorgegebenen Zahlen die übrigen drei zu ziehen. Damit ist $P(A) = \frac{|A|}{|\Omega|} = \frac{24\,692}{13\,983\,816}$. Analog ergibt sich für $k = 4, 5, 6$ die Wahrscheinlichkeit für k Richtige als

$$P(A_k) = \frac{\binom{6}{k} \cdot \binom{43}{6-k}}{\binom{49}{6}}$$

1.5 Spiegelungsprinzip

Das *Spiegelungsprinzip* ist ein kombinatorisches Hilfsmittel zum Abzählen von Elementen einer Menge. Betrachte hierzu Gitterpunkte $(u, v) \in \mathbb{Z} \times \mathbb{Z}$. Ausgehend von einem Startpunkt $(0, a)$ kann man in n Schritten einen Punkt $(n, a + 2k - n)$ durch einen Pfad mit k Aufwärtssprüngen und $n - k$ Abwärtssprüngen erreichen. Sprünge haben nur die Höhe ± 1.

Die Anzahl der Pfade, die $(0, a)$ mit $(n, a + 2k - n)$ verbindet, beträgt $\binom{n}{k} = \binom{n}{n-k}$. Seien $a > 0$ und $a + 2k - n > 0$.

> **Frage:** Wie viele Pfade gibt es, die $(0, a)$ mit $(n, a + 2k - n)$ verbinden und dabei die x-Achse erreichen?

Die Anzahl stimmt mit der Anzahl der Pfade, die $(0, a)$ mit $(n, -(a+2k-n))$ verbinden, überein und beträgt $\binom{n}{k+a} = \binom{n}{n-(k+a)}$, denn $-(a + 2k - n) = a - 2a - 2k + n = a + 2(n - (k + a)) - n$.

1.5.1 Anwendung des Spiegelungsprinzips

Zwei Kandidaten A, B in einer Wahl mit $2n$ Wahlberechtigten. Nach n ausgezählten Stimmen führt A mit a Stimmen $(0 < a \leq n)$.

> **Frage:** Wie groß ist die Wahrscheinlichkeit, daß A bis zum Ende der Wahl immer führt, also nicht mehr von B eingeholt wird?

Hier ist $\Omega = \{(\omega_1, \ldots, \omega_n) \mid w_i \in \{-1, +1\}\} = \{-1, +1\}^n$, wobei $\omega_i = 1$ eine Stimme für A, $\omega_i = -1$ eine Stimme für B bei der $(n+i)$-ten Auszählung bedeutet. Damit ist $|\Omega| = 2^n$, $P(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{2^n}$. Der Ergebnisraum ist

$$E = \left\{ \omega \in \Omega \;\middle|\; a + \sum_{i=1}^{m} \omega_i > 0 \;\forall\; 1 \leq m \leq n \right\}$$

E wird disjunkt zerlegt in Ereignisse $E_k = \{\omega \in E \mid a + \sum_{i=1}^{n} \omega_i = a + 2k - n\}$ für jedes k mit $n \geq k > \frac{n-a}{2}$ (d.h. k ist die Anzahl der Stimmen). $\omega \in E_k$ entspricht einem Ziehungsvorgang, bei dem A immer führt und am Ende einen Vorsprung von $a + 2k - n > 0$ hat.

- Für $n - k \geq a$ entspricht $\omega \in E_k$ genau einem Pfad, der $(0, a)$ und $(n, a + 2k - n)$ verbindet, ohne die x-Achse zu erreichen. Also ist $|E_k| = \binom{n}{k} - \binom{n}{a+k}$ und damit

$$P(E_k) = \frac{\binom{n}{k} - \binom{n}{a+k}}{2^n}$$

- Für $n - k < a$ kann B den Vorsprung nicht mehr einholen und $|E_k|$ ist $\binom{n}{k}$.

Zusammen erhält man:

$$P(E) = \sum_{k=\lceil \frac{n-a}{2} \rceil}^{n} P(E_k) = \left(\sum_{k=\lceil \frac{n-a}{2} \rceil}^{n-a} \frac{\binom{n}{k} - \binom{n}{a+k}}{2^n} \right) + \left(\sum_{k=n-a+1}^{n} \binom{n}{k} \right)$$

1.6 Das Nadelproblem von Buffon

Bislang wurden diskrete Zufallsexperimente betrachtet, nun kommen Experimente mit überabzählbar vielen Ausgängen vor. Betrachte eine Nadel der Länge 1 und eine Ebene, die durch Parallelen im Abstand 1 in Streifen unterteilt ist.

> **Frage:** Wie groß ist die Wahrscheinlichkeit, daß bei einem zufälligen Wurf die Nadel keine der Parallelen trifft?

Beschreibung des Zufallsexperiments durch

- Den Abstand $a \in [0, 1)$ des Mittelpunktes der Nadel zur nächstgelegenen unteren Parallelen; $a = 0$ bedeutet: Mittelpunkt ist auf einer Parallelen

- Der Winkel $\varphi \in [-\frac{\pi}{2}, \frac{\pi}{2})$ der Nadel mit der Senkrechten zur Parallelen, die durch den Mittelpunkt geht

Der Ereignisraum ist $\Omega = [0,1) \times [-\frac{\pi}{2}, \frac{\pi}{2})$. Naiv gesehen entsprechen Ereignisse meßbaren Flächen, die in Ω enthalten sind und zu einem Laplace-Experiment analog ist die Festlegung der Wahrscheinlichkeit eines Ereignisses A

$$P(A) = \frac{\text{Flächeninhalt von } A}{\text{Flächeninhalt von } \Omega}$$

Hierzu muß präzisiert werden, was meßbare Flächen und Flächeninhalte sind (siehe weitere Vorlesung). Die Nadel trifft genau dann eine der Parallelen, wenn $\frac{1}{2}\cos\varphi < a$ und $\frac{1}{2}\cos\varphi < 1-a$ ist, also $\frac{1}{2}\cos\varphi < \min\{a, 1-a\}$. Das Ereignis $A = \left\{ (a,\varphi) \mid \frac{1}{2}\cos\varphi < \min\{a,1-a\} \right\}$ und das Komplement dazu in Ω ist

$$A^c = \left\{ (a,\varphi) \mid a \leq \frac{1}{2}\cos\varphi \right\} \cup \left\{ (a,\varphi) \mid a \geq 1 - \frac{1}{2}\cos\varphi \right\}$$

Der Flächeninhalt von A^c ist $2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}\cos\varphi \, d\varphi = 2$, der Flächeninhalt von Ω ist π. Also gilt $P(A^c) = \frac{2}{\pi}$ und $P(A) = 1 - \frac{2}{\pi}$.

Das Nadelproblem kann zur näherungsweisen Bestimmung der Zahl π benutzt werden: Werfe hierzu n mal die Nadel. $X(n)$ bezeichne die zufällige Anzahl der Würfe, die keine der Parallelen treffen. Dann gilt mit dem Gesetz der großen Zahlen:

$$\lim_{n\to\infty} \frac{X(n)}{n} = P(A)$$

Somit läßt sich π näherungsweise berechnen durch $\pi \approx \frac{2\cdot n}{n - X(n)}$.

1.7 Grundbegriffe der Kombinatorik

- *Variationen:* Die Anzahl der k-Tupel von Elementen aus $\{1,\ldots,n\}$ beträgt $\left|\{1,\ldots,n\}^k\right| = n^k$, jedes k-Tupel ist eine Variation.

- *Permutationen:* Die Anzahl der k-Tupel mit unterschiedlichen Einträgen aus $\{1,\ldots,n\}$ beträgt $n\cdot(n-1)\cdot\ldots\cdot(n-k+1) = \frac{n!}{(n-k)!}$ Bei $k < n$ spricht man von einer *Permutation vom Umfang* k, für $k = n$ ist es eine eigentliche *Permutation*

- *Kombinationen:* Für $1 \leq k \leq n$ ist die Anzahl der k-elementigen Teilmengen von $\{1,\ldots,n\}$ gleich $\binom{n}{k} = \frac{n!}{(n-k)!}$. Dabei ist $\binom{n}{k}$ der *Binomialkoeffizient*, es gilt $\binom{n}{0} = 1$. Jede k-elementige Teilmenge entspricht einer Kombination vom Umfang k.

Verallgemeinerung ist der *Multinomialkoeffizient*: Gegeben sei eine n-elementige Menge B und Mächtigkeiten $k_1, \dots, k_m \geq 1$ mit $\sum_{i=1}^m k_i = n$. Ein Tupel $A = (A_1, \dots, A_m)$ ist eine Partition zu den vorgegebenen Mächtigkeiten $k_1, \dots, k_m$, wann gilt

1. $A_1, \dots, A_m$ ist eine disjunkte Zerlegung, d.h. $B = \bigcup_{i=1}^m A_i$ und $A_i \cap A_j = \emptyset$ für $i \neq j$.
2. $|A_i| = k_i$ für $1 \leq i \leq m$

Die Anzahl solcher Partitionen beträgt

$$\frac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_m!} =: \binom{n}{k_1, \dots, k_m}$$

Argumentation: $\binom{n}{k_1}$ Möglichkeiten für A_1, $\binom{n-k_1}{k_2}$ Möglichkeiten für A_2, $\dots$, $\binom{n-(k_1+\ldots+k_{m-2})}{k_{m-1}}$ Möglichkeiten für A_{m-1}, damit auch A_m festgelegt. Die Gesamtzahl ist dann

$$\frac{n!}{k_1!(n-k_1)!} \cdot \frac{(n-k_1)!}{k_2!(n-(k_1+k_2))!} \cdot \ldots \cdot \frac{(n-(k_1+\ldots+k_{m-2}))}{k_{m-1}! \cdot k_m!} = \frac{n!}{k_1! \cdot \ldots \cdot k_m!}$$

Man beachte, daß es bei den Partitionen auf die Reihenfolge ankommt, d.h. $(\{1\},\{2\}) \neq (\{2\},\{1\})$

2 Wahrscheinlichkeitsräume

Ziel ist jetzt die Motivation der Axiomatik von Колмогоров. Es stellen sich folgende Fragen:

1. Welchen Ereignissen sollen Wahrscheinlichkeiten zugerechnet werden?
2. Was für Eigenschaften sollte diese Zuordnung haben?

Bestandteile eines Zufallsexperimentes sind der *Ergebnisraum* Ω, die *Menge aller möglichen Ereignisse* $\mathfrak{A} \subseteq \mathcal{P}(\Omega)$ und eine *Zuordnungsfunktion* P, die Ereignissen Wahrscheinlichkeiten aus $[0, 1]$ zuordnet. **Naheliegende Forderungen** ist:

- Ω ist ein Ereignis.
- Ist A ein Ereignis, so ist auch $\Omega \setminus A = A^c$ ein Ereignis, das *Komplementärereignis*.
- Mit Ereignissen A, B ist auch $A \cup B$ ein Ereignis.

DEFINITION: Ein System $\mathfrak{A}$ von Teilmengen von Ω heißt *Mengenalgebra*, falls:

1. $\Omega \in \mathfrak{A}$
2. $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$
3. $A, B \in \mathfrak{A} \Rightarrow A \cup B \in \mathfrak{A}$

Aus der dritten Eigenschaft folgt, daß $\mathfrak{A}$ abgeschlossen ist gegenüber endlichen Vereinigungsbildungen. Für $|\Omega| < \infty$ ist die Begriffsbildung ausreichend. Startend mit den Elementarereignissen $\mathcal{E} = \{\{w\} \mid w \in \Omega\}$ erhält man die gesamte Potenzmenge als Menge von Ereignissen. Ist $\mathfrak{A}$ eine Mengenalgebra mit $\mathcal{E} \subset \mathfrak{A}$, so folgt: $\mathfrak{A} = \mathcal{P}(\Omega)$.

Für $|\Omega| = +\infty$ ist der Begriff der Mengenalgebra jedoch nicht ausreichend. Für $\Omega = \mathbb{N}$ beispielsweise ist ausgehend von $\mathcal{E}$ die Mengenalgebra $\mathfrak{A}$ beschränkt als:

$$\mathfrak{A} = \{A \subset \mathbb{N} \mid (|A| < \infty) \vee (|A^c| < \infty)\}$$

Damit sind aber etwa die geraden, die ungeraden oder die Primzahlen keine Ereignisse.

DEFINITION: Ein System $\mathfrak{A}$ von Teilmengen von Ω heißt *σ-Algebra*, falls:

1. $\Omega \in \mathfrak{A}$
2. $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$
3. $A_i \in \mathfrak{A} \; \forall \, i \in I \Rightarrow \bigcup_{i \in I} A_i \in \mathfrak{A}$ für alle abzahälbaren Indexmengen I

Dabei heißt $(\Omega, \mathfrak{A})$ *meßbarer Raum.*

Sei $\Omega = \mathbb{N}$, ist $\mathfrak{A}$ eine σ-Algebra mit $\mathcal{E} \subset \mathfrak{A}$, so folgt $\mathfrak{A} = \mathcal{P}(\Omega)$. Eine naheliegende Forderung an die Wahrscheinlichkeitszuordnung ist die Additivität, d.h. $P(A + B) = P(A) + P(B)$ für disjunkte Ereignisse A, B.

DEFINITION: Sei $\mathfrak{A}$ eine Mengenalgebra auf Ω. Eine Funktion $\mu : \mathfrak{A} \to [0, \infty]$ heißt *Inhalt*, falls gilt:

1. $\mu(\emptyset) = 0$
2. $\mu(A + B) = \mu(A) + \mu(B)$ für disjunkte $A, B \in \mathfrak{A}$

Durch Induktion erhält man die endliche Additivität $\mu(A_1 + \ldots + A_n) = \sum_{i=1}^{n} \mu(A_i)$ für disjunkte $A_1, \ldots, A_n$. Dies ist nicht ausreichend, um Wahrscheinlichkeitstheorie sinnvoll zu betreiben, man benötigt eine Konsistenz bezüglich der abzählbaren Vereinigungsbildung.

DEFINITION: Sei $(\Omega, \mathfrak{A})$ ein meßbarer Raum. Eine Abbildung $\mu : \mathfrak{A} \to [0, \infty]$ heißt *Maß*, falls gilt:

1. $\mu(\emptyset) = 0$
2. $\mu(\sum_{i \in I} A_i) = \sum_{i \in I} \mu(A_i)$ für abzählbare Familien $(A_i)_{i \in I}$ von paarweise disjunkten Ereignissen.

Gilt $\mu(\Omega) < \infty$, so spricht man von einem *endlichen Maß*. Für $\mu(\Omega) = +\infty$ liegt ein *unendliches Maß* vor. Ist $\mu(\Omega) = 1$, so spricht man von einem *Wahrscheinlichkeitsmaß* und bezeichnet es oft als P. Das Tripel $(\Omega, \mathfrak{A}, P)$ heißt dann *Wahrscheinlichkeitsraum*

2.1 Axiomatik von Колмогоров

Ein Zufallsexperiment wird beschrieben in einem mathematischen Modell durch einen Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit der Menge der potentiell möglichen Ergebnisse Ω, einem System der Ereignisse $\mathfrak{A}$[1] und einer Zuordnung $P : \mathfrak{A} \to [0,1]$. Prinzipielles Vorgehen:

$$\begin{array}{ccc} \textbf{Zufallsexperiment (reale Welt)} & \to & \textbf{Wahrscheinlichkeitsraum} \\ \textit{Lösung für reales Prob.} \uparrow & & \downarrow \textit{Analyse} \\ \textbf{Ergebnis} & \xleftarrow{\textit{Interpretation}} & \textbf{Lösung im Modell} \end{array}$$

Beispiel: Sie werfen nacheinander eine Münze. Bei jedem Wurf erhalten Sie mit der Wahrscheinlichkeit p Kopf und mit $1-p$ Zahl. Das Zufallsexperiment ist die Anzahl an Würfen bis das erste Mal Zahl fällt.

Setze $\omega_n = (1, \ldots, 1, 0)$, d.h. $\omega_{n,i} = 1$ bedeutet Kopf beim i-ten Wurf, $\omega_{n,i} = 0$ Zahl. Dann ist $\Omega = \{\omega_n \mid n \in \mathbb{N}\}$ und $|\Omega| = |\mathbb{N}|$ sowie $\mathfrak{A} = \mathcal{P}(\Omega)$. P wird nun definiert durch $P(\{\omega_n\}) = p^{n-1}(1-p)$. Somit ist für $A \subseteq \Omega$ die Wahrscheinlichkeit $P(A) := \sum_{\omega \in A} P(\{\omega\})$ und $P(\Omega) = \sum_{n \in \mathbb{N}} p^{n-1}(1-p) = 1$.

Für $A := \{\omega_{2k} \mid k \in \mathbb{N}\}$ und $B := \{\omega_{2k-1} \mid k \in \mathbb{N}\}$ ist

$$P(A) = \sum_{k=1}^{\infty} p^{2k-1}(1-p) = p \sum_{k=1}^{\infty} p^{(2k-1)-1}(1-p) = p \cdot P(B)$$

Mit $P(A) + P(B) = 1$ folgt $P(B) = \frac{1}{1+p}$.

2.2 Nicht-Existenz einer Flächenfunktion für $\mathbb{R}^k$

Satz: Es existiert keine Abbildung $\mu : \mathcal{P}(\mathbb{R}^k) \to [0, \infty]$ mit den Eigenschaften

1. μ ist Maß,
2. μ ist translationsinvariant (d.h. $\mu(A) = \mu(x + A)$ wobei $x + A = \{x + y \mid y \in A\}$),
3. falls A beschränkt ist, dann ist $\mu(A) < \infty$,
4. $\mu(\mathbb{R}^k) > 0$.

Beweis: Angenommen, ein solches μ würde existieren. Definiere eine Äquivalenzrelation $\sim$ durch $x \sim y :\Leftrightarrow x - y \in \mathbb{Q}^k$. Sei $\mathcal{K}$ die Menge aller Äquivalenzklassen. Das Auswahlaxiom liefert eine Abbildung $h : \mathcal{K} \to [0,1]^k$ mit

[1] Die Einführung einer σ-Algebra aller Teilmengen des Ergebnisraums, denen Wahrscheinlichkeiten zugeordnet werden, ist nur bei überabzählbaren Ergebnisräumen wirklich notwendig.

$h(K) \in K$ für alle $K \in \mathcal{K}$. Sei $E = \{ h(K) \mid K \in \mathcal{K} \}$.[2] Es wird nachgewiesen: Die Menge $\{ r + E \mid r \in \mathbb{Q}^k \}$ ist eine Partition von $\mathbb{R}^k$:

1. Die Menge bildet eine Zerlegung:

$$\bigcup_{r \in \mathbb{Q}^k} (r + E) = \bigcup_{r \in \mathbb{Q}^k} \bigcup_{e \in E} \{r + e\} = \bigcup_{e \in E} \bigcup_{r \in \mathbb{Q}^k} \{r + e\} = \bigcup_{e \in E} [e] = \mathbb{R}^k$$

2. Die Zerlegung ist disjunkt: Seien $r, r' \in \mathbb{Q}^k$ mit $r \neq r'$. Zu zeigen: $r + E \cap r' + E = \emptyset$. Annahme: $r + E \cap r' + E \neq \emptyset$. Dann existieren $x, x' \in E$ mit $r + x = r' + x'$ mit $x \neq x'$. Damit ist $x - x' = r' - r \in \mathbb{Q}^k$, also $x \sim x'$, damit können aber x und x' nicht beide in E liegen, Widerspruch!

Damit wurde eine Menge $E \subset \mathbb{R}^k$ gefunden mit $\mathbb{R}^k = \sum_{r \in \mathbb{Q}^k} (r + E)$. Die geforderten Eigenschaften von μ liefern:[3]

$$0 < \mu(\mathbb{R}^k) = \mu\left(\sum_{r \in \mathbb{Q}^k} (r + E) \right) \overset{(\star)}{=} \sum_{r \in \mathbb{Q}^k} \mu(r + E) = \sum_{r \in \mathbb{Q}^k} \mu(E)$$

Somit ist $\mu(E) > 0$. Sei $Q := \mathbb{Q}^k \cap [0,1]^k$. Dann ist $\sum_{r \in Q} (r + E) \subseteq [0,2]^k$ und damit

$$\infty > \mu\left([0,2]^k\right) \geq \mu\left(\sum_{r \in Q} (r + E) \right) = \sum_{r \in Q} \mu\left(r + E\right) = \sum_{r \in Q} \mu(E) = \infty$$

Widerspruch! Also existiert keine solche Abbildung μ. □

An dem Problem gearbeitet haben E. Borel (1898) und H. Lebesgue (1902/1904), wir lernen Borelsche Mengen und das Lebesgue-Integral kennen.

2.3 Borelsche σ-Algebra

Die *Borelsche σ-Algebra* wird definiert als die kleinste σ-Algebra, die sämtliche Intervalle enthält. Elemente der Borelschen σ-Algebra heißen *Borelsche Mengen*.

Diese Definition beruht auf folgendem Sachverhalt:

---

[2] „E steht für *eklige Menge*.“

[3] wobei $(\star)$ durch die Abzählbarkeit von $\mathbb{Q}^k$ aus den Maßeigenschaften von μ folgt

SATZ: Ω sei eine Menge, $\mathcal{E} \subset \mathcal{P}(\Omega)$. Definiere

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{E} \subset \mathfrak{A} \\ \mathfrak{A}\ \sigma-\text{Algebra}}} \mathfrak{A}$$

Dann ist $\sigma(\mathcal{E})$ eine σ-Algebra, und für jede σ-Algebra $\mathfrak{A}$ mit $\mathcal{E} \subset \mathfrak{A}$ ist $\mathfrak{A} \supset \sigma(\mathcal{E})$, d.h. $\sigma(\mathcal{E})$ ist die kleinste σ-Algebra, die $\mathcal{E}$ umfaßt.

BEWEIS: Folgt sofort aus der Definition. Als Erzeugendensystem der Borelschen σ-Algebra wird $\mathcal{E}^k$ benutzt:

$$\mathcal{E}^k = \left\{ (a,b] \mid a,b \in \mathbb{R}^k;\ a \leq b \right\}$$

Mit $a = (a_1, \ldots, a_k)$ und $b = (b_1, \ldots, b_k)$ ist dabei

$$(a,b] := \{ x \mid a_i < x_i \leq b_i\ \forall\ i = 1, \ldots, k \}$$

Ein Erzeugendensystem $\mathcal{E}^k$ hat folgende Eigenschaften:

- Es ist ist $\cap$-stabil: $A, B \in \mathcal{E}^k \Rightarrow A \cap B \in \mathcal{E}^k$.
- Für $A, B \in \mathcal{E}^k$ mit $A \subseteq B$ ist $A \setminus B = \sum_{j=1}^n A_j$ mit $A_1, \ldots, A_n \in \mathcal{E}^k$.

Ein Mengensystem mit diesen beiden Eigenschaften heißt *Halbring*.

DEFINITION: Die Borelsche σ-Algebra ist $\mathcal{B}^k = \sigma(\mathcal{E}^k)$.

Faustregel: Alle Teilmengen, die im Mathematikstudium auftreten, liegen in $\mathcal{B}^k$ (bis auf die oben definierte Menge E)[4].

2.4 Das Lebeguesche Maß

Ziel: Teilmengen des $\mathcal{R}^k$ einen Inhalt, d.h. Längen/Flächen/... zuzuordnen. Der *Hauptsatz der Maßtheorie* lautet:

SATZ: Es existiert genau ein Maß $\lambda^k : \mathcal{B}^k \to [0, \infty]$ mit

$$\lambda^k((a,b]) = \prod_{i=1}^k (b_i - a_i)\ \forall\ (a,b] \in \mathcal{E}^k$$

D.h. für jedes Intervall I ist $\lambda^k(I)$ der Inhalt des Intervalls.

[4]und bis auf Prof. Spinas' „Axiomatische Mengenlehre“

Der Satz beinhaltet die Existenz- und die Eindeutigkeitsaussage. Letzere benutzt eine Beweismethode von Wichtigkeit für die Wahrscheinlichkeitstheorie („Dynken-Systeme“), folgt später. Der Beweis der Existenzaussage (von CARATHÉODORY 1914)als Skizze (ausführlich in jedem Buch über Maßtheorie):

BEWEIS: Starte mit einem Halbring $\mathcal{H}$ (Beispiel: $\mathcal{E}^k$) und einer Abbildung $\nu : \mathcal{H} \to [0, \infty)$ mit den Eigenschaften $\nu(\emptyset) = 0$, aus $A \subseteq B$ folgt $\nu(A) \leq \nu(B)$ und $\nu\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n \nu(A_i)$ für $A_1, \ldots, A_n \in \mathcal{H}$ mit $\sum_{i=1}^n A_i \in \mathcal{H}$. Definiere zu ν eine Abbildung $\nu^* : \mathcal{P}(\Omega) \to [0, \infty]$ mit

$$\nu^* : \ A \mapsto \inf \left\{ \sum_{i \in \mathbb{N}} \nu(A_i) \ \middle| \ (A_i)_{i \in \mathbb{N}} \in \mathcal{H}^{\mathbb{N}} \text{ mit } A \subseteq \bigcup_{i \in \mathbb{N}} A_i \right\}$$

LEMMA: In der obigen Situation gilt $\nu^*(\emptyset) = 0$, $\nu^*(A) \leq \nu^*(B)$ für $A \subseteq B$ und $\nu^*\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \nu^*(A_n)$. Damit heißt ν^* *äußeres Maß*.

SATZ: Sei ν^* äußeres Maß. Sei

$$\mathfrak{A} = \{ A \subseteq \Omega \mid \nu^*(B) = \nu^*(B \cap A) + \nu^*(B \cap A^c) \ \forall \ B \subseteq \Omega \}$$

Dann gilt: $\mathfrak{A}$ ist σ-Algebra und $\nu^*|_{\mathfrak{A}}$ ist ein Maß.

Mit dieser Vorgehensweise erhält man das *Lebesguesche Maß* λ^k:

1. Starte mit Halbring $\mathcal{E}^k$, definiere $\lambda^k((a, b]) = \prod_{i=1}^k (b_i - a_i)$.
2. Definiere zu $\nu = \lambda^k$ das äußere Maß ν^*.
3. Betrachte $\mathfrak{A}$ (Carathéodory) zu ν^*, beachte dabei $\sigma(\mathcal{E}^k) \subseteq \mathfrak{A}$
4. Definiere $\lambda^k = \nu^*|_{\mathcal{B}^k}$

3 Rechnen mit Wahrscheinlichkeiten

3.1 Rechenregeln für Wahrscheinlichkeitsmaße

Diese Regeln basieren auf σ-Additivität $P(\sum_{i\in I} A_i) = \sum_{i\in I} P(A_i)$ für abzählbar viele, disjunkte A_i.

- $P(A^c) = 1 - P(A)$
- $A \subseteq B \Rightarrow P(B) = P(A) + P(B \setminus A)$
- für zwei Mengen:

$$\begin{aligned} P(A \cup B) &= P(A \setminus (A \cap B)) + P(B \setminus (A \cap B)) + P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

  für drei Mengen:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) - P(A \cap C) + P(C) \\ &\quad -P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

  für n Mengen:

$$P(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{(i_1,\ldots,i_k) \\ 1\le i_1<\ldots<i_k\le n}} P(A_{i_1} \cap \ldots \cap A_{i_k})$$

  Beweis: mit Induktion: Induktionsanfang für $n = 2$ siehe oben. Induktionsschritt:

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right) \\
&= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right) \\
\text{(IV)} \quad &= \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{(i_1,\ldots,i_k) \\ 1\le i_1<\ldots<i_k\le n}} P(A_{i_1} \cap \ldots \cap A_{i_k}) \\
&\quad + P(A_{n+1}) - P\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right) \\
&\quad + \sum_{k=1}^{n} (-1)^{k} \sum_{\substack{(i_1,\ldots,i_k) \\ 1\le i_1<\ldots<i_k\le n}} P(A_{i_1} \cap \ldots \cap A_{i_k} \cap A_{n+1}) \\
&= P(A_1) + \ldots P(A_n) + P(A_{n+1}) \\
&\quad + \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{(i_1,\ldots,i_k) \\ 1\le i_1<\ldots<i_k\le n}} P(A_{i_1} \cap \ldots \cap A_{i_k}) \\
&\quad + \sum_{k=2}^{n} (-1)^{k+1} \sum_{\substack{(i_1,\ldots,i_{k-1}) \\ 1\le i_1<\ldots<i_{k-1}\le n}} P(A_{i_1} \cap \ldots \cap A_{i_{k-1}} \cap A_{n+1}) \\
&\quad + (-1)^n P(A_1 \cap \ldots \cap A_n \cap A_{n+1}) \\
&= \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{(i_1,\ldots,i_k) \\ 1\le i_1<\ldots<i_k\le n}} P(A_{i_1} \cap \ldots \cap A_{i_k})
\end{aligned}$$

Die Rechenregeln gelten für allgemeine Maße, falls „$\infty - \infty$“ nicht auftritt.

3.2 Modellierung eines parapsychologischen Experiments

Betrachtet[5] seien die Zahlen $1, \dots, N$ in einer Anordnung $\pi = (\pi_1, \dots, \pi_N)$, d.h. betrachtet wird eine spezifische Permutationen der Zahlen $1, \dots, N$ Zufällig wird eine weitere Permutation $\sigma = (\sigma_1, \dots, \sigma_n)$ gewählt, wobei jede Permutation σ gleichwahrscheinlich sei, d.h. $P(A) = \frac{|A|}{N!}$.

Also ist $\Omega = \{\sigma \mid \sigma \text{ Permutation von } (1, \dots, N)\}$, $\mathfrak{A} = \mathcal{P}(\Omega)$ und $P(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{N!}$ für $A \in \mathfrak{A}$. Die Frage nach der Wahrscheinlicheit von

$$B_k = \{\sigma \mid |\{i \mid \sigma_i = \pi_i\}| = k\}$$

für alle $k = 0, \dots, N$. Ohne Einschränkung sei $\pi = (1, \dots, N)$. Falls $\sigma_i = i$, so nennt man i einen Fixpunkt von σ, d.h.

$$B_k = \{\sigma \mid \sigma \text{ besitzt } k \text{ Fixpunkte}\}$$

Frage: Wie groß ist die Wahrscheinlichkeit keiner Übereinstimmung? Es ist $B_0 = \{\sigma \mid \sigma(i) \neq i \ \forall\, i\}$ und

$$B_0^c = \{\sigma \mid \ \exists\, i \in \{1, \dots, N\} \ : \ \sigma(i) = i\} = \bigcup_{i=1}^{N} \{\sigma \mid \sigma(i) = i\}$$

Anwendung der Rechenregeln liefert

$$P(B_0^c) = P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{k=1}^{N} (-1)^{k+1} \sum_{1 \le i_1 < \dots < i_k \le N} P(A_{i_1} \cap \dots \cap A_{i_k})$$

Dabei ist $A_{i_1} \cap \dots \cap A_{i_k} = \{\sigma \mid \sigma(i_j) = i_j \ \forall\, j = 1, \dots, k\}$ und $|A_{i_1} \cap \dots \cap A_{i_k}| = (N-k)!$, also ist $P(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(N-k)!}{N!}$ und damit

$$P(B_0^c) = \sum_{k=1}^{N} (-1)^{k+1} \cdot \binom{N}{k} \cdot \frac{(N-k)!}{N!}$$

Damit ist $P(B_0) = 1 - P(B_0^c) = \sum_{k=0}^{N} (-1)^{k+1} \frac{1}{k!}$. Beachte $e = \sum_{n=0}^{\infty} \frac{1}{k!}$, also ist

$$\lim_{N \to \infty} P(B_0) = \frac{1}{e} \approx 36.7879\%$$

[5]„Sie gewinnen doch ganz bestimmt mit hoher Wahrscheinlichkeit…! (…) Also ich würde nicht mit Ihnen spielen, das wäre ja unfair, wenn ich Ihnen das ganze Geld aus der Tasche ziehe!“

Für $N = 5$ ist $P(B_0) = 36.7\%$, bei $N = 10$ ist es 36.788%.

Alternative Möglichkeit: Betrachte Abhängigkeit in N: Sei d_N die Anzahl der Permutationen ohne Fixpunkte zu $(1, \ldots, N)$. Offensichtlich ist $d_1 = 0$, $d_2 = 1$, $d_3 = 2$. **Versuch:** Finde eine Rekursion, d.h. suche d_N durch $d_{N-1}, \ldots, d_{N-k}$ (mit k unabhängig von N!) auszudrücken:

$$\begin{aligned} d_N &= (N-1) \cdot |\{\sigma \mid \sigma \text{ fixpunktfrei}, \sigma(1) = 2\}| \\ &= (N-1) \cdot (d_{N-2} + d_{N-1}) \end{aligned}$$

Weitere Auswertung:

$$\begin{aligned} d_N - N \cdot d_{N-1} &= (-1)(d_{N-1} - (N-1)d_{N-2}) \\ &= (-1)(-1)(d_{N-2} - (N-2)d_{N-3}) \\ &= (-1)(-1)(-1)(d_{N-3} - (N-3)d_{N-3}) \\ &= (-1)^N(d_2 - 2d_1) = (-1)^N \end{aligned}$$

Damit ist $d_N = N d_{N-1} + (-1)^N$, Lösung also

$$d_N = N! \cdot \left(1 - \frac{1}{1!} + \frac{1}{2!} - \ldots + (-1)^N \frac{1}{N!}\right)$$

Es bleibt zu bestimmen: $P(B_k)$ für $k = 1, \ldots, N$. Es ist

$$\begin{aligned} B_k &= \{\sigma \mid |\{i \mid \sigma_i = i\}| = k\} \\ &= \sum_{1 \le i_1 < \ldots < i_k \le N} \{\sigma \mid \sigma(i_j) = i_j \; \forall\, j = 1, \ldots, k;\; \sigma(j) \neq j \text{ sonst}\} \\ |B_k| &= \binom{N}{k} |\{\text{Permutation von } N-k \text{ Elementen ohne Fixpunkte}\}| \\ P(B_k) &= \frac{1}{N!}\binom{N}{k}(N-k)! P(\{\pi \mid \pi \text{ Perm. von } (1, \ldots, N-k) \text{ o. FP}\}) \\ &= \frac{1}{N!}\binom{N}{k}(N-k)! \sum_{j=0}^{N-k} (-1)^j \frac{1}{j!} \\ &= \frac{1}{k!} \sum_{j=0}^{N-k} (-1)^j \frac{1}{j!} \end{aligned}$$

Es ist also $\lim_{N\to\infty} P(B_k) = \frac{1}{k!e}$

3.3 Unendliche Folgen von Ereignissen

Sei $(A_i)_{i \in \mathbb{N}}$ eine Folge von Ereignissen. Von Interesse sind (die Beispiele beziehen sich auf eine Folge von Würfel-Würfen mit A_i als Ereignis, daß 1 geworfen wird):

Ereignis	*Ergebnis liegt...*	*Beispiel*
$\bigcup_{i \in \mathbb{N}} A_i$	in mindestens einem A_i	mindestens eine 1 geworfen
$\bigcap_{i \in \mathbb{N}} A_i$	in allen A_i	sämtliche Würfe 1
$\limsup A_i = \bigcap_{i \in \mathbb{N}} \bigcup_{n \geq i} A_n$	in unendlich vielen A_i	unendlich viele 1 geworfen
$\liminf A_i = \bigcup_{i \in \mathbb{N}} \bigcap_{n \geq i} A_n$	in allen bis auf endlich viele A_i	ab irgednwann nur noch 1

Um solche Ereignisse mit Wahrscheinlichkeiten zu versehen, benötigen wir die folgende Eigenschaft für abzählbar-unendliche disjunkte Vereinigungen:

$$P\left(\sum_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i)$$

Satz: Sei $(A_i)_{i \in \mathbb{N}}$ eine Folge von Ereignissen. Dann gilt:

1. Falls $(A_i)_{i \in \mathbb{N}}$ *monoton wachsend* ist ($A_i \subseteq A_{i+1}$), so gilt

$$P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \lim_{n \to \infty} P(A_n)$$

2. Falls $(A_i)_{i \in \mathbb{N}}$ *monoton fallend* ist ($A_i \supseteq A_{i+1}$), so gilt

$$P\left(\bigcap_{i \in \mathbb{N}} A_i\right) = \lim_{n \to \infty} P(A_n)$$

Beweis:

1. Da $(A_i)_{i\in\mathbb{N}}$ wachsend ist, gilt

$$\begin{aligned}
\bigcup_{i\in\mathbb{N}} A_i &= \underbrace{A_1 + (A_2 \setminus A_1) + (A_3 \setminus A_2) + \ldots + (A_n \setminus A_{n-1})}_{A_n} + \ldots \\
&= A_1 + \sum_{i=2}^{\infty} (A_i \setminus A_{i-1}) \\
P\left(\bigcup_{i\in\mathbb{N}} A_i\right) &= P(A_1) + \sum_{i=2}^{\infty} P(A_i \setminus A_{i-1}) \\
&= P(A_1) + \lim_{n\to\infty} \sum_{i=2}^{n} P(A_i \setminus A_{i-1}) \\
&= \lim_{n\to\infty} P(A_n)
\end{aligned}$$

2. Es gilt

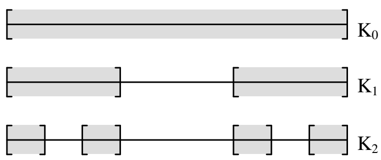
$$P\left(\bigcap_{i\in\mathbb{N}} A_i\right) = 1 - P\left(\left(\bigcap_{i\in\mathbb{N}} A_i\right)^c\right) = 1 - P\left(\bigcup_{i\in\mathbb{N}} (A_i^c)\right)$$

   Da $(A_i)_{i\in\mathbb{N}}$ fallend ist, ist $(A_i^c)_{i\in\mathbb{N}}$ monoton wachsend und mit eben gezeigtem ersten Teil ergibt sich

$$P\left(\bigcap_{i\in\mathbb{N}} A_i\right) = 1 - \lim_{n\to\infty} P(A_n^c) = 1 - \lim_{n\to\infty} (1 - P(A_n)) = \lim_{n\to\infty} P(A_n)$$

3.3.1 Cantor-Menge

Sei $K_0 = [0,1]$, $K_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, …und K_n ergibt sich aus K_{n-1} durch Entfernen der mittleren Drittel, wobei die Randpunkte nicht entfernt werden[6]:



Die Cantor-Menge C ist definiert als $C := \bigcap_{n=0}^{\infty} K_n$, wobei $(K_n)_{n\in\mathbb{N}}$ eine fallende Menge bildet. Mit obigem Satz ergibt sich

$$\lambda(C) = \lambda\left(\bigcap_{n=0}^{\infty} K_n\right) = \lim_{n\to\infty} \lambda(K_n) = \lim_{n\to\infty} \left(\frac{2}{3}\right)^n = 0$$

Die Menge C ist jedoch überabzählbar, so daß wir eine überabzählbare Menge mit Lebesgue-Maß 0 gefunden haben!

[6] „…können Sie ja formal aufschreiben, wenn Sie Lust haben!“

3.3.2 Borell-Cantelli-Lemma

Satz: Sei $(A_i)_{i\in\mathbb{N}}$ eine Folge von Ereignissen. Dann gilt:

1. $$P\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leq \sum_{i\in\mathbb{N}} P(A_i)$$

2. $$P\left(\limsup_{i\in\mathbb{N}} A_i\right) = 0 \text{ falls } \sum_{i\in\mathbb{N}} P(A_i) < \infty$$

Beweis:

1. Für $A_1, \ldots, A_n$ gilt stets $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$, wie man mit Induktion sofort einsieht:
$$\begin{aligned} P\left(\bigcup_{i=1}^{n+1} A_i\right) &= P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \\ &\leq P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) \\ &= \sum_{i=1}^{n+1} P(A_i) \end{aligned}$$
Es folgt mit $B_i = \bigcup_{k=1}^i A_k$ (aufsteigend):
$$\begin{aligned} P(\bigcup_{i=1}^{\infty} A_i) &= P(\bigcup_{i=1}^{\infty} B_i) \\ &= \lim_{i\to\infty} P(B_i) = \lim_{i\to\infty} P\left(\bigcup_{k=1}^{i} A_k\right) \\ &\leq \lim_{i\to\infty} \sum_{k=1}^{i} P(A_k) = \sum_{i=1}^{\infty} P(A_i) \end{aligned}$$

2. Für $\limsup A_i = \bigcap_{n=1}^\infty \bigcup_{i\geq n} B_i$ mit $\bigcup_{i\geq n} B_i$ fallend in n, also ist
$$P\left(\limsup_{i\to\infty} A_i\right) = \lim_{n\to\infty} P\left(\bigcup_{i\geq n} A_i\right) \leq \lim_{n\to\infty} \sum_{i=n}^{\infty} P(A_i)$$
Dabei geht $\sum_{i=n}^\infty P(A_i)$ gegen 0, falls es konvergiert.

4 Bedingte Wahrscheinlichkeiten

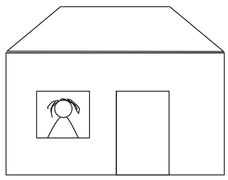
Betrachte Ereignisse A, B mit $P(B) > 0$. Wie groß ist die Wahrscheinlichkeit von A, gegeben das Eintreten von B. Wir sprechen dann von einer *bedingten Wahrscheinlichkeit*:

> Definition: Für A, B mit $P(B) > 0$ ist die *bedingte Wahrscheinlichkeit* von A gegeben B definiert durch $P(A|B) = \frac{P(A\cap B)}{P(B)}$.

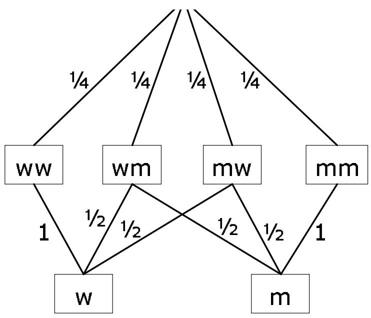
Beispiele:

1. *Kartenspiel:* drei zweiseitig markierte Karten: rot-rot, grün-grün und rot-grün; es wird zufällig eine der drei Karten gewählt und zufällig eine Seite der Karte ausgewählt; zu erraten ist, was auf der anderen Seite ist. Gewinnstrategie: die gleiche Farbe nennen, die auch zu sehen ist; Gewinn in $\frac{2}{3}$ der Fälle

2. *Zwei-Kinder-Problem:* Es ziehen neue Nachbarn mit zwei Kindern ein, man sieht am Fenster ein Mädchen (w):

   Wie groß ist die Wahrscheinlichkeit, daß auch das andere Kind ein Mädchen ist? Unter der Annahme, daß ww, wm, mw und mm gleichwahrscheinlich sind und mm nicht auftreten kann, könnte man die Wahrscheinlichkeit als $\frac{1}{3}$ beziffern - es gilt jedoch: $P(ww|w) = \frac{P(ww)}{P(w)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$

3. *Ziegenproblem:* Kandidat in einer Quizshow, drei Türen, hinter zweien sind Ziegen, hinter einer ein Auto. Der Kandidat wählt eine, der Quizmaster öffnet eine der beiden anderen Türen (mit einer Ziege dahinter)

und der Kandidat erhält die Möglichkeit zu wechseln. Soll er wechseln? Ja, wenn er wechselt, gewinnt er mit der Wahrscheinlichkeit $\frac{2}{3}$! Einfache Argumentation: Wenn er zu Anfang eine Ziege trifft (Wahrscheinlichkeit $\frac{2}{3}$), so muß der Quizmaster die andere Ziege zeigen und der Kandidat wechselt zum Auto!

Empfehlung: G. v. Randow: „Das Ziegenproblem"(rororo) und N. Henze: „Stochastik für Einsteiger"(vieweg)

4.1 Rechenregeln für bedingte Wahrscheinlichkeiten

1. Nach Definition gilt

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

2. Wegen $P\left(\bigcap_{i=1}^{n} A_i\right) = P\left(A_n \cap \bigcap_{i=1}^{n-1} A_i\right) = P\left(A_n | \bigcap_{i=1}^{n-1} A_i\right) \cdot P\left(\bigcap_{i=1}^{n-1} A_i\right)$ gilt

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right)$$

3. Seien $A_1, \ldots, A_n$ disjunkte Ereignisse mit $\sum_{i=1}^{n} A_i = \Omega$. Dann gilt für jedes Ereignis A:

$$P(A) = \sum_{i=1}^{n} P(A|A_i)P(A_i)$$

denn $\sum_{i=1}^{n} P(A|A_i)P(A_i) = \sum_{i=1}^{n} P(A \cap A_i) = P(\sum_{i=1}^{n}(A \cap A_i)) = P(A \cap \sum_{i=1}^{n} A_i) = P(A)$.

4. Seien $A_1, \ldots, A_n$ wieder disjunkte Ereignisse mit $\sum_{i=1}^{n} A_i = \Omega$. Dann gilt für jedes Ereignis A und $j = 1, \ldots, n$ die *Formel von Bayes*:

$$P(A_j|A) = \frac{P(A|A_j)P(A_j)}{\sum_{i=1}^{n} P(A|A_i)P(A_i)}$$

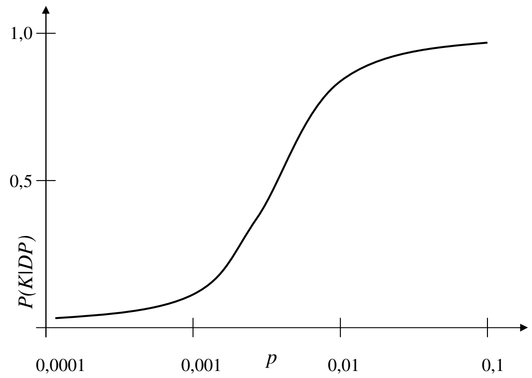
denn auf der rechten Seite liegt vor: $\frac{P(A \cap A_j)}{P(A)}$.

4.1.1 Aussagekraft von medizinischen Testverfahren

Eine Person hat sich einem Test zur Früherkennung unterzogen und das Ergebnis ist positiv. Mit welcher Wahrscheinlichkeit liegt die Erkrankung vor? Sei $P(K) = p$ die Wahrscheinlichkeit des Auftretens in der Bevölkerungsgruppe und $P(N) = 1 - p$. Die Wahrscheinlichkeit für eine positive Diagnose bei tatsächlicher Erkrankung sei $P(DP|K) = q_1$, die Wahrscheinlichkeit für eine (falsche) positive Diagnose bei gesunden Personen sei $P(DP|N) = q_2$ (d.h. q_1 und q_2 sind die Güteparameter für das Testverfahren). Anwendung der Formel von Bayes liefert:

$$P(K|DP) = \frac{P(DP|K)P(K)}{P(DP|K)P(K) + P(DP|N)P(N)} = \frac{q_1 p}{q_1 p + q_2(1-p)}$$

Im Beispiel eines Tests für HIV-Infektionen ist $q_1 = 0.998$ und $q_2 = 0.002$, die Wahrscheinlichkeit hängt dann stark von p ab:



4.2 Stochastische Unabhängigkeit

Definition: Zwei Ereignisse A und B sind *stochastisch unabhängig*, falls

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

Daraus folgt (falls $P(A)$ und $P(B)$ größer 0) $P(A|B) = P(A)$ und $P(B|A) = P(B)$.

4.2.1 Ziehen mit und ohne Zurücklegen

In einer „Urne“ seien k weiße und l grüne Kugeln, d.h. insgesamt $n = k + l$. Es wird zwei mal gezogen, damit ist $\Omega = \{(w, w), (g, w), (w, g), (g, g)\}$.

- Beim Ziehen mit Zurücklegen: $P(\{(w, w)\}) = \frac{k}{n} \cdot \frac{k}{n}$, $P(\{(w, g)\}) = P(\{(g, w)\}) = \frac{k}{n} \cdot \frac{l}{n}$, $P(\{(g, g)\}) = \frac{l}{n} \cdot \frac{l}{n}$

- Beim Ziehen ohne Zurücklegen: $P(\{(w,w)\}) = \frac{k}{n} \cdot \frac{k-1}{n-1}$, $P(\{(w,g)\}) = \frac{k}{n} \cdot \frac{l}{n-1}$, $P(\{(g,w)\}) = \frac{l}{n} \cdot \frac{k}{n-1}$, $P(\{(g,g)\}) = \frac{l}{n} \cdot \frac{l-1}{n-1}$

Seien nun die Ereignisse A als *weiß im ersten Zug* und B als *weiß im zweiten Zug* definiert, d.h. $A = \{(w,w),(w,g)\}$ und $B = \{(w,w),(g,w)\}$. Es gilt $P(A \cap B) = \{(w,w)\}$.

- Mit Zurücklegen ergibt sich:

$$\begin{aligned} P(A) &= \frac{k}{n} \cdot \frac{k}{n} + \frac{k}{n} \cdot \frac{l}{n} = \frac{k}{n} \\ P(B) &= \frac{k}{n} \cdot \frac{k}{n} + \frac{l}{n} \cdot \frac{k}{n} = \frac{k}{n} \\ P(A \cap B) &= \frac{k}{n} \cdot \frac{k}{n} = P(A) \cdot P(B) \end{aligned}$$

- Ohne Zurücklegen kommt heraus:

$$\begin{aligned} P(A) &= \frac{k}{n} \cdot \frac{k-1}{n-1} + \frac{k}{n} \frac{l}{n-1} = \frac{k}{n} \\ P(B) &= \frac{k}{n} \cdot \frac{k-1}{l-1} + \frac{l}{n} \frac{l}{n-1} = \frac{k}{n} \\ P(A \cap B) &= \frac{k}{n} \frac{k-1}{n-1} \neq P(A)P(B) \end{aligned}$$

4.2.2 Verallgemeinerung der Definition

DEFINITION: Sei $(A_i)_{i \in I}$ eine Familie von Ereignissen. Diese wird als *stochastisch unabhängig* bezeichnet, falls gilt:

$$P\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} P(A_j) \ \forall \ J \subseteq I$$

SATZ: Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge von Ereignissen, die stochastisch unabhängig sind. Dann gilt:

$$\sum_{n \in \mathbb{N}} P(A_n) = \infty \Rightarrow P\left(\limsup_n A_n\right) = 1$$

BEWEIS: Es gilt:

$$\begin{aligned}
P\left(\left(\limsup_n A_n\right)^c\right) &= P\left(\left(\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} A_k\right)^c\right) \\
&= P\left(\bigcup_{n\in\mathbb{N}}\bigcap_{k\geq n} A_k^c\right) \\
&\leq \sum_{n\in\mathbb{N}} P\left(\bigcap_{k\geq n} A_k^c\right) \\
&= \sum_{n\in\mathbb{N}} \underbrace{\lim_{\substack{m\to\infty \\ m\geq n}} P\left(\bigcap_{k=n}^m A_k^c\right)}
\end{aligned}$$

Es genügt also zu zeigen, daß der hintere Teil gleich 0 ist. Es gilt (mit der Voraussetzung der stochastischen Unabhängigkeit und der Abschätzung $1 - x \leq e^{-x}$):

$$\begin{aligned}
\lim_{\substack{m\to\infty \\ m\geq n}} P\left(\bigcap_{k=n}^m A_k^c\right) &= \lim_{\substack{m\to\infty \\ m\geq n}} \prod_{k=n}^m (1 - P(A_n)) \\
&\leq \lim_{\substack{m\to\infty \\ m\geq n}} \prod_{k=n}^m e^{-P(A_n)} \\
&\leq \lim_{\substack{m\to\infty \\ m\geq n}} e^{-\sum_{k=n}^m P(A_n)} = 0
\end{aligned}$$

In diesem Beweis wurde die (noch nicht bewiesene) Aussage benutzt, daß mit der Unabhängigkeit der Ereignisse ebenfalls die Unabhängigkeit der Komplemente vorliegt. Betrachtet man zunächst die Ereignisse, so ergibt sich

$$P(A\cap B^c) = P(A) - P(A\cap B) = P(A) - P(A)\cdot P(B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

SATZ: Sei $(A_i)_{i\in I}$ eine Familie von stochastisch unabhängigen Ereignissen. Dann gilt für alle endlichen $J, K \subseteq I$ mit $J \cap K = \emptyset$:

$$P\left(\left(\bigcap_{j\in J} A_j\right) \cap \left(\bigcap_{k\in K} A_k^c\right)\right) = \prod_{j\in J} P(A_j) \cdot \prod_{k\in K} P(A_k^c)$$

BEWEIS: mit Induktion über $|K|$.

- Induktionsanfang: $K = \emptyset$ ist klar nach Definition der stochastischen Unabhängigkeit.
- Induktionsvoraussetzung: Es gelte für jedes endliche K mit $|K| = n$ und jedes endliche J die Behauptung.
- Induktionsschritt: Sei $K \subseteq I$, $|K| = n + 1$ und $J \subseteq I$ endlich. Sei $k_0 \in K$ und $K' = K \setminus \{k_0\}$. Dann ist

$$
\begin{aligned}
& P\left(\bigcap_{j\in J} A_j \cap \bigcap_{k\in K} A_k^c\right) \\
= \; & P\left(\bigcap_{j\in J} A_j \cap A_{k_0}^c \cap \bigcap_{k\in K'} A_k^c\right) \\
= \; & P\left(\bigcap_{j\in J} A_j \cap \bigcap_{k\in K'} A_k^c\right) - P\left(\left(\bigcap_{j\in J} A_j \cap \bigcap_{k\in K'} A_k^c\right) \cap A_{k_0}\right) \\
= \; & \prod_{j\in J} P(A_j) \cdot \prod_{k\in K'} P(A_k^c) - \prod_{j\in J} P(A_j) P(A_{k_0}) \cdot \prod_{k\in K'} P(A_k^c) \\
= \; & \prod_{j\in J} P(A_j) \underbrace{\prod_{k\in K'} P(A_k^c) \underbrace{(1 - P(A_{k_0}))}_{P(A_{k_0}^c)}}_{\prod_{k\in K} P(A_k^c)}
\end{aligned}
$$

4.2.3 Das Werfen einer Münze

Betrachte das wiederholte Werfen einer Münze, dann ist $\Omega = \{0, 1\}^N$, wobei 0 Kopf und 1 Zahl sei. Als σ-Algebra ist $\mathcal{P}(\mathbb{R})$ nicht möglich für eine sinnvolle Wahrscheinlichkeitsdefinition, da dies „in gewisser Weise“ äquivalent ist zur Einführung des Längenbegriffs in $\mathbb{R}$. Die Definition der „richtigen“ σ-Algebra erfolgt in der folgenden Weise: Zu endlichen $I \subseteq \mathbb{N}$ wähle $z = (z_i)_{i\in I} \in \{0, 1\}^I$ und setze

$$B_z = \left\{\omega \in \{0, 1\}^{\mathbb{N}} \;\middle|\; \omega_i = z_i \;\forall\, i \in I\right\}$$

Als Erzeugendensystem wird nun betrachtet $\mathcal{E} := \left\{ B_z \mid I \subseteq \mathbb{N} \text{ endlich}, z \in \{0, 1\}^I \right\}$. Als Wahrscheinlichkeit gilt nun $P(B_z) = 2^{-|I|}$. Der Satz von Carathéodory liefert, daß genau ein Wahrscheinlichkeitsmaß P auf $\Omega = \{0, 1\}^{\mathbb{N}}$ existiert mit $\mathfrak{A} = \sigma(\mathcal{E})$ mit der Eigenschaft $P(B_z) = 2^{-|I|}$ für endliche $I \subseteq \mathbb{N}$ mit $z \in \{0, 1\}^I$.

Sei nun A_i das Ereignis *im i-ten Wurf Kopf*, dann ist $P(A_i) = \frac{1}{2}$ und

$$P(A_{i_1} \cap \ldots \cap A_{i_n}) = 2^{-n} = \prod_{j=1}^{n} P(A_{i_j})$$

Das Borel-Cantelli-Lemma liefert[7] $P\left(\limsup_n A_n\right) = 1$ (d.h. die Wahrscheinlichkeit für unendlich oft Kopf), da $\sum_{n=1}^{\infty} P(A_n) = \infty$ und $(A_n)_{n\in\mathbb{N}}$ stochastisch unabhängig.

[7] „…Goethes gesammelte Werke - oder noch was viel Schlimmeres, z.B. das Microsoft-Word-2000-Handbuch!"

5 Diskrete Wahrscheinlichkeitsmaße

Wahrscheinlichkeitsmaße werden auch oft als *Wahrscheinlichkeitsverteilungen* oder einfach *Verteilungen* bezeichnet (englisch probability *measure, (probability) distribution*).

DEFINITION: Sei Ω abzählbar, $\mathfrak{A} = \mathcal{P}(\Omega)$, P sei Wahrscheinlichkeitsmaß. Dann heißt $(\Omega, \mathfrak{A}, P)$ *diskreter Wahrscheinlichekeitsraum*, P *diskretes Wahrscheinlichkeitsmaß*.

Für jedes $A \subseteq \Omega$ gilt $A = \sum_{\omega \in A} \{\omega\}$, also $P(A) = \sum_{\omega \in A} P(\{\omega\})$. Sei definiert $p(\omega) = P(\{\omega\})$ für $\omega \in \Omega$, bezeichne $(p(\omega))_\omega$ als *stochastischen Vektor* zu P, damit ist $P(A) = \sum_{\omega \in \Omega} p(\omega)$.

Jedes diskrete Wahrscheinlichkeitsmaß ist eindeutig bestimmt durch den dazugehörigen stochastischen Vektor.

5.1 Laplace-Verteilung

Für eine endliche Menge Ω ist gegeben durch $p(\omega) = \frac{1}{|\Omega|}$, also $P(A) = \frac{|A|}{|\Omega|}$.

5.1.1 Das Geburtstagsproblem

Sei $\Omega = \{1, \ldots, N\}^n$ mit $N = 365$ für das Geburtstagsproblem („Mit welcher Wahrscheinlichkeit haben zwei Personen in einem Raum mit n Personen am gleichen Tag Geburtstag?") und $n = 60$ für die Stochastik-1-Studenten, betrachte vereinfacht eine Laplace-Verteilung. Dann ist für $\omega = (\omega_1, \ldots, \omega_n)$ $p(\omega) = \frac{1}{N^n}$. Betrachte also

$$\begin{aligned} A &= \{\omega \in \Omega \mid \ \exists\, (i,j) \ : (\omega_i = \omega_j) \land (i \neq j)\} \\ A^c &= \{\omega \in \Omega \mid \omega_i \neq \omega_j \ \forall\, i \neq j\} \end{aligned}$$

Damit ist

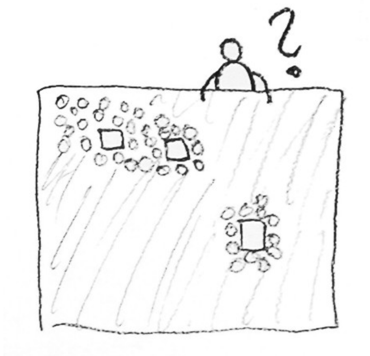
$$\begin{aligned} P(A) &= 1 - P(A^c) = 1 - \frac{N(N-1)\ldots(N-(n-1)}{N^n} \\ &= 1 - \frac{N!}{(N-n)!N^n} = 1 - \frac{365!}{305! \cdot 365^{60}} \end{aligned}$$

Es ergeben sich folgende Werte:

n	5	10	30	60
$P(A)$	2.7%	11.7%	70.6%	99.4%

5.2 Hypergeometrische Verteilung

Qualitätskontrolle:



Sei eine Sendung mit N Stück vorhanden, von der M defekt sind. Man zieht eine Stichprobe von n Stück, wie groß ist nun die Wahrscheinlichkeit, daß k defekte Stücke in der Stichprobe sind. Dies ist offensichtlich 0 für $k > M$ und $n - k > N - M$. Insgesamt gilt[8]:

$$P(A_k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

Die *Hypergeometrische Verteilung* $\mathrm{H}(N, M, n)$ ist gegeben durch $\Omega = \{0, \dots, n\}$ und $P(\omega) = \binom{M}{k} \cdot \binom{N-M}{n-k} \cdot \binom{N}{n}^{-1}$ für $\omega = 0, 1, \dots n$.

5.3 Bernoulli-Verteilung

Es geht um ein Experiment mit den zwei Möglichkeiten Erfolge ($\widehat{=}$ 1) und Mißerfolg ($\widehat{=}$ 0), somit $\Omega = \{0,1\}^n$. Die $\mathrm{Ber}(n,p)$-Verteilung *(Bernoulli-Verteilung)* mit Parametern $n \in \mathbb{N}, p \in [0,1]$ ist gegeben durch

$$p(\omega) = p^{|\{\, i \mid \omega_i = 1 \}|}(1-p)^{|\{\, i \mid \omega_i = 0 \}|} = p^{\sum_{i=1}^n \omega_i} \cdot (1-p)^{n - \sum_{i=1}^n \omega_i}$$

Bei k möglichen Ausgängen ist $\mathrm{Ber}(n, k, p_1, \dots, p_k)$ durch $\Omega = \{1, \dots, k\}^n$ und

$$p(\omega) = \prod_{i=1}^{k} p_i^{|\{\, j \mid \omega_j = i \}|}$$

[8] mit $\binom{k}{j} = 0$ für $j > k$

5.4 Binomialverteilung

Wie groß ist die Wahrscheinlichkeit, daß k Erfolge bei der $\mathrm{Ber}(n,p)$-Verteilung eintreten?

$$(\mathrm{Ber}(n,p))\left(\left\{\omega \;\middle|\; \sum_{i=1}^{n}\omega_i = k\right\}\right) = \binom{n}{k}p^k(1-p)^{n-k}$$

Die *Binomialverteilung* $B(n,p)$ ist gegeben durch $\Omega = \{0,1,\ldots,n\}$ und $p(\omega) = \binom{n}{\omega}p^{\omega}(1-p)^{n-\omega}$ für ω Erfolge bei n-facher Durchführung des Experiments.

5.4.1 Mehrfacher Münzwurf

Betrachte einen 100-fachen Münzwurf, wie groß ist die Wahrscheinlichkeit, daß genau 50 mal Kopf fällt? Es ist $\mathrm{B}(100, 0.5)(50) = \binom{100}{50}\frac{1}{2^{100}}$. Dies liefert die allgemeinere Frage nach einer Approximation für $\mathrm{B}(2n, 0.5)(\{n\})$. Es gilt mit der Stirlingschen Formel $n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$:

$$\begin{aligned}
\mathrm{B}(2n, 0.5)(\{n\}) &= \frac{(2n)!}{2(n!)}\frac{1}{2^{2n}}\\
&\approx \frac{\sqrt{4n\pi}}{2\left(\sqrt{2n\pi}\left(\frac{n}{e}\right)^n\right)}\frac{1}{2^{2n}}\\
&= \frac{1}{\sqrt{\pi n}}
\end{aligned}$$

5.5 Multinomialverteilung

Entsprechend zur Binomialverteilung $\mathrm{B}(n,p)$ ergibt sich die *Multinomialverteilung* $\mathrm{M}(n,k,p_1,\ldots,p_k)$ aus der $\mathrm{Ber}(n,k,p_1,\ldots,p_n)$-Verteilung:

$$\Omega = \left\{\omega \in \{0,\ldots,n\}^k \;\middle|\; \sum_{i=1}^{k}\omega_i = n\right\} \text{ mit } p(\omega) = \binom{n}{\omega_1,\ldots,\omega_k}\prod_{i=1}^{k}p_i^{\omega_i}$$

5.6 Geometrische Verteilung

Startend mit $\{0,1\}^{\mathbb{N}}$ Wiederholung von Münzwürfen. Frage nach der Wahrscheinlichkeit, daß beim k-ten Wurf die 1 zum ersten Mal auftritt: $p(k) = (1-p)^{k-1}p$, die *geometrische Verteilung* $\mathrm{Geo}(p)$ ist gegeben durch $\Omega = \mathbb{N}$ und $p(\omega) = (1-p)^{\omega-1}p$. Die Frage nach der Wahrscheinlichkeit, daß beim k-ten Wurf zum ersten Mal n-mal auftritt: gegeben durch $p(k) = pp^{n-1}(1-p)^{k-1-(n-1)}\binom{n-1}{k-1}$, d.h. die Verteilung $G(n,p)$ ist gegeben durch $\Omega = \mathbb{N}$ und $p(\omega) = \binom{\omega-1}{n-1}p^n(1-p)^{\omega-n}$.

5.7 Poisson-Verteilung

Für $\Omega = \{0, 1, 2, \ldots\}$ ist $p(\omega)$ gegeben durch $e^{-\beta}\frac{\beta^{\omega}}{\omega!}$ für $\omega = 0, 1, 2, \ldots$ mit Parameter β, die ist die *Poisson-Verteilung* $\mathrm{P}(\beta)$. Dies ist eine wichtige Verteilung zur Modellierung von Zählvorgängen: die Anzahl von radioaktiven Zerfallsprozessen in einer Zeiteinheit, die Anzahl eintreffender Kunden in einem Bedienungssystem, die Anzahl von Druckfehler in einem Manuskript. Eine Disziplin der angewandten Wahrscheinlichkeitstheorie ist die *Warteschlangentheorie*:

5.7.1 *Exkurs:* Stirling-Formel

Wir zeigen:

$$n! \sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi} = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

Es ist $\log n! = \log 1 + \ldots + \log n$. Wir benutzen die folgende Abschätzung, die wegen der Konkavität der Logarithmus-Funktion gilt:

$$\int\limits_{k-\frac{1}{2}}^{k+\frac{1}{2}} \log x \, dx \leq \log k \leq \int\limits_{k}^{k+1} \log x \, dx$$

Nun gilt:

$$\int\limits_{\frac{1}{2}}^{n+\frac{1}{2}} \log x \, dx \leq \log n! \leq \int\limits_{1}^{n+1} \log x \, dx$$

Durch Bilden der Stammfunktion $x \log x - x$ erhält man

$$(n + \frac{1}{2}) \log \left(n + \frac{1}{2}\right) - n - \frac{1}{2} \log \left(\frac{1}{2}\right) \leq \log n! \leq (n+1) \log (n+1) - n$$

Es folgt daraus $d_n = \log n! - \left(n + \frac{1}{2}\right) \log n + n \geq -\frac{1}{2} \log \left(\frac{1}{2}\right)$, also ist $(d_n)_{n \in \mathbb{N}}$ nach unten beschränkt. Weiter gilt:

$$\begin{aligned} d_n - d_{n+1} &= \left(n + \frac{1}{2}\right) \log \left(\frac{n+1}{n}\right) - 1 \\ &= \int_n^{n+1} \log x \, dx - \frac{1}{2} \left(\log(n+1) + \log n\right) \geq 0 \end{aligned}$$

Damit ist die Folge $(d_n)_{n \in \mathbb{N}}$ monoton fallend, d.h. sie besitzt einen Grenzwert $\tau = \lim_{n \to \infty} d_n$, also

$$e^{\tau} = \lim_{n \to \infty} e^{\log n! - \left(n + \frac{1}{2}\right) \log n + n} = \lim_{n \to \infty} \frac{n!}{n^{n+\frac{1}{2}} e^{-n}}$$

Wir haben also gezeigt: Es exitiert γ mit

$$n! \sim n^{n+\frac{1}{2}} e^{-n} \gamma$$

Schwieriger ist zu zeigen, daß $\gamma = \sqrt{2\pi}$ ist, der Beweis erfolgt hier nicht.

6 Wahrscheinlichkeitsmaße auf $\mathbb{R}$

Betrachtet wird in diesem Abschnitt stets $\mathbb{R}$ mit σ-Algebra $\mathcal{B}$ der Borelschen Mengen. Dabei ist $\mathcal{B} = \sigma(\mathcal{E})$, wobei $\mathcal{E} = \{\,(a,b] \mid a < b\}$ ist.

6.1 Verteilungsfunktion

DEFINITION: Sei P ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$, d.h. $P\colon \mathcal{B} \to [0,1]$. Zu P wird definiert

$$F_P\colon \mathbb{R} \to [0,1] \text{ mit } t \mapsto P((-\infty,t])\ \forall\, t \in \mathbb{R}$$

Dann heißt F_P *Verteilungsfunktion zu P*.

Die Verteilungsfunktion F_P hat folgende Eigenschaften:

- sie ist monoton wachsend, denn für $t_1 \leq t_2$ gilt $P((\infty,t_1]) \leq P((\infty,t_2])$
- sie ist rechtsseitig stetig, denn für $t_n \downarrow t$ gilt
  $$F_P(t) = P((-\infty,t]) = P\left(\bigcap_n (-\infty,t_n]\right) = \lim_{n\to\infty} P((-\infty,t_n]) = \lim_{n\to\infty} F_P(t_n)$$
- $\lim_{t\to-\infty} F_P(t) = 0$, da für $t_n \downarrow -\infty$ gilt:
  $$0 = P(\emptyset) = P\left(\bigcap_n (-\infty,t_n]\right) = \lim_{n\to-\infty} F_P(t_n)$$
- analog $\lim_{t\to+\infty} F_P(t) = 1$, da für $t_n \uparrow +\infty$ gilt:
  $$1 = P(\mathbb{R}) = \lim_{n\to\infty} P\left((-\infty,t_n]\right) = \lim_{n\to\infty} F_P(t_n)$$

DEFINITION: Eine Funktion $F : \mathbb{R} \to [0,1]$ mit den Eigenschaften

1. F ist monoton wachsend,
2. F ist rechtsseitig stetig,
3. es gilt $\lim_{t\to-\infty} F_P(t) = 0$ und $\lim_{t\to+\infty} F_P(t) = 1$

heißt *Verteilungsfunktion*.

SATZ: Sei F Verteilungsfunktion. Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $\mathbb{R}$ so, daß $F = F_P$ gilt, d.h. daß P die Verteilungsfunktion F besitzt, d.h. $F(t) = P((-\infty,t])$ für alle $t \in \mathbb{R}$.

Der Beweis erfolgt in mehreren Schritten:

6.1.1 Beweis der Eindeutigkeit

Beginne mit der *Eindeutigkeitsaussage:*

SATZ: Seien P, Q Wahrscheinlichkeitsmaße auf $\mathbb{R}$ mit $P((-\infty, t]) = Q((-\infty, t])$ für alle $t \in \mathbb{R}$. Dann gilt $P = Q$, d.h. $P(B) = Q(B)$ für alle $B \in \mathcal{B}$.

BEWEIS: Definiere $\mathcal{C} = \{B \in \mathcal{B} \mid P(B) = Q(B)\}$. Zu zeigen ist: $\mathcal{C} = \mathcal{B}$. Ausgehend von $\mathcal{B} = \sigma(\mathcal{E})$ (d.h. die kleinste σ-Algebra, die $\mathcal{E}$ umfaßt) genügt zu zeigen:

($\star$) $\mathcal{E} \subseteq \mathcal{C}$

($\star\star$) $\mathcal{C}$ ist σ-Algebra.

Es gilt:

($\star$) Es gilt $(a, b] = (-\infty, b] \setminus (-\infty, a]$. Also gilt:

$$\begin{aligned} P((a,b]) &= P((-\infty, b]) - P((-\infty, a]) \\ &= Q((-\infty, b]) - Q((-\infty, a]) \\ &= Q((a,b]) \end{aligned}$$

($\star\star$) Es gelten alle Eigenschaften der σ-Algebra:

1. Es sind $\emptyset, \Omega \in \mathcal{C}$, da $P(\emptyset) = 0 = Q(\emptyset)$ und $P(\mathbb{R}) = 1 = Q(\mathbb{R})$ ist.
2. Mit $A \in \mathcal{C}$ ist auch $A^C \in \mathcal{C}$, da mit $P(A) = Q(A)$ ist auch $P(A^c) = 1 - P(A) = 1 - Q(A) = Q(A^c)$.
3. Für $(A_n)_{n\in\mathbb{N}}$ mit $A_n \in \mathcal{C}$ ist auch $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{C}$, denn:

$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} Q(A_n) = Q\left(\bigcup_{n\in\mathbb{N}} A_n\right)$$

wobei o.B.d.A. $(A_n)_{n\in\mathbb{N}}$ als wachsend angenommen wird.

Ganz so einfach geht's aber nicht: Sei $(A_n)_{n\in\mathbb{N}}$ gegeben mit $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n$ mit $B_n = \bigcup_{k\le n} A$ monoton wachsend. Nun ist aber $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. Aber was ist hier $Q(A_1 \cup A_2) = Q(A_1) + Q(A_2) - Q(A_1 \cap A_2)$? Also kann man das „o.B.d.A.“ nicht einfach so verwenden.

Diese Eigenschaft läßt sich nicht direkt nachweisen, die Mathematik mußte sich etwas einfallen lassen!

6.1.2 Dynkin-Systeme

Die Dynkin-Systeme entstanden zur Lösung der Schwierigkeiten in eben gezeigtem Beweis.

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

- $\Omega \in \mathcal{D}$
- $A, B \in \mathcal{D} \Rightarrow B \setminus A \in \mathcal{D}$
- Für $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkt mit $A_n \in \mathcal{D}$, so ist $\sum_{n\in\mathbb{N}} A_n \in \mathcal{D}$

Bemerkungen:

- Jede σ-Algebra ist ein Dynkin-System
- Definiere
$$\delta(\mathcal{E}) := \bigcap_{\substack{\mathcal{E}\subset\mathcal{D} \\ \mathcal{D}\text{ Dynkin}-\text{System}}} \mathcal{D}$$
Dann ist $\delta(\mathcal{E})$ ein Dynkin-System mit $\delta(\mathcal{E}) \subseteq \sigma(\mathcal{E})$.

Nun gilt im obigen Beweis:

($\star\star$) $\mathcal{C}$ ist Dynkin-System, denn

1. Es sind $\emptyset, \Omega \in \mathcal{C}$, da $P(\emptyset) = 0 = Q(\emptyset)$ und $P(\mathbb{R}) = 1 = Q(\mathbb{R})$ ist.
2. Mit $A, B \in \mathcal{C}$ und $A \subseteq B$ ist $B \setminus A \in \mathcal{D}$ wegen $P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$.
3. Für $(A_n)_{n\in\mathbb{N}}$ mit paarweise disjunkten $A_n \in \mathcal{C}$ gilt
$$P\left(\sum_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n) = \sum_{n\in\mathbb{N}} Q(A_n) = Q\left(\sum_{n\in\mathbb{N}} A_n\right)$$

Daraus folgt $\delta(\mathcal{E}) \subseteq \mathcal{C}$.

Damit ist obiger Satz gezeigt für $\delta(\mathcal{E}) = \sigma(\mathcal{E})$. Beachte dazu folgenden Satz:

Satz: Die Menge $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ sei $\cap$-stabil (d.h. mit $A, B \in \mathcal{E}$ sei auch $A \cap B \in \mathcal{E}$). Dann gilt $\delta(\mathcal{E}) = \sigma(\mathcal{E})$.

Beweis:

- Man beachte zunächst: Falls $\mathcal{D}$ ein $\cap$-stabiles Dynkin-System ist, so ist $\mathcal{D}$ schon σ-Algebra. Es gilt für $D_1, D_2 \in \mathcal{D}$, daß auch D_1^c und $D_2^c \in \mathcal{D}$ sind, mit der $\cap$-Stabilität gilt $D_1^c \cap D_2^c \in \mathcal{D}$, also $D_1 \cup D_2 = (D_1^c \cap D_2^c)^c \in \mathcal{D}$. Per Induktion sind also alle endlichen Vereinigungen von Elementen aus $\mathcal{D}$ wieder in $\mathcal{D}$. Für abzählbare $D_1, D_2, \ldots \in \mathcal{D}$ ist

$$\bigcup_{n\in\mathbb{N}} D_n = \sum_{n=1}^{\infty} \left(\left(\bigcup_{j=1}^{n} D_j \right) \setminus \left(\bigcup_{j=1}^{n-1} D_j \right) \right) \in \mathcal{D}$$

- Zeige nun: $\delta(\mathcal{E})$ ist $\cap$-stabil und somit wie eben gezeigt auch σ-Algebra. Sei $E \in \mathcal{E}$. Definiere nun

$$\mathcal{F}_E = \{ D \in \delta(\mathcal{E}) \mid D \cap E \in \delta(\mathcal{E}) \}$$

Es gilt:

1. $\mathcal{E} \subseteq \mathcal{F}_E$, da $\mathcal{E}$ $\cap$-stabil ist (d.h. für alle $E' \in \mathcal{E}$ ist $E' \cap E \in \delta(\mathcal{E})$, also in $\mathcal{F}_E$)
2. $\mathcal{F}_E$ ist Dynkin-System[9]:
   - $\emptyset$ und Ω liegen in $\mathcal{F}_E$
   - mit D_1, $D_2 \in \mathcal{F}_E$ liegen auch $(D_2 \setminus D_1) \cap E = (D_2 \cap E) \setminus (D_1 \cap E) \in \delta(\mathcal{E})$, also $D_2 \setminus D_1 \in \mathcal{F}_E$,
   - und für $D_1, D_2, \ldots$ paarweise disjunkt aus $\mathcal{F}_E$ ist $\left(\sum_{i\in\mathbb{N}} D_i\right) \cap E = \sum_{i\in\mathbb{N}} (D_i \cap E) \in \delta(\mathcal{E})$, also $\sum_{i\in\mathbb{N}} D_i \in \mathcal{F}_E$.

Dies liefert $\delta(\mathcal{E}) \subseteq \mathcal{F}_E$, d.h. $E \cap D \in \delta(\mathcal{E})$ für alle $E \in \mathcal{E}$ und $D \in \delta(\mathcal{E})$. Sei nun $D \in \delta(\mathcal{E})$, definiere analog

$$\mathcal{F}_D = \{ D' \in \delta(\mathcal{E}) \mid D' \cap D \in \delta(\mathcal{E}) \}$$

Dann gilt wieder:

1. $\mathcal{E} \subseteq \mathcal{F}_D$, denn gemäß vorherigem gilt $E \cap D \in \delta(\mathcal{E})$ für alle $E \in \mathcal{E}$.
2. $\mathcal{F}$ ist Dynkin-System (wie eben)

Somit wieder $\delta(\mathcal{E}) \subseteq \mathcal{F}_D$, d.h. $D \cap D' \in \delta(\mathcal{E})$ für alle $D, D' \in \delta(\mathcal{E})$.

- Schon gezeigt: $\delta(\mathcal{E}) \subseteq \sigma(\mathcal{E})$, wie eben gezeigt gilt $\sigma(\mathcal{E}) \subseteq \delta(\mathcal{E})$, insgesamt also $\delta(\mathcal{E}) = \sigma(\mathcal{E})$.

[9] „Das ist so trivial wie es trivialer nicht gehen kann!“

6.1.3 Beweis der Eindeutigkeit (korrekt)

Beweis von (6.1.1): Sei $\mathcal{C} = \{B \in \mathcal{B} \mid P(B) = Q(B)\}$. Zu zeigen ist: $\mathcal{C} = \mathcal{B}$. Ausgehend von $\mathcal{B} = \sigma(\mathcal{E})$ (d.h. die kleinste σ-Algebra, die $\mathcal{E}$ umfaßt) genügt zu zeigen:

($\star$) $\mathcal{E} \subseteq \mathcal{C}$

($\star\star$) $\mathcal{C}$ ist σ-Algebra.

Es gilt:

($\star$) Es gilt $(a, b] = (-\infty, b] \setminus (-\infty, a]$. Also gilt:

$$\begin{aligned} P((a,b]) &= P((-\infty, b]) - P((-\infty, a]) \\ &= Q((-\infty, b]) - Q((-\infty, a]) \\ &= Q((a,b]) \end{aligned}$$

($\star\star$) $\mathcal{C}$ ist Dynkin-System, denn:

1. Es sind $\emptyset, \Omega \in \mathcal{C}$, da $P(\emptyset) = 0 = Q(\emptyset)$ und $P(\mathbb{R}) = 1 = Q(\mathbb{R})$ ist.
2. Für $A, B \in \mathcal{C}$ mit $A \subseteq B$ ist $P(B \setminus A) = P(B) - P(A) = Q(B) - Q(A) = Q(B \setminus A)$.
3. Für paarweise disjunkte $A_1, A_2, \ldots \in \mathcal{C}$ ist

$$P\left(\sum_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i) = \sum_{i \in \mathbb{N}} Q(A_i) = Q\left(\sum_{i \in \mathbb{N}} A_i\right)$$

Wende nun den Satz von Dynkin an, beachte dabei, daß $\mathcal{E}$ $\cap$-stabil ist, und erhalte damit $\mathcal{C} = \sigma(\mathcal{E}) = \mathcal{B}$, d.h. $P(B) = Q(B)$ für alle $B \in \mathcal{B}$.

6.1.4 Bierdeckelexperiment

Betrachte ein Experiment, bei dem auf zwei Bierdeckel zwei reelle Zahlen geschrieben werden. Einer der Bierdeckel wird aufgedeckt, zu raten ist, ob die andere Zahl größer oder kleiner ist. Das reine Raten liefert Gewinnwahrscheinlichkeit $\frac{1}{2}$, theoretisch ist jedoch folgende Gewinnwahrscheinlichkeit möglich (mit P eine normalverteilte Zufallsgröße):

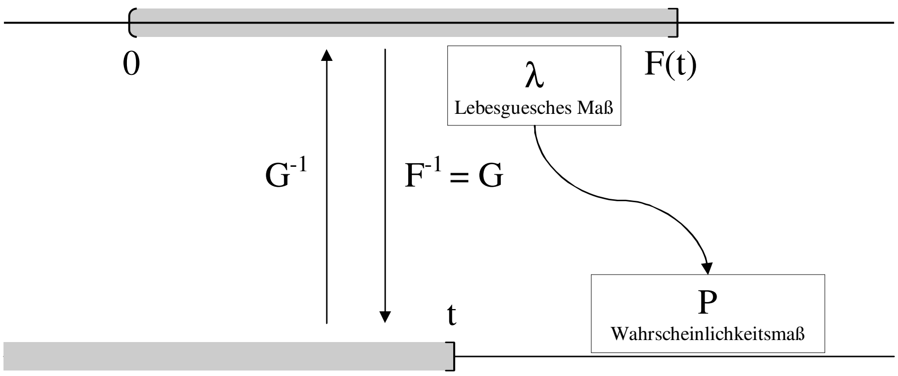
$$\begin{aligned} P(\text{richtig}) &= P(\text{richtig}|k \text{ aufgedeckt}) + P(\text{richtig}|g \text{ aufgedeckt}) \\ &= \frac{1}{2}P([k, \infty)) + \frac{1}{2}P((-\infty, g]) \\ &= \frac{1}{2} + \frac{1}{2}P((k, g]) > \frac{1}{2} \end{aligned}$$

Strategie ist also: Nachdem ein Bierdeckel mit der Zahl x umgedreht wurde, zückt man sein Noteboook ;o), lässt diesen den Wert y einer normalverteilten Zufallsgröße Y erzeugen, vergleicht x und y und entscheidet sich für den aufgedeckten Deckel, falls $x > y$ ist und andernfalls für den noch zugedeckten.

6.1.5 Beweis der Existenzaussage

Satz: Sei F Verteilungsfunktion. Dann existiert ein Wahrscheinlichkeitsmaß P auf $\mathbb{R}$ mit $F(t) = P((-\infty, t])$ für alle $t \in \mathbb{R}$.

Beweis: Die Existenz ergibt sich mit dem Satz von Carathéodory. *Alternative Möglichkeit:* Benutzung der verallgemeinerten Inversen F^{-1} und des Lebesgueschen Maßes λ. Dies wird wie folgt durchgeführt:



Definition: Zu $F\colon \mathbb{R} \to [0,1]$ definiere die *verallgemeinerte Inverse*

$$F^{-1}\colon (0,1) \to \mathbb{R} \text{ mit } F^{-1}(x) = \inf\{t \mid F(t) \geq x\}$$

Es gilt: $F(t) \geq x \Leftrightarrow t \geq F^{-1}(x)$. Mit $G = F^{-1}$ ist für alle $i \in \mathbb{R}$:

$$G^{-1}((-\infty, t]) = \{x \mid G(x) \leq t\} = \{x \mid x \leq F(t)\} = (0, F(t)]$$

Für $B \in \mathcal{B}$ definiere $P(B) = \lambda(G^{-1}(B))$. Insbesondere gilt für $B = (-\infty, t]$:

$$P((-\infty, t]) = \lambda(G^{-1}((-\infty, t])) = \lambda((0, F(t)]) = F(t)$$

Fragen:

1. Ist diese Definition sinnvoll? Darf $\lambda(G^{-1}(B))$ gebildet werden, d.h. ist $G^{-1}(B) \in \mathcal{B}$?

2. Ist P ein Wahrscheinlichkeitsmaß?

Falls diese Fragen mit *ja!* beantwortet werden können, so ist diese Existenzaussage bewiesen.

1. Ja! Definiere wieder ein System $\mathcal{C}$, welches alle $B \in \mathcal{B}$ umfaßt, für die $G^{-1}(B) \in \mathcal{B}$ gebildet werden kann, es ist wieder $\mathcal{E} \subseteq \mathcal{C}$ und ist eine σ-Algebra, also gleich $\sigma(\mathcal{E}) = \mathcal{B}$.

2. Ja! Folgt aus den Eigenschaften der Inversen G^{-1}.

6.2 Berechnung von Wahrscheinlichkeiten unter Benutzung von Verteilungsfunktionen

DEFINITION: Es sei

$$f(x+) := \lim_{s \downarrow x} f(s) \quad f(x-) := \lim_{s \uparrow x} f(s) \quad f(\infty-) := \lim_{s \to \infty} f(s)$$

LEMMA: Es gilt für eine Verteilungsfunktion F_P zum Wahrscheinlichkeitsmaß P:

1. $P((a,b]) = F_p(b) - F_p(a)$
2. $P((a,b)) = F_p(b-) - F_p(a)$
3. $P([a,b]) = F_p(b) - F_p(a-)$
4. $P([a,b)) = F_p(b-) - F_p(a-)$
5. F ist stetig in t genau dann, wenn $P(\{t\}) = 0$ ist.

BEWEIS: Zunächst ist

$$P((-\infty,t)) = P\left(\bigcup_{n\in\mathbb{N}}\left(-\infty, t-\frac{1}{n}\right]\right) = \lim_{n\to\infty} P\left(\left(-\infty, t-\frac{1}{n}\right]\right)$$

2. $P((a,b)) = P((-\infty,b)) - P((-\infty,a]) = F_P(b-) - F_P(a-)$, die nächsten beiden entsprechend

5. Es ist $P(\{t\}) = P([t,t]) = F_P(t) - F_P(t-)$, d.h. bei $P(\{t\}) = 0$ ist die Funktion im Punkt t linksseitig stetig und somit stetig.

DEFINITION: Ein Wahrscheinlichkeitsmaß heißt *stetig*, falls $P(\{t\}) = 0$ für alle $t \in \mathbb{R}$.

Die am häufigsten benutzten Wahrscheinlichkeitsmaße auf $\mathbb{R}$ sind folgenden Typs:

6.3 Wahrscheinlichkeitsmaße mit stetigen Dichten

Seien $-\infty \leq a < b \leq \infty$, $f\colon \mathbb{R} \to [0, \infty)$ mit $f|_{(a,b)}$ stetig, $f|_{(a,b)^c} = 0$ und $\int_{-\infty}^{\infty} f(x)\,dx = \int_a^b f(x)\,dx = 1$. Sei definiert:

$$F\colon \mathbb{R} \to [0,1] \text{ mit } t \mapsto \int\limits_{-\infty}^{t} f(x)\,dx$$

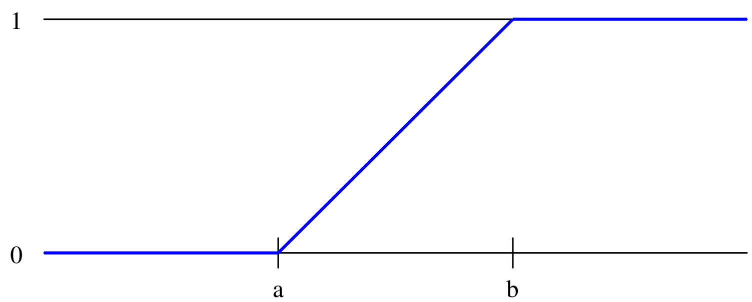
Damit ist F Verteilungsfunktion. Also existiert genau ein Wahrscheinlichkeitsmaß P mit dieser Verteilungsfunktion. P heißt dann *Wahrscheinlichkeitsmaß mit der Dichte f.*

6.3.1 Rechtecksverteilung

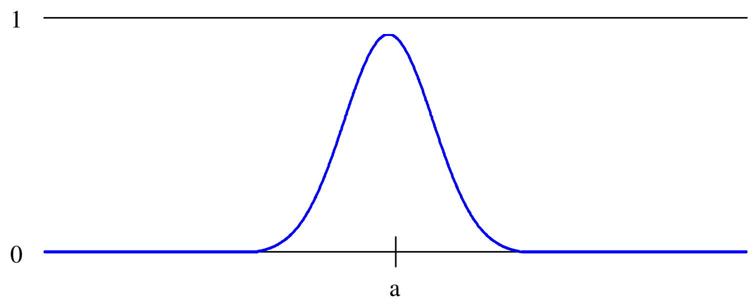
DEFINITION: Die *Indikatorfunktion* sei gegeben durch

$$1_A(t) := \begin{cases} 0 & \text{falls} \quad t \in A \\ 1 & \text{falls} \quad t \notin A \end{cases}$$

Die *Rechteckverteilung* $R(a,b)$ ist gegeben durch die Dichte $f(t) = \frac{1}{b-a} 1_{(a,b)}(t)$



6.3.2 Normalverteilung



Die *Normalverteilung (Gaußverteilung)* ist gegeben durch die Dichte

$$N(a,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Sie wird häufig angewandt zur Beschreibung von Meßfehlern und Erträgen, in der Ökonometrie etc., es ist *die* Verteilung der Statistik!

Nachzuweisen ist noch $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (und gehe zur neuen Variable $\frac{x}{\sigma}$ über in zweiten Schritt):

$$\begin{aligned} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}}\,dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx \end{aligned}$$

Benötigt wird $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \frac{1}{\sqrt{2\pi}}$. Zum Nachweis betrachte (Trick! mit Polarkoordinaten $dxdy = rdrd\vartheta$)

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx\right) \cdot \left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,dy\right) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\,dxdy \\ &= \int_0^{2\pi} \underbrace{\int_0^{\infty} rdrd\vartheta}_{=1} \\ &= 2\pi \end{aligned}$$

6.3.3 Gammaverteilung

Definiere die Gammafunktion

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x}\,dx$$

Für $\alpha, \beta > 0$ sei

$$f(x) = \begin{cases} \beta^\alpha x^{\alpha-1} e^{-\beta x} \frac{1}{\Gamma(\alpha)} & \text{falls} \quad x > 0 \\ 0 & \text{falls} \quad x \leq 0 \end{cases}$$

Für $\alpha = 1$ ist $f(x) = \beta e^{-\beta x}$ die *Exponentialverteilung* $\mathrm{Exp}(\beta)$. Sie wird benutzt zur Modellierung von Wartezeiten, Bedienzeiten und radioaktiven Zerfällen. Die Exponentialverteilung ist charakterisiert durch die *Gedächtnislosigkeit*.

Mathematische Formulierung der Gedächtnislosigkeit zur Modellierung von Wartezeiten, Lebensdauern:

$$\frac{P((s+t,\infty))}{P((s,\infty))} = P((s+t,\infty)|(s,\infty)) = P((t,\infty))$$

Die Wahrscheinlichkeit dafür, noch t Zeiteinheiten zu warten, ist unabhängig davon, ob man schon s Zeiteinheiten gewartet hat.

1. Falls P Exponentialverteilung $\text{Exp}(\beta)$ ist, so ist P gedächtnislos, denn:

   $$\text{Exp}(\beta)((s,\infty)) = \int_s^\infty \beta e^{-\beta x}\, dx = e^{-\beta s}$$

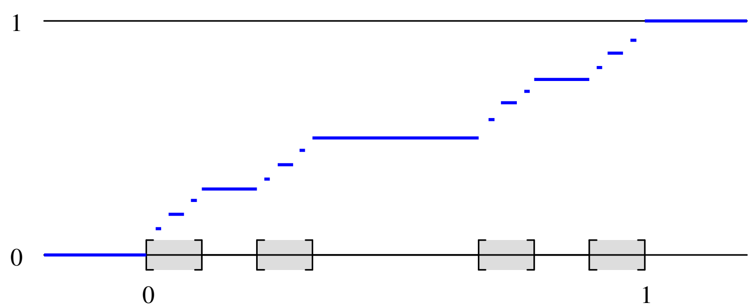
   Die Behauptung folgt durch Einsetzen.

2. Aus P gedächtnislos und $P((0,\infty)) = 1$ folgt P Exponentialverteilung, denn: Setze $G(t) = P((t,\infty))$, Gedächtnislosigkeit besagt $G(s+t) = G(s)G(t)$ für alle $s,t > 0$, diese Funktionalgleichung für G liefert als Lösung $G(s) = e^{\gamma s}$ mit $\gamma < 0$, da $G(t) \to 0$ für $t \to \infty$. Es folgt: $P((-\infty,s]) = 1 - G(s) = 1 - e^{-\beta x}$ mit $\beta = -\gamma > 0$. Eindeutigkeitssatz besagt: P ist Exponentialverteilung.

6.3.4 eine ungewöhnliche Verteilungsfunktion

Definiere eine Verteilungsfunktion mit Hilfe der Cantormenge C:

- induktiver Vorgang liefert Funktion auf Komplement der Cantormenge C:



- stetige Fortsetzung auf C liefert $F\colon \mathbb{R} \to [0,1]$ mit $F(t) = 0$ für $t \le 0$ und $F(t) = 1$ für $t \ge 1$, wobei F monoton wachsend und stetig.

Die Cantormenge hat die Eigenschaft $\lambda(C) = 0$. Zur Cantorfunktion gehört eindeutig ein Wahrscheinlichkeitsmaß P.

Frage: Wie groß ist $P(C)$? Äquivalent: Wie groß ist $P(C^c)$? Für jedes Intervall $I = [a, b]$ ist die Wahrscheinlichkeit $P(I) = 0$, da die Verteilungsfunktion $F = 0$ ist auf den herausgenommenen Intervallen. Somit ist $P(C^c) = 0$, also $P(C) = 1$.

DEFINITION: Seien P, Q Wahrscheinlichkeitsmaße

- P heißt *singulär* zu Q (in Zeichen $P \perp Q$), falls gilt: Es existiert $A \in \mathfrak{A}$ mit $P(A) = 0$ und $Q(A) = 1$.
- P heißt *absolut stetig* bezüglich Q (in Zeichen $P \ll Q$), falls für alle $A \in \mathfrak{A}$ gilt: $Q(A) = 0 \Rightarrow P(A) = 0$.

Bezüglich dieser Definition gilt: $P_{\text{Cantor}} \perp \lambda$ betrachtet als Wahrscheinlichkeitsmaße auf $(0, 1)$, zudem gilt z.B. $B(n, p) \perp R(0, n)$. Für Maße P, P' läßt sich nachweisen: $P = Q + Q'$ mit $Q \perp P'$ und $Q' \ll P'$.

7 Zufallsvariablen

In diesem Abschnitt werden die in der Wahrscheinlichkeitstheorie und Statistik stets benutzten Bezeichnungsweisen eingeführt.

7.1 Bezeichnungsweisen

Die Symbole X, Y, Z bezeichnen in der Regel Abbildungen, d.h. $X : \Omega \to \mathcal{X}$, weiter ist $X^{-1} : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\Omega)$ mit $X^{-1}(A) = \{\omega \mid X(\omega) \in A\}$. Weitere Bezeichnungen:

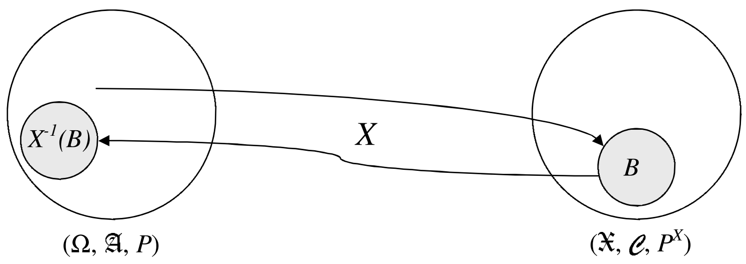
$$\begin{array}{rll} \{X \in A\} & \text{steht für} & \{\omega \mid X(\omega) \in A\} = X^{-1}(A) \\ \{X \leq t\} & \text{steht für} & \{\omega \mid X(\omega) \leq t\} \\ P(X \in A) & \text{steht für} & P(X^{-1}(A)) = P(\{\omega \mid X(\omega) \in A\}) \end{array}$$

Definition: $(\Omega, \mathcal{A}), (\mathcal{X}, \mathcal{C})$ seien messbare Räume. Eine Abbildung $X : \Omega \to \mathcal{X}$ heißt *messbar*, falls gilt: $X^{-1}(A) \in \mathcal{A}$ für alle $A \in \mathcal{C}$. Falls $(\Omega, \mathcal{A}, P)$ einen Wahrscheinlichkeitsraum bildet, so wird eine messbare Abbildung $X : \Omega \to \mathcal{X}$ als *Zufallsvariable* bezeichnet (*random variable*). Im Falle $\mathcal{X} = \mathbb{R}$ sprechen wir von einer *Zufallsgröße*.

Definition: $X : \Omega \to \mathcal{X}$ sei eine Zufallsvariable. Dann ist die Abbildung $P^X : \mathcal{C} \to [0, 1]$ mit $P^X(A) = P(X \in A) = P(X^{-1}(A))$ ein Wahrscheinlichkeitsmaß und wird als *Verteilung von X* bezeichnet. Ist X Zufallsgröße, also P^X Wahrscheinlichkeitsmaß auf $\mathbb{R}$, so wird die zugehörige Verteilungsfunktion $F_X = F_{P^X}$ als *Verteilungsfunktion von X* bezeichnet; entsprechend f_X als *Dichte* von X bei Vorliegen einer solchen.

Beweis: Wir zeigen, dass P^X ein Maß ist

$$\begin{aligned} P^X(\emptyset) &= P(X^{-1}(\emptyset)) = P(\emptyset) = 0 \\ P^X(\mathcal{X}) &= P(X^{-1}(\mathcal{X})) = p(\Omega) = 1 \\ P^X\left(\sum_i A_i\right) &= P(X^{-1}(\sum_i A_i)) \\ &= P(\sum_i X^{-1}(A_i)) = \sum_i P(X^{-1}(A_i)) = \sum_i P^X(A_i) \end{aligned}$$



7.2 Modellierung

Zur Modellierung von zufälligen Vorgängen werden üblicherweise Zufallsvariablen herangezogen, wobei *nur deren Verteilung* spezifiziert wird, jedoch *nicht* der zugrundeliegende Wahrscheinlichkeitsraum.

Beispiel: Wir spielen Roulette und registrieren bei n Spielen die Anzahl des auftretenden „Rot". Dann ist $\Omega = \{ (\omega_1, \dots, \omega_n) \mid \omega_i \in \{0,1\}\}$, Schwarz entspricht 0 und Rot 1. Weiter ist

$$p(\omega) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

Sei $X : \Omega \to \mathcal{X} = \{0,1,\dots,n\}$ mit $X(\omega) = \sum_{i=1}^n \omega_i$. P^X ist ein Wahrscheinlichkeitsmaß auf $\{0,1,\dots,n\}$ gegeben durch

$$P(X=k) = P^X(\{k\}) = P(\{\omega \mid \omega_1 + \ldots + \omega_n = k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

Verteilung von X ist also die Binomialverteilung mit Parametern (n,p), also $P^X = B(n,p)$.

7.3 Eigenschaften von meßbaren Abbildungen

Im Folgenden betrachten wir einige einfachen Tatsachen über Meßbarkeit.

(i) Sind $X : \Omega \to \mathcal{X}$ und $Y : \mathcal{X} \to \mathcal{Y}$ meßbar, so ist $Y \circ X$ meßbar.

(ii) Sei $\mathcal{C} = \sigma(\mathcal{F})$ für ein Erzeugendensystem $\mathcal{F}$. Für $X : \Omega \to \mathcal{X}$ gelte $X^{-1}(F) \in \mathcal{A}$ für alle $F \in \mathcal{F}$. Dann gilt: X ist messbar. Setze $\mathcal{C}' = \{ A \in \mathcal{C} \mid X^{-1}(A) \in \mathcal{A}\}$. Nach Voraussetzung gilt: $\mathcal{C}' \subseteq \mathcal{C}$. Weiter gilt: $\mathcal{C}'$ ist eine σ-Algebra. Daraus folgt: $\mathcal{C}' \supseteq \sigma(\mathfrak{F}) = \mathcal{C}$

Sei nun $X = (X_1, \dots, X_n) : \mathbb{R} \to \mathbb{R}^n$.

(iii) Es gelte $\{\omega \mid X(\omega) \le t\} \in \mathcal{A}$ für alle $t \in \mathbb{R}^n$. Dann ist X messbar. Dies folgt mit (ii) und $\mathcal{F} = \{ ((-\infty, \dots, -\infty), t] \mid t \in \mathbb{R}^n\}$

(iv) X ist genau dann messbar, wenn X_i messbar ist für $i = 1, \ldots, n$. *Beweis:*

„$\Rightarrow$“ $X_i^{-1}(B) = X^{-1}(\mathbb{R} \times \ldots \times \mathbb{R} \times B \times \mathbb{R} \times \ldots \times \mathbb{R})$.

„$\Leftarrow$“ Wende (ii) an, beachte dabei:

$$\begin{aligned} \{\omega \mid X(\omega) \leq t\} &= \{\omega \mid X_i(\omega) \leq t_i \text{ für } i = 1, \ldots, n\} \\ &= \bigcap_{i=1}^{n} \{\omega \mid X_i(\omega) \leq t_i\} \end{aligned}$$

(v) Ist $f : \mathbb{R}^k \to \mathbb{R}^n$ stetig, so ist f messbar. Wende dazu (ii) an mit $\mathcal{F} = \big\{ O \mid O \subseteq \mathbb{R}^N \text{ offen } \big\}$.

(vi) Sind $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ messbar, so sind es auch $X_1 + \ldots + X_n$ und $\min\{X_1, \ldots, X_n\}$.

Beweis: $X = (X_1, \ldots, X_n)$ ist laut (iii) messbar. Betrachte nun

$$\Omega \xrightarrow{X} \mathbb{R}^n \xrightarrow[]{\substack{+\ldots+, \\ \min\{\ldots\}}} \mathbb{R}$$

Die Funktionen $+ \ldots +$ und min sind stetig, somit ist die gesamte Abbildung messbar.

(vii) Sei f stetig, $B \subseteq \mathbb{R}^n$ offen, dann ist $f^{-1}(B)$ offen, also $\in \mathcal{B}^k$.

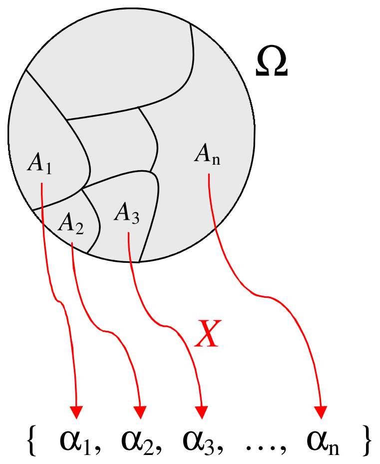
„Zusammenfassung“ der Eigenschaften: sämtliche Abbildungen in Stochastik I und Stochastik II sind meßbar, d.h. Zufallsvariablen.

7.4 Meßbare Abbildungen mit endlichem Wertebereich

Zu $A \subseteq \Omega$ ist die *Indikatorfunktion* definiert als

$$1_A : \Omega \to \mathbb{R} \text{ mit } 1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Es gilt 1_A ist genau dann meßbar, wenn $A \in \mathcal{A}$. Seien nun $A_i = \{\omega \mid X(\omega) = \alpha_i\}$, dann gilt $X = \sum_{i=1}^{n} \alpha_i 1_{A_i}$ ist meßbar, falls $A_1, \ldots, A_n \in \mathcal{A}$ (siehe oben Eigenschaft (vi)). Eine Zufallsvariable mit endlichem Wertebereich besitzt also die Darstellung $X = \sum_{i=1}^{n} \alpha_i 1_{A_i}$ mit meßbaren, paarweise disjunkten $A_i \in \mathcal{A}$.

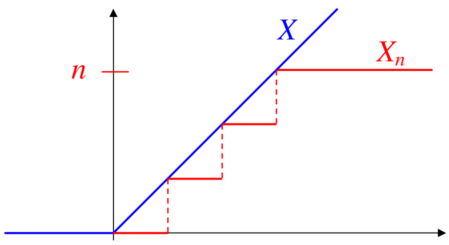


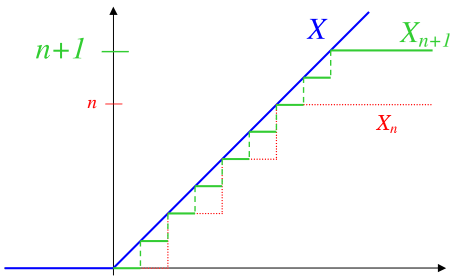
Strukturaussage über meßbare reellwertige Abbildungen ≥ 0: Diese sind darstellbar als Limites von monoton wachsenden Folgen von meßbaren Abbildungen mit endlichem Wertebereich:

SATZ: Sei X eine meßbare Abbildung, $X : \Omega \to [0, \infty)$. Dann existiert eine Folge von meßbaren Abbildungen $X_n : \Omega \to [0, \infty)$ mit endlichem Wertebereich, so daß gilt $0 \leq X_1 \leq X_2 \leq \ldots$ und $\lim_{n\to\infty} X_n = X$ (punktweise Konvergenz).

BEWEIS: Definiere

$$X_n(\omega) = \begin{cases} n & \text{falls} \qquad\qquad X(\omega) \geq n \\ \frac{i}{2^n} & \text{falls} \quad \frac{i}{2^n} \leq X(\omega) < \frac{i+1}{2^n} \quad i = 0, 1, \ldots, n2^n - 1 \end{cases}$$





D.h. es ist

$$X_n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} 1_{\{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\}} + n 1_{\{X \geq n\}}$$

Dann gilt:

1. $X_1, X_2, \ldots$ sind meßbar mit endlichen Wertebereich.
2. Monotonie: $X_n(\omega) \leq X_{n+1}(\omega)$ für alle n, ω.
3. Für $n \geq X(\omega)$ gilt: $|X_n(\omega) - X(\omega)| \leq \frac{1}{2^n}$, also $X_n(\omega) \to X(\omega)$.

Sind $X_1, X_2, \ldots$ meßbar reellwertig, so sind auch $\sup_n X_n$, $\inf_n X_n$, $\limsup_n X_n$, $\liminf_n X_n$ meßbar, denn:[10]

1. $\{\sup_n X_n \leq t\} = \bigcap \{X_n \leq t\}$
2. $\inf X_n = -\sup_n(-X_n)$
3. $\limsup_n X_n = \inf_n \sup_{k \geq n} X_k$
4. $\liminf_n X_n = -\limsup_n(-X_n)$

Beachte aber: i.a. ist $\sup X_n \colon \Omega \to \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, wobei $\bar{\mathcal{B}} = \sigma(\mathcal{B} \cup \{-\infty, \infty\})$

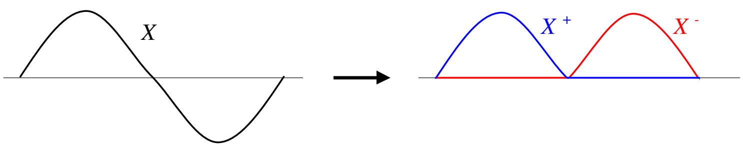
7.4.1 Beweismethodik

Sie haben die Aufgabe, für eine Aussage $(\star)$ herzuleiten, daß sie für jedes meßbare $X \geq 0$ gilt.

Beweismethode: Sei $\mathcal{X}$ die Menge aller meßbaren $X \geq 0$. Definiere zunächst die Menge $\mathcal{M} = \{X \in \mathcal{X} \mid (\star) \text{ gilt für } X\}$. Zeige:

1. $\{X \in \mathcal{X} \mid X \text{ hat endlichen Wertebereich } \} \subseteq \mathcal{M}$
2. Sind $X_1, X_2, \ldots \in \mathcal{M}$ mit $0 \leq X_1 \leq X_2 \leq \ldots$, so ist auch $\lim_{n\to\infty} X_n \in \mathcal{M}$

Aus beiden Eigenschaften folgt: $\mathcal{M} = \mathcal{X}$. Um schließlich etwas für beliebige meßbare Abbildungen nachzuweisen, schreibe man $X = X^+ - X^-$ (mit $X^+ = \max\{X, 0\}$ und $X^- = -\min\{X, 0\}$) als Differenz zweiter meßbarer Abbildungen ≥ 0:



[10] „Das wissen Sie alle: Mogeln spart Zeit, das kennen wir schon aus der Schule.“

7.4.2 Praktische Erwägungen zur Verteilung von Zufallsgrößen

1. Die Verteilung von Zufallsgrößen kann insbesondere festgelegt werden durch

   (a) Angabe der Verteilungsfunktion F^X

   (b) Angabe der Dichte f^X

   Von (a) nach (b) gelangt man durch Differenzieren, sei beispielsweise F gegeben mit $I = \{t \mid 0 < F(t) < 1\}$ und $F \in \mathcal{C}^1(I)$, ableiten ergibt $f^X = (F^X)'$, auf I und $f^X = 0$ sonst.

   Von (b) nach (a) gelangt man durch Integrieren, $F^X(t) = \int_{-\infty}^t f^X(x)\,dx$.

2. Gegeben seien Abbildungen $\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$, X besitze die Dichte f^X, welche Dichte hat dann $g(X)(= g \circ X)$? Bei g streng monoton

   $$P(g(x) \leq t) = P(X \leq g^{-1}(t)) = F^X(g^{-1}(t))$$

   Dichte durch Ableitung, falls F^X stetig differenzierbar, g differenzierbar:

   $$f^{g(X)}(x) = f^X(g^{-1}(x))(g^{-1})'(x)$$

Beispiel: Dichte von X^2? Es ist

$$\begin{aligned} P(X^2 \leq t) &= P(-\sqrt{t} \leq X \leq \sqrt{t}) \\ &= P(X \leq \sqrt{t}) - P(X < -\sqrt{t}) \\ &= F^X(\sqrt{t}) - F^X(-\sqrt{t}) \end{aligned}$$

Durch Ableiten:

$$f^{X^2}(x) = f^X(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} - f^X(-\sqrt{x}) \cdot \frac{1}{2\sqrt{x}}$$

Ist zum Beispiel X N$(0,1)$-verteilt, so besitzt X^2 die Dichte (Gamma-Verteilung):

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \frac{1}{2\sqrt{x}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \cdot \frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$$

8 Erwartungswerte und Integral

8.1 Untersuchung eines Algorithmus'

Seien Zahlen $\omega_1, \ldots, \omega_n$ gegeben. Betrachte den Algorithmus:

$$
\begin{array}{l}
j := n; \\
m := \omega_n; \\
\texttt{for}(k := n-1; k > 0; k--) \\
\quad \texttt{if}(\omega_k > m) \\
\quad \{ \\
\quad\quad j := k; \\
\quad\quad m := \omega_k; \\
\quad \{ \\
\texttt{write}(j, m);
\end{array}
$$

Dieser sucht das Maximum und die Maximalstelle. Untersuche nun das **Laufzeitverhalten** des Algorithmus! Dies führt zu Fragen der Wahrscheinlichkeitstheorie: Bei der Untersuchung von *mittleren Laufzeiten* wird die Eingabe als *zufällig* angesehen und es wird die *durchschnittliche Laufzeit* bestimmt.

Die Anzahl X von Austausschritten ($j := k$; und $m := \omega_k$) ist abhängig vom Vektor $(\omega_1, \ldots, \omega_n)$. Die minimale Anzahl $X = 0$ ergibt sich, falls $\omega_n = \max_{j \leq n} \omega_j$; die maximale Anzahl $X = n-1$ ergibt sich für $\omega_1 > \omega_2 > \ldots > \omega_n$.

Die Aufwandsanalyse von Programmen ergibt sich dadurch, daß die mittlere Anzahl von Programmschritten ermittelt wird, und zwar in einem geeigneten wahrscheinlichkeitstheoretischen Modell.

Modellannahme: $\omega_i \neq \omega_j$ für alle $1 \leq i \leq j \leq n$ mit $i \neq j$, sämtliche Permutationen der n Zahlen sollen gleichwahrscheinlich sein. Betrachten wir $n = 3$:

$$
\begin{array}{lcl}
\omega_1 < \omega_2 < \omega_3 & \Rightarrow & X(\omega) = 0 \\
\omega_1 < \omega_3 < \omega_2 & \Rightarrow & X(\omega) = 1 \\
\omega_2 < \omega_1 < \omega_3 & \Rightarrow & X(\omega) = 0 \\
\omega_2 < \omega_3 < \omega_1 & \Rightarrow & X(\omega) = 1 \\
\omega_3 < \omega_1 < \omega_2 & \Rightarrow & X(\omega) = 1 \\
\omega_3 < \omega_2 < \omega_1 & \Rightarrow & X(\omega) = 2
\end{array}
$$

Jeder Fall tritt dabei mit der Wahrscheinlichkeit von $\frac{1}{6}$ auf, so daß sich ergibt:

$$P(X = 0) = \frac{1}{3} \quad P(X = 1) = \frac{1}{2} \quad P(X = 2) = \frac{1}{6}$$

Als mittleren Wert betrachtet man dabei $0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) = \frac{5}{6}$.

Dieser Wert soll im Allgemeinen bestimmt werden, also $\sum_{k=0}^{n-1} kP(X=k)$. Wie man sieht, ist nur die Anordnung der ω_i, jedoch nicht ihre tatsächliche Größe von Bedeutung, so daß wir ohne Einschränkung der Allgemeinheit annehmen können, daß ω eine Permutation der Zahlen $1, \ldots, n$ ist. Betrachtet wird also das Zufallsexperiment

$$\Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega \text{ Permutation von } \{1, \ldots, n\}\}$$

mit Laplace-Verteilung, also $P_n(A) = \frac{|A|}{n!}$ für alle $A \subseteq \Omega_n$. Ferner sei

$$X_n \colon \Omega \to \{0, \ldots, n-1\} \text{ mit } \omega \mapsto \# \text{ Austauschschritte bei } \omega$$

Eine Definition von Funktionen durch Programme, Flußdiagramme bzw. Automaten ist in der Informatik üblich. Um eine für die Mathematik gebräuchliche Form zu erhalten, benutzen wir einen induktiven Vorgang:

> Es ist $X_1 = 0$, für $n > 1$ definiere $X_n(\omega) = X_{n-1}(\tilde{\omega})$, falls $\omega_1 \neq n$; und $X_n(\omega) = X_{n-1}(\tilde{\omega}) + 1$, falls $\omega_1 = n$, wobei $\tilde{\omega}$ Permutation von $\{1, \ldots, n-1\}$ mit $\tilde{\omega}_i < \tilde{\omega}_j$ genau dann, wenn $\omega_{i+1} < \omega_{j+1}$ für alle $i, j \in \{1, \ldots, n\}$.

Ebenso wie eine explizite Angabe von X_n schwierig ist, ist dies auch für die Wahrscheinlichkeiten $P_{n,k} = P_n(X_n = k)$. Wir benutzen somit wieder ein rekursives Vorgehen: Offensichtlich ist $P_{1,0} = 1$ und $P_{1,1} = 0$. Ferner gilt für jedes $k \in \{0, \ldots, n-1\}$ mit $n \geq 2$

$$\begin{aligned}
P_{n,k} &= P_n(\{\omega \mid X_n(\omega) = k,\ \omega_1 = n\}) + P_n(\{\omega \mid X_n(\omega) = k; \omega_1 \neq n\}) \\
&= \frac{|\{\omega \mid X_n(\omega) = k,\ \omega_1 = n\}|}{n!} + \frac{|\{\omega \mid X_n(\omega) = k; \omega_1 \neq n\}|}{n!} \\
&= \frac{1}{n} P_n(X_{n-1} = k-1) + \frac{n-1}{n} P_n(X_{n-1} = k) \\
&= \frac{1}{n} P_{n-1,k-1} + \frac{n-1}{n} P_{n-1,k}
\end{aligned}$$

wobei $P_{n-1,-1} := 0 = P_{n-1,n-1}$. Mit den vorgegebenen Anfangsbedingungen können wir mit dieser Rekursion, in der die $P_{n,k}$ numerisch berechnet werden und damit auch $\sum_{k=0}^{n-1} kP_{n,k}$, die mittlere Anzahl an Austausschritten.

Um einen expliziten Ausdruck für $\sum_{k=0}^{n-1} kP_{n,k}$ zu ermitteln, gehen wir über zu einem weiteren Hilfsmittel: Wir bilden

$$G_n \colon (0, \infty) \to (0, \infty) \text{ mit } z \mapsto \sum_{k=0}^{n-1} z^k p_{n,k} = \sum_{k=0}^{n-1} z^k P(X_n = k)$$

Dabei ist G_n dier erzeugende Funktion von P^{X_n}, bzw. von X_n. Insbesondere gilt $G_n(1) = 1$.

Bildet man dann die Ableitung $G'_n(z) = \sum_{k=0}^{n-1} kz^{k-1}P_{n,k}$, so erhält man durch $G'_n(1) = \sum_{k=0}^{n} kp_{n,k}$ den gewünschten Wert. In unserem Fall erhält man dann: $G(z) = P_{1,0} + zP_{1,1} = 1$ und mittels der gebildeten rekursiven Beziehung auch

$$\begin{aligned}
G_n(z) &= \sum_{k=0}^{n} z^k P_{n,k} = z\sum_{k=0}^{n-1} z^{k-1}\frac{1}{n}P_{n-1,k-1} + \sum_{k=0}^{n-1} z^k \frac{n-1}{n}P_{n-1,k} \\
&= \frac{z}{n}\sum_{k=1}^{n-2} z^k \frac{1}{n}P_{n-1,k} + \frac{n-1}{n}\sum_{k=0}^{n-1} z^k P_{n-1,k} \\
&= \frac{z}{n}G_{n-1}(z) + \frac{n-1}{n}G_{n-1}(z) \\
&= \frac{z+n-1}{n}G_{n-1}(z)
\end{aligned}$$

Damit erhält man

$$G_n(z) = \frac{z+n-1}{n}G_{n-1}(z) = \frac{z+n-1}{n}\frac{z+n-2}{n-1}G_{n-2}(z) = \ldots = \frac{1}{n!}(z+n-1)\cdot\ldots\cdot(z+1)$$

und $G'_n(z) = \frac{1}{n}G_{n-1}(z) + \frac{z+n-1}{n}G'_{n-1}(z)$, wegen $G'_n(1) = 0$ folgt

$$G'_n(1) = \frac{1}{n} + \frac{1}{n-1} + \ldots + \frac{1}{2} = \sum_{k=2}^{n} \frac{1}{k}$$

Also ist $\sum_{k=0}^{n-1} kP_{n,k} = \sum_{k=2}^{n} \frac{1}{k}$. Die Größenordnung erhält man durch den Vergleich von Summe und Integral:

$$\log(n+1) - \log(2) = \int_1^n \frac{1}{x+1}\,dx \le \sum_{k=2}^{n} \frac{1}{k} \le \int_1^n \frac{1}{x}\,dx = \log n$$

Wir merken an, daß zur Bestimmung des mittleren Wertes von X_n gebildet wurde:

$$\begin{aligned}
\sum_{k=0}^{n-1} kP(X_n = k) &= \sum_{k=0}^{n-1} kP(\{\omega \mid X_n(\omega) = k\}) \\
&= \sum_{k=0}^{n-1} \sum_{\omega\in\{X_n=k\}} kP(\{\omega\}) \\
&= \sum_{k=0}^{n-1} \sum_{\omega\in\{X_n=k\}} X_n(\omega)P(\{\omega\}) \\
&= \sum_{\omega\in\Omega} X_n(\omega)P(\{\omega\})
\end{aligned}$$

8.2 Erwartungswert

DEFINITION: Sei X eine Zufallsgröße mit endlichem Wertebereich. Dann sei der *Erwartungswert* $E(X)$ von X definiert als:

$$E(X) = \sum_{x \in X(\Omega)} xP(X = x)$$

BEISPIELE:

1. Sei X die $\mathrm{B}(n, p)$-Verteilung. Dann gilt[11]:

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k \cdot P(X = k) \\ &= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} n \cdot \binom{n-1}{k-1} p p^{k-1} (1-p)^{n-1-(k-1)} \\ &= n \cdot p \cdot \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k}}_{=(p+(1-p))^{n-1}=1} \\ &= np \end{aligned}$$

2. Ziehen mit und ohne Zurücklegen: Seien N Kugeln gegeben, darunter M weiße und $N - M$ grüne. Bei n Zügen ist X die Anzahl der gezogenen weißen Kugeln.
   - mit Zurücklegen: $X = B(n, \frac{M}{N})$, also $E(X) = n\frac{M}{N}$
   - ohne Zurücklegen: $X = H(N, M, n)$ hypergeometrische Verteilung, also $P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$. Der Erwartungswert berechnet sich

[11] mit $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$

als

$$
\begin{aligned}
E(X) &= \sum_{k=0}^{n} \frac{k\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \\
&= \sum_{k=1}^{n} \frac{M\binom{M-1}{k-1}\binom{(N-1)-(M-1)}{n-1-(k-1)}}{\frac{N}{n}\binom{N-1}{n-1}} \\
&= n\frac{M}{N} \cdot \sum_{k=0}^{n-1} \frac{\binom{M-1}{k}\binom{(N-1)-(M-1)}{n-1-k}}{\binom{N-1}{n-1}} \\
&= n\frac{M}{N} \cdot \sum_{k=0}^{n-1} (H(N-1, M-1, n-1))(\{k\}) \\
&= n\frac{M}{N}
\end{aligned}
$$

8.2.1 Eigenschaften

1. Sei X Zufallsgröße mit endlichem Wertebereich, $X = \sum_{i=1}^{n} \alpha_i 1_{A_i}$ mit paarweise verschiedenen A_i und $\sum_{i=1}^{n} A_i = \Omega$. Dann gilt:

$$
\begin{aligned}
E(X) &= \sum_{x \in X(\Omega)} xP(\{\omega \mid X(\omega) = x\}) \\
&= \sum_{x \in X(\Omega)} \left(x \sum_{i,\, x=\alpha_i} P(A_i) \right) \\
&= \sum_{x \in X(\Omega)} \sum_{i,\, x=\alpha_i} \alpha_i P(A_i) \\
&= \sum_{i=1}^{n} \alpha_i P(A_i)
\end{aligned}
$$

2. Seien X, Y Zufallsgrößen mit endlichem Wertebereich,

$$
X = \sum_{i=1}^{n} \alpha_i 1_{A_i} = \sum_{i,j} \alpha_i 1_{A_i \cap B_j} \text{ und } Y = \sum_{j=1}^{m} \beta_j 1_{B_j} = \sum_{i,j} \beta_i 1_{A_i \cap B_j}
$$

   (a) Aus $X \leq Y$ folgt $E(X) \leq E(Y)$, denn wegen $\alpha_i \leq \beta_j$ auf $A_i \cap B_j$ ist

$$
E(X) = \sum_{i,j} \alpha_i 1_{A_i \cap B_j} \leq \sum_{i,j} \beta_i 1_{A_i \cap B_j} = E(Y)
$$

(b) Für $a, b \in \mathbb{R}$ gilt

$$\begin{aligned} E(aX + bY) &= \sum_{i,j} (a\alpha_i + b\beta_j) 1_{A_i \cap B_j} \\ &= \sum_{i,j} a\alpha_i 1_{A_i \cap B_j} + \sum_{i,j} b\beta_i 1_{A_i \cap B_j} \\ &= aE(X) + bE(Y) \end{aligned}$$

Damit ist das Bilden des Erwartungswertes **linear und monoton**!

DEFINITION: Für Zufallsgrößen $X \geq 0$ sei definiert:

$$E(X) = \sup \{ E(Y) \mid Y \text{ Zufallsgr. mit endl. Wertebereich } 0 \leq Y \leq X \}$$

Ist X allgemeine Zufallsgröße, so schreibe $X = X^+ - X^-$. X heißt *regulär*, falls $E(X^+) < \infty$ oder $E(X^-) < \infty$. Es wird dann definiert:

$$E(X) = E(X^+) - E(X^-)$$

X heißt *integrierbar*, falls gilt: $E(X^+) < \infty$ und $E(X^-) < \infty$. Man beachte: Wegen $|X| = X^+ + X^-$ ist X integrierbar genau dann, wenn $E(|X|) < \infty$ ist. Ist diese allgemeinere Erwartungswertbildung immer noch **linear und monoton**?

8.2.2 Monotonie und Linearität

Aus $X \geq 0$, $Y \geq 0$, $X \leq Y$ folgt

$$\begin{aligned} E(X) &= \{ E(Z) \mid Z \text{ Zufallsgr. mit endl. Wertebereich } 0 \leq Z \leq X \} \\ &\leq \{ E(Z) \mid Z \text{ Zufallsgr. mit endl. Wertebereich } 0 \leq Z \leq Y \} \\ &= E(Y) \end{aligned}$$

Schwierig ist der Nachweis der Linearität! Es gilt:

$$\sup_{x \in \mathcal{X}} (f(x) + g(x)) \leq \sup_{x \in \mathcal{X}} f(x) + \sup_{x \in \mathcal{X}} g(x)$$

Seien $(X_n)_{n \in \mathbb{N}}$ und $(Y_n)_{n \in \mathbb{N}}$ Folgen mit $X_n \uparrow X$ und $Y_n \uparrow Y$, dann ist $E(X_n) + E(Y_n) = E(X_n + Y_n)$. Wir benötigen noch folgende Aussage:

$$X_n \uparrow X \Longrightarrow E(X_n) \uparrow E(X) \qquad (\star)$$

Dann haben wir:

$$EX + EY = \lim_{n \to \infty} (E(X_n) + E(Y_n)) = \lim_{n \to \infty} E(X_n + Y_n) = E(X + Y)$$

Der Schlüssel zum Umgang mit Erwartungswertbildung ist also eine Aussage vom Typ $(\star)$, bekannt als der *Satz von der monotonen Konvergenz* aus der Maßtheorie.

8.3 Das allgemeine Maßintegral

Statt eines Wahrscheinlichkeitsmaßes betrachten wir ein allgemeines Maß μ. Definiere wie bisher ($\int X\,d\mu \to E(X)$) mit $X \geq 0$ mit endlichem Wertebereich:

$$\int X\,d\mu = \sum_{x \in X(\Omega)} x\mu(\{\omega \mid X(\omega) = x\})$$

und für allgemeines $X \geq 0$:

$$\int X\,d\mu = \sup\left\{\int Y\,d\mu \;\middle|\; Y \text{ Zufallsgr. mit endl. Wertebereich } 0 \leq Y \leq X\right\}$$

und allgemein

$$\int X\,d\mu = \int X^+\,d\mu - \int X^-\,d\mu$$

8.3.1 Satz von der monotonen Konvergenz

Satz: Seien $0 \leq X_1 \leq X_2 \leq \ldots \leq X$ mit $\lim_{n\to\infty} X_n = X$. Dann gilt:

$$\int \lim_{n\to\infty} X_n\,d\mu = \lim_{n\to\infty} \int X_n\,d\mu$$

Beweis: Die Monotonie von $\int \cdot\,d\mu$ zeigt

$$0 \leq \int X_1\,d\mu \leq \int X_2\,d\mu \leq \ldots \leq \int X\,d\mu$$

Also existiert $\lim_{n\to\infty} \int X_n\,d\mu \leq \int X\,d\mu$. Zu zeigen ist: Für beliebiges $Y = \sum_{i=1}^n \alpha_i 1_{A_i} \leq X$ gilt: $\lim_{n\to\infty} \int X_n\,d\mu \geq \int Y\,d\mu$.

Wähle ein solches Y. Falls $\int Y\,d\mu = 0$, so ist die Behauptung klar. Sei also $\int Y\,d\mu > 0$. Sei

$$0 < \beta < \int Y\,d\mu = \sum_{i=1}^n \alpha_i \mu(A_i)$$

Zu zeigen ist: $\lim_{n\to\infty} \int X_n\,d\mu \geq \beta$. Wähle $c > 1$ mit $\beta < \sum_{i=1}^n \frac{\alpha_i}{c}\mu(A_i)$. Es gilt:

$$\lim_{n\to\infty} X_n = X \geq \sum_{i=1}^n \alpha_i 1_{A_i} \geq \sum_{i=1}^n \frac{\alpha_i}{c} 1_{A_i}$$

Für jedes ω existiert ein m mit $X_k(\omega) > \sum_{i=1}^n \frac{\alpha_i}{c} 1_{A_i}(\omega)$ für $k > m$ Definiere

$$A_{i,k} = \left\{\omega \mid \omega \in A_i, X_k(\omega) \geq \sum_{l=1}^n \frac{\alpha_l}{c} 1_{A_l}(\omega)\right\}$$

Es gilt für jedes $i = 1, \dots, n$:

$$A_{i,1} \subseteq A_{i,2} \subseteq \ldots \Rightarrow \bigcup_{k \in \mathbb{N}} A_{i,k} = A_i \tag{$*$}$$

Es gilt

$$\mu(A_i) = \mu\left(\bigcup_{m \in \mathbb{N}} A_{i,m}\right) = \lim_{m \to \infty} \mu(A_{i,m})$$

Damit gilt:

$$X_m \overset{\text{Def.}A_{i,m}}{\geq} \sum_{i=1}^{n} \frac{\alpha_i}{c} 1_{A_{i,m}}$$

Also gilt weiter: $\int X_m \, d\mu \geq \sum_{i=1}^{n} \frac{\alpha_i}{c} \mu(A_{i,m})$ und

$$\lim_{m \to \infty} \int X_m \, d\mu \geq \sum_{i=1}^{n} \frac{\alpha_i}{c} \lim_{m \to \infty} \mu(A_{i,m}) \overset{(*)}{=} \sum_{i=1}^{n} \frac{\alpha_i}{c} \mu(A_i) > \beta$$

Satz:

1. mit $X, Y \geq 0$ und $a, b \geq 0$ ist

$$\int (aX + bY) \, d\mu = a \int X \, d\mu + b \int Y \, d\mu$$

2. für allgemeine, integrierbare X, Y und $a, b \in \mathbb{R}$ ist $aX + bY$ integrierbar und

$$\int (aX + bY) \, d\mu = a \int X \, d\mu + b \int Y \, d\mu$$

Beweis:

1. Zu X, Y wähle Folgen $(X_n)_{n \in \mathbb{N}}$ und $(Y_n)_{n \in \mathbb{N}}$ mit $0 \leq X_1 \leq X_2 \leq \ldots \uparrow X$ und $0 \leq Y_1 \leq Y_2 \leq \ldots \uparrow Y$ Dann gilt: $aX_1 + bY_1 \leq aX_2 + bY_2 \leq \ldots \uparrow aX + bY$. Mit dem Satz von der monotonen Konvergenz folgt die

Behauptung:

$$\begin{aligned}
a\int X\,d\mu + b\int Y\,d\mu &= a\lim_{n\to\infty}\int X_n\,d\mu + b\lim_{n\to\infty}\int Y_n\,d\mu \\
&= \lim_{n\to\infty}\left(a\int X_n\,d\mu + b\int Y_n\,d\mu\right) \\
&= \lim_{n\to\infty}\int (aX_n + bY_n)\;d\mu \\
&= \int \lim_{n\to\infty}(aX_n + bY_n)\;d\mu \\
&= \int aX_n + bY_n\,d\mu
\end{aligned}$$

2. Die Behauptung folgt aus folgenden Aussagen:

   (a) $\int X\,d\mu = -\int(-X)\,d\mu$

   (b) $\int aX\,d\mu = a\int X\,d\mu$ für $a \geq 0$

   (c) $\int(X+Y)\,d\mu = \int X\,d\mu + \int Y\,d\mu$

   Es gilt:

$$\begin{aligned}
\int |aX + bY|\;d\mu &\leq \int(|a|\,|X| + |b|\,|Y|)\,d\mu \\
&= |a|\int \underbrace{|X|\;d\mu}_{<\infty} + |b|\int \underbrace{|Y|\;d\mu}_{<\infty} < \infty
\end{aligned}$$

   Es gilt nun:

   (a) Es ist $(-X)^+ = X^-$ und $(-X)^- = X^+$, also:

$$\begin{aligned}
\int(-X)\,d\mu &= \int(-X)^+\,d\mu - \int(-X)^-\,d\mu \\
&= \int X^-\,d\mu - \int X^+\,d\mu = -\int X\,d\mu
\end{aligned}$$

   (b) Sei $a \geq 0$, dann ist $(aX)^+ = aX^+$ und $(aX)^- = aX^-$, also:

$$\begin{aligned}
\int aX\,d\mu &= \int aX^+\,d\mu - \int aX^-\,d\mu \\
&= a\int X^+\,d\mu - a\int X^-\,d\mu = a\int X\,d\mu
\end{aligned}$$

(c) Es ist $X+Y = X^+ - X^- + Y^+ - Y^- = (X+Y)^+ - (X+Y)^-$, daraus folgt

$$(X+Y)^+ + X^- + Y^- = (X+Y)^- + X^+ + Y^+$$

Damit folgt:

$$\int (X+Y)^+ \, d\mu + \int X^- \, d\mu + \int Y^- \, d\mu = \int (X+Y)^- \, d\mu + \int X^+ \, d\mu + \int Y^+ \, d\mu$$

Also ist $\int (X+Y) \, d\mu = \int X \, d\mu + \int Y \, d\mu$.

Satz: (Lemma von Fatou) Seien $X_1, X_2, \ldots$ meßbare Abbildungen. Es existiere ein integrierbares Y mit $Y \leq X_n$ für alle n. Dann gilt:

$$\int \liminf_n X_n \leq \liminf_n \int X_n$$

Beweis: Zur Erinngerung: Es ist

$$\liminf_n X_n = \sup_n \underbrace{\inf_{k \geq n} X_k}_{z_n} = lim_n Z_n \text{ mit } Z_1 \leq Z_2 \ldots$$

Nun ist

$$\begin{aligned}
\int \liminf_n (X_n - Y) \, d\mu &= \int \lim_n (Z_n - Y) \, d\mu \\
&\overset{(8.3.1)}{=} \lim_n \int (Z_n - Y) \, d\mu \\
&= \lim_n \int Z_n \, d\mu - \int Y \, d\mu \\
&= \lim_n \int \inf_{k \geq n} X_k \, d\mu - \int Y \, d\mu \\
\textit{Monotonie des Integrals} \quad &\leq \lim_n \inf_{k \geq n} \int X_k \, d\mu - \int Y \, d\mu
\end{aligned}$$

Satz: (Satz von Lebesgue, Satz von der dominanten Konvergenz) Seien $X, X_1, X_2, \ldots$ meßbare Abbildungen mit $\lim_n X_n = X$. Es existieren integrierbare Abbildungen Y_1, Y_2 mit $Y_1 \leq X_n \leq Y_2$ für alle n. Dann gilt:

$$\int \lim_n X_n \, d\mu = \lim_n \int X_n \, d\mu$$

Beweis: Mit dem Lemma von Fatou gilt:

$$\int \liminf_n X_n \, d\mu \leq \liminf_n \int X_n \, d\mu$$

und

$$\int \liminf_n (-X_n) \, d\mu \leq \liminf_n \int (-X_n) \, d\mu$$

Also ist[12]

$$-\int \limsup_n X_n \, d\mu \leq -\limsup_n \int X_n \, d\mu$$

Zusammen ergibt sich:

$$\int \underbrace{\liminf_n X_n}_{=X} \, d\mu \leq \liminf_n \int X_n \, d\mu \leq \limsup_n \int X_n \, d\mu \leq \int \underbrace{\limsup_n X_n}_{=X} \, d\mu$$

Damit ist also

$$\int X \, d\mu = \underbrace{\liminf_n \int X_n \, d\mu = \limsup_n \int X_n \, d\mu}_{=\lim \int X_n \, d\mu} = \int X \, d\mu$$

Bemerkungen:

- Es wird definiert $\int_A X \, d\mu = \int X 1_A \, d\mu$, Schreibweisen:

$$\int_A X \, d\mu = \int_A X(\omega)\mu(d\omega)$$

- Mit obiger Definition ergibt sich z.B. für integrierbares X bzw. $X \geq 0$:

$$\int_{\sum_n A_n} X \, d\mu = \sum_n \int_{A_n} X \, d\mu$$

Begründung dazu: Mit obigen Sätzen (und $X 1_{\sum_{n=1}^k A_n} \to X 1_{\sum_n A_n}$)

[12] mit $\liminf(-X_n) = \limsup X_n$

folgt:

$$\begin{aligned}
\sum_n \int_{A_n} X\, d\mu &= \lim_{k\to\infty} \sum_{n=1}^{k} \int X 1_{A_n}\, d\mu \\
&= \lim_{k\to\infty} \int \sum_{n=1}^{k} X 1_{A_n}\, d\mu \\
&= \lim_{k\to\infty} \int X 1_{\sum_{n=1}^{k} A_n}\, d\mu \\
&= \lim_{k\to\infty} \int X 1_{\sum_{n=1}^{k} A_n}\, d\mu \\
&\overset{(\star)}{=} \int \lim_{k\to\infty} X 1_{\sum_{n=1}^{k} A_n}\, d\mu \\
&= \int X 1_{\sum_n A_n}\, d\mu
\end{aligned}$$

Dabei folgt $(\star)$ bei integrierbarem X aus dem Satz von der dominanten Konvergenz, bei $X \geq 0$ aus dem Satz der monotonen Konvergenz.

8.3.2 Wahrscheinlichkeitsmaße mit Dichten - der allgemeine Fall

Sei μ ein Maß auf $(\Omega, \mathfrak{A})$. Sei $f\colon \Omega \to \mathbb{R}$ eine meßbare Abbildung $f \geq 0$ mit $\int f\, d\mu = 1$. Definiere

$$P : \mathfrak{A} \to [0,1] \text{ durch } P(A) = \int_A f\, d\mu$$

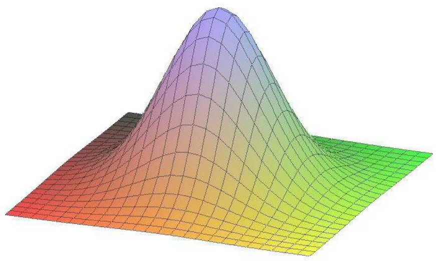
Dann ist P ein Wahrscheinlichkeitsmaß, beachte dazu die Bemerkung oben. Offensichtlich gilt:

$$\begin{aligned}
P(\emptyset) &= \int f 1_{\emptyset}\, d\mu = \int 0\, d\mu = 0 \\
P(\Omega) &= \int f 1_{\Omega}\, d\mu = \int f\, d\mu = 1 \\
P\left(\sum_n A_n\right) &= \sum_n \int_{A_n} f\, d\mu = \sum_n P(A_n)
\end{aligned}$$

Anwendung: Wahrscheinlichkeitsmaße auf $\mathbb{R}^n$ werden oft unter Benutzung dieser Möglichkeit angegeben. Benutze dazu $(\Omega, \mathfrak{A}) = (\mathbb{R}^n, \mathcal{B}^n)$ mit $\mu = \lambda^n$.

Beispiel: Zweidimensionale Glockenkurve:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}} = \frac{1}{2\pi} e^{-\frac{x_1^2 + x_2^2}{2}}$$



n-dimensionale Glockenkurve:

$$f(x_1,\dots,x_n) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \;\forall\; (x_1,\dots,x_n) \in \mathbb{R}^n$$

Es ist

$$\begin{aligned}\int_{\mathbb{R}^n} f(x)\lambda^n(dx) &= \int_{\mathbb{R}^n} \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}\, dx_1 \dots dx_n \\ &= \prod_{i=1}^n \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}\, dx_i = 1\end{aligned}$$

Das zugehörige Wahrscheinlichkeitsmaß heißt *n-dimensionale Standard-Normalverteilung.*

8.3.3 Lebesgue-Integral

Sei $g\colon \mathbb{R} \to \mathbb{R}$ stetig. Dann existieren das Riemann-Integral $\int_a^b g(x)\,dx$ und das Integral von $g \cdot 1_{[a,b]}$ bezüglich des Lebesgue-Maßes λ: $\int_{[a,b]} g\,d\lambda$. Beide sind „natürlich“ gleich: Sei $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$ eine Zerlegung $\mathcal{Z}$ des Intervalls $[a,b]$ mit $w(\mathcal{Z}) = \max\{\,|x_i - x_{i-1}| \mid i = 1,\dots,n\}$. Dann gilt:

$$\int_a^b g(x)\,dx = \lim_{w(\mathcal{Z})\to 0} \sum_{i=1}^n g(x_{i-1})\,|x_i - x_{i-1}|$$

Betrachte

$$g_{\mathcal{Z}} = g(a)1_{[a,x_1)} + \sum_{i=2}^{n-1} g(x_{i-1})1_{[x_{i-1},x_i)} + g(x_{n-1})1_{[x_{n-1},b]}$$

Dann gilt:

$$\int g_{\mathcal{Z}}\,d\lambda = \sum_{i=1}^n g(x_{i-1})\,|x_i - x_{i-1}|$$

Für eine Folge von Zerlegungen $(\mathcal{Z}_n)_{n\in\mathbb{N}}$ mit $w(\mathcal{Z}_n) \longrightarrow 0$ gilt: $g_{\mathcal{Z}_n} \longrightarrow g \cdot 1_{[a,b]}$. Mit dem Satz von der dominanten Konvergenz gilt: $\int g_{\mathcal{Z}_n}\,d\lambda \longrightarrow \int g \cdot 1_{[a,b]}\,d\lambda$, andererseits gilt $\int g_{\mathcal{Z}_n}\,d\lambda \longrightarrow \int_a^b g(x)\,dx$, also $\int_a^b g(x)\,dx = \int_{[a,b]} g\,d\lambda$.

Deshalb schreiben wir statt $\int_B g\,d\lambda$ zumeist $\int_B g(x)\,dx$.

8.3.4 Integration bzgl. durch Dichten definierte Wahrscheinlichkeitsmaße

Schon kennengelernt: die Angabe von Wahrscheinlichkeitsmaßen durch Dichten: Seien $(\Omega, \mathfrak{A})$ und μ gegeben, sei $f\colon \Omega \to \mathbb{R}$ meßbar, $f \geq 0$ mit $\int f\,d\mu = 1$. Dann definiert $P(A) = \int_A f\,d\mu$ ein Wahrscheinlichkeitsmaß. f heißt auch μ-Dichte von P ($f = \frac{dP}{d\mu}$). Dabei gilt für eine Zufallsgröße X (regulär bezüglich P):

$$\int X\,dP = \int X \cdot f\,d\mu$$

Beweis:

(a) Für $X = 1_A$ steht auf der linken Seite $\int 1_A\,dP = P(A)$, auf der rechten Seite $\int 1_A f\,d\mu = \int_A f\,d\mu = P(A)$.

(b) Für $X = \sum_{i=1}^n \alpha_i 1_{A_i}$ ist die linke Seite

$$\int X\,dP = \int \sum_{i=1}^n \alpha_i 1_{A_i}\,dP = \sum_{i=1}^n \alpha_i P(A_i)$$

Die rechte Seite ergibt:

$$\int Xf\,dP = \int \left(\sum_{i=1}^n \alpha_i 1_{A_i}\right) f\,d\mu = \sum_{i=1}^n \alpha_i \int 1_{A_i} f\,d\mu = \sum_{i=1}^n \alpha_i P(A_i)$$

(c) Für $X \geq 0$: Es existiert eine Folge $(X_n)_{n\in\mathbb{N}}$ gemäß (c) mit $0 \leq X_n \uparrow X$. Es gilt:

$$= \int X\,dP = \lim_{n\to\infty} \int X_n\,dP = \lim_{n\to\infty} \int X_n f\,d\mu = \int \lim_{n\to\infty} (X_n f)\,d\mu = \int Xf\,d\mu$$

8.3.5 Integration bzgl. der Verteilung von Zufallsvariablen

Sei Ω mit Wahrscheinlichkeitsmaß P gegeben, $X\colon \Omega \to \mathcal{X}$ und $g\colon \mathcal{X} \to \mathbb{R}$. Dann ist

$$Eg(X) = \int g(X)\,dP = \int g\,dP^X$$

Beachte:

> *Erwartungswertbildung ist Integration bezüglich des zugrundeliegenden Wahrscheinlichkeitsmaßes!*

Beweis:

(a) Für $g = 1_A$ steht auf der linken Seite $\int 1_A(X)\,dP = \int 1_{X^{-1}(A)}\,dP = P^X(A)$, auf der rechten Seite $\int 1_A\,dP^X = P^X(A)$

(b) Für $g = \sum_{i=1}^n \alpha_i 1_{A_i}$ folgt die Behauptung aus (a) mit der Linearität des Integrals

(c) Für $g \geq 0$: Folgt gemäß (b) aus dem Satz der monotonen Konvergenz mit einer Folge $0 \leq g_n \uparrow g$.

(d) Für g benutze (c) mit $g = g^+ - g^-$.

8.3.6 Anwendungen

1. Sei X $N(a, \sigma^2)$-verteilt. Dann ergibt sich[13] mit $g(x) = x$:

$$\begin{aligned}
E(X) &= \int xP^X(dx) \\
&= \int xN(a,\sigma^2)(dx) \\
&= \int xf(x)\,dx \\
&= \int x\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx \\
&= \int_{-\infty}^{+\infty}(x+a)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\underbrace{\int_{-\infty}^{+\infty} xe^{-\frac{x^2}{2\sigma^2}}\,dx}_{=0} + a\underbrace{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}\,dx}_{=1} \\
&= a
\end{aligned}$$

2. Allgemein können wir formulieren: Besitzt eine Zufallsgröße X eine Verteilung mit stetiger Dichte f, so ergibt sich ihr Erwartungswert als $E(X) = \int xf(x)\,dx$.

3. Für X $\mathrm{R}(a,b)$-verteilt ist

$$E(X) = \frac{1}{b-a}\int_a^b x\,dx = \frac{b+a}{2}$$

[13] zur Schreibweise: $\int X\,dP = \int \mathrm{id}\,dP^X = \int xP^X(dx)$

8.4 Varianz

Der Erwatungswert ist als der mittlere Wert einer Zufallsgröße anzusehen, um den die tatsächlich registrierten betrachteten Werte streuen. Als Maßzahl dafür wird die *Varianz* definiert:

DEFINITION: Die *Varianz von X* ist

$$\mathrm{Var}(X) = \int (X - E(X))^2 \, dP = E((X - E(X))^2)$$

Die *Standardabweichung* oder *Streuung von X* ist

$$\sigma(X) = \sqrt{\mathrm{Var}(X)}$$

Unter Benutzung der Verteilung von X schreibt sich dies als $\int (x - E(X))^2 p^X(dx)$, dies ist bei Vorliegen einer stetigen Dichte f gleich $\int_{-\infty}^{+\infty} (x - E(X))^2 f(x) \, dx$.

8.4.1 Anmerkungen

Sei X Zufallsgröße mit $E(X) < \infty$ und $\mathrm{Var}(X) < \infty$.

- Es gilt:

$$\begin{aligned} \mathrm{Var}(aX + b) &= E(((aX + b) - E(aX + b))^2) \\ &= E((aX - E(aX) + b - b)^2) \\ &= \mathrm{Var}(aX) = a^2 \, \mathrm{Var}(X) \end{aligned}$$

- Es gilt:

$$\begin{aligned} E((X - E(X))^2) &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

  Das zeigt $E(X^2) \geq E(X)^2$.

- Weiter ist $\mathrm{Var}(X) = 0$ genau dann, wenn $P(X = E(X)) = 1$ (vgl. Übungsaufgabe)

- Für alle $a \in \mathbb{R}$ ist $E((X - a)^2) \geq E((X - E(X))^2)$, denn

$$\begin{aligned} E((X - a)^2) &= E((X - E(X) + E(X) - a)^2) \\ &= E((X - E(X))^2) + 2E((X - E(X))(E(X) - a)) + (E(X) - a)^2 \\ &= E((X - E(X))^2) + 2(E(X) - a)E(X - E(X)) + (E(X) - a)^2 \\ &= E((X - E(X))^2) + 2(E(X) - a)(E(X) - E(X)) + (E(X) - a)^2 \\ &= E((X - E(X))^2) + (E(X) - a)^2 \\ &\geq E((X - E(X))^2) \end{aligned}$$

8.4.2 Чебышев-Ungleichung

SATZ: Sei X Zufallsgröße mit endlichem Erwartungswert. Dann gilt für jedes $\varepsilon > 0$:[14]

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

Es gilt:

$$\varepsilon^2 1_{\{|X-E(X)|\geq\varepsilon\}} \leq (X - E(X))^2$$

Dabei ist die linke Seite

$$\begin{cases} \varepsilon^2 & \text{falls} \quad |X(\omega) - E(X)| \geq \varepsilon \\ 0 & \text{falls} \quad |X(\omega) - E(X)| < \varepsilon \end{cases}$$

dabei ist im ersten Fall $(X(\omega) - E(X))^2 \geq \varepsilon^2$, damit gilt die Ungleichung. Mit der Monotonie des Integrals folgt:

$$\begin{aligned} \int \varepsilon^2 1_{\{|X-E(X)|\geq\varepsilon\}}\, dP &\leq \int (X - E(X))^2\, dP \\ \Rightarrow \varepsilon^2 P(|X - E(X)| \geq \varepsilon) &\leq \mathrm{Var}(X) \end{aligned}$$

[14]ausführlich:

$$\begin{aligned} P(\{\omega \mid |X(\omega) - E(X)| \geq \varepsilon\}) &= P^X(\{x \mid |x - E(X)| \geq \varepsilon\}) \\ &= P^X((-\infty, E(X) - \varepsilon]) + P^X([E(X) + \varepsilon, \infty)) \end{aligned}$$

9 Momente und stochastische Ungleichungen

9.1 Momente, Median

DEFINITION: Sei X Zufallsgröße. Dann heißt

1. $E(X^n)$ *n-tes Moment*
2. $E((X - E(X))^n)$ *n-tes zentriertes Moment*

Dann ist der Erwartungsmoment das erste Moment, die Varianz das zweite zentrierte Moment und die Schiefe das dritte zentrierte Moment.

Eine weitere Kenngröße ist der Median:

DEFINITION: Sei X Zufallsgröße. Dann heißt folgende Zahl *Median*:

$$\operatorname{med}(X) := \inf\left\{ t \mid P(X \le t) \ge \frac{1}{2} \right\}$$

9.2 Ungleichungen

Frage: Wie läßt sich $|\operatorname{med}(X) - E(X)|$ abschätzen? Zur Beantwortung dieser Frage wird die folgende Eigenschaft des Medians benötigt:

SATZ: Sei X Zufallsgröße. Dann gilt:

$$E(|X - \operatorname{med}(X)|) \le E(|X - a|) \ \forall\, a \in \mathbb{R}$$

BEWEIS: Sei $a \in \mathbb{R}$, $a > \operatorname{med}(X) =: m$. Dann gilt:

$$\begin{aligned}
& |X-a| - |X-m| \\
= \ & 1_{\{X \le m\}} \cdot (a - X - (m - X)) + 1_{\{X \ge a\}} \cdot (X - a - (X - m)) \\
& \quad + 1_{\{m < X < a\}} \cdot (a - X - (X - m)) \\
= \ & 1_{\{X \le m\}} \cdot (a - m) + 1_{\{X \ge a\}} \cdot (m - a) + 1_{\{m < X < a\}} \cdot (m + \underbrace{a - X}_{>0} - X) \\
\ge \ & 1_{\{X \le m\}} \cdot (a - m) + 1_{\{X \ge a\}} \cdot (m - a) + 1_{\{m < X < a\}} \cdot (m - a) \\
= \ & 1_{\{X \le m\}} \cdot (a - m) + 1_{\{X > m\}} \cdot (m - a)
\end{aligned}$$

Es folgt für die Erwartungswerte:

$$\begin{aligned}
E(|X-a| - |X-m|) \ & \ge \ (a-m) \cdot \underbrace{P(X \le m)}_{\ge \frac{1}{2}} - (a-m) \cdot \underbrace{P(X > m)}_{< \frac{1}{2}} \\
& \ge \ (a-m) \cdot P(X \le m) - (a-m) \cdot P(X > m)
\end{aligned}$$

Damit ist für $a > \mathrm{med}(X)$ (und analog für $a < \mathrm{med}(X)$):

$$E(|X-a|) \geq E(|X-\mathrm{med}(X)|)$$

SATZ: Sei X Zufallsgröße. Dann gilt

$$|E(X)-\mathrm{med}(X)| \leq \sqrt{\mathrm{Var}(X)}$$

BEWEIS:

$$\begin{aligned} |E(X)-\mathrm{med}(X)| &= |E(X-\mathrm{med}(X))| \leq E(|X-\mathrm{med}(X)|) \\ &\leq E(|X-E(X)|) \leq \sqrt{E((X-E(X))^2)} \\ &= \sqrt{\mathrm{Var}(X)} \end{aligned}$$

9.2.1 CAUCHY-SCHWARZ-Ungleichung

SATZ: Seien X, Y Zufallsgrößen mit $E(X^2) < \infty$ und $E(Y^2) < \infty$. Dann ist XY integrierbar, und es gilt

$$E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$$

BEWEIS: Es gilt $0 \leq (-|X|+|Y|)^2 = X^2+Y^2-2|X||Y|$, also ist

$$E(|XY|) \leq \frac{E(X^2)+E(Y^2)}{2} < \infty$$

Für jedes $c \in \mathbb{R}$ gilt: $0 \leq (-c|X|+|Y|)^2 = c^2|X|^2+|Y|^2-2c|XY|$. Also folgt mit $c = E(|XY|)\cdot E(X^2)^{-1}$ (falls $E(X^2) > 0$):

$$0 \leq c^2E(X^2)+E(Y^2)-2c\cdot E(|XY|) = \frac{E(|XY|)^2}{E(X^2)}+E(Y^2)-2\frac{E(|XY|)^2}{E(X^2)}$$

Die Behauptung folgt für $E(X^2) > 0$ also aus:

$$E(|XY|)^2 \leq E(X^2)E(Y^2)$$

Für $E(X^2) = 0$ folgt $P(X=0) = 1$, also $P(XY=0) = 1$, also $E(|XY|) = 0 = \sqrt{E(X^2)E(Y^2)}$.

9.2.2 JENSENsche Ungleichung

SATZ: Sei $I \subseteq \mathbb{R}$ ein Intervall, $f\colon I \to \mathbb{R}$ sei konvex. Dann gilt für jede Zufallsgröße $X\colon \Omega \to I$:

$$f(E(X)) \leq E(f(X))$$

BEWEIS: Da $X\colon \Omega \to I$ vorliegt, gilt - wie man sich sofort überlegt - $E(X) \in I$. Für alle $x \in I$ gilt

$$f(x) \geq f(E(X)) + f'(E(X)) \cdot (x - E(X))$$

Durch Integration bezüglich P^X gilt:

$$\int f(x)P^X(dx) \geq f(E(X)) + f'(E(X)) \cdot \int (x - E(X))P^X(dx)$$

also

$$\begin{aligned} E(f(X)) &= \int f(x)P^X(dx) \geq \ldots \\ &= f(E(X)) + f'(E(X)) \cdot \underbrace{E(X - E(X))}_{E(X)-E(X)=0} \\ &= f(E(X)) \end{aligned}$$

Falls f in $E(X)$ nicht differenzierbar ist, so benutze man die bei konvexen Funktionen stets existierende rechtsseitige Ableitung in $E(X)$.

Anmerkung: Ist g konkav, so ist $f = -g$ konvex und mit obiger Ungleichung gilt

$$-g(E(X)) \leq -E(g(X)) \text{ also } g(E(X)) \geq E(g(X))$$

Damit ist z.B. bei $X \geq 0$: $\log(E(X)) \geq E(\log(X))$ und

$$(E(|Z|))^p \begin{cases} \leq E(|Z|^p) & \text{falls } p \geq 1 \\ \geq E(|Z|^p) & \text{falls } p \leq 1 \end{cases}$$

Falls $x_i = \log y_i$ ist für $y_1, \ldots, y_n > 0$, und $f(x) = e^x$, so ist

$$\prod_{i=1}^n y_i^{p_i} = e^{\sum_{i=1}^n p_i \log y_i} \leq \sum_{i=1}^n p_i y_i$$

Damit gilt mit $p_i = \frac{1}{n}$ die *Ungleichung zwischen geometrischen und arithmetischem Mittel*:

$$\left(\prod_{i=1}^n y_i\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^n p_i$$

10 Stochastische Unabhängigkeit

Ereignisse A, B sind stochastisch unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$.

DEFINITION: Seien $X_1 \colon \Omega \to \mathcal{X}_1$ und $X_2 \colon \Omega \to \mathcal{X}_2$ Zufallsvariablen. X_1 und X_2 heißen *stochastisch unabhängig*, falls für alle meßbaren $D_1 \subseteq \mathcal{X}_1$ und $D_2 \subseteq \mathcal{X}_2$ gilt:

$$P(X_1 \in D_1, X_2 \in D_2) = P(X_1 \in D_1) \cdot P(X_2 \in D_2)$$

BEMERKUNG: Beachte dabei

$$\begin{aligned} \{X_1 \in D_1, X_2 \in D_2\} &= \{\omega \mid X_1(\omega) \in D_1, X_2(\omega) \in D_2\} \\ &= \{\omega \mid X_1(\omega) \in D_1\} \cap \{\omega \mid X_2(\omega) \in D_2\} \\ &= X_1^{-1}(D_1) \cap X_2^{-1}(D_2) \end{aligned}$$

Also: X_1, X_2 sind stochastisch unabhängig genau dann, wenn sämtliche Ereignisse der Form $X_1^{-1}(D_1)$, $X_2^{-1}(D_2)$ stochastisch unabhängig sind. Diese Definition läßt sich wie folgt erweitern:

DEFINITION: $X_i \colon \Omega \to \mathcal{X}_i$ für $i = 1, \ldots, n$ heißen stochastisch unabhängig, falls für alle meßbaren $D_i \subseteq \mathcal{X}_i$ gilt:

$$P(X_i \in D_i, \ldots, X_i \in D_i) = \prod_{i=1}^{n} P(X_i \in D_i)$$

Dies läßt sich wiederum schreiben als:

$$P\left(\bigcap_{i=1}^{n} X_i^{-1}(D_i)\right) = \prod_{i=1}^{n} P(X_i^{-1}(D_i))$$

Eine (beliebige) Familie $(X_i)_{i \in I}$ von Zufallsvariablen heißen stochastisch unabhängig, falls $(X_j)_{j \in J}$ stochastisch unabhängig ist für jedes endliche $J \subseteq I$. Stochastisch Unabhängig heißt die Möglichkeit, Wahrscheinlichkeiten, die sich durch Zusammenwirken verschiedener Zufallsgrößen ergeben, einfach zu errechnen.

BEISPIEL: $X_1, X_2 \colon \Omega \to \mathbb{Z}$ seien stochastisch unabhängig. **Frage:** Was ist die Verteilung von $X = X_1 + X_2$? Zu berechnen ist

$$\begin{aligned}
P(X_1 + X_2 = k) &= P\left(\sum_{j\in\mathbb{Z}} \{X_1 + X_2 = k \mid X_1 = j\}\right) \\
&= \sum_{j\in\mathbb{Z}} P\left(X_1 + X_2 = k, X_1 = j\right) \\
&= \sum_{j\in\mathbb{Z}} P\left(X_2 = k - j, X_1 = j\right) \\
&= \sum_{j\in\mathbb{Z}} P\left(X_2 = k - j\right) \cdot P\left(X_1 = j\right)
\end{aligned}$$

Ohne stochastische Unabhängigkeit wird dies zu: $\sum_{j\in\mathbb{Z}} P\left(X_2 = k - j | X_1 = j\right) \cdot P\left(X_1 = j\right)$.

10.1 Faltung, Faltungsstabilität

10.1.1 Binomial-Verteilung

Beispiel: Man betrachte zwei Urnen mit schwarzen und weißen Kugeln, aus beiden wird gleichzeitig mit zurücklegen gezogen. Gefragt ist nach der Anzahl $X_1 + X_2$ der gezogenen weißen Kugeln, wobei X_1, X_2 stochastisch unabhängig sind. Sei X_1 $\mathrm{B}(n_1, p_1)$-verteilt, X_2 entsprechend $\mathrm{B}(n_2, p_2)$. Für $k \in \{0, \ldots, n_1 + n_2\}$ ist

$$\begin{aligned}
P(X_1 + X_2 = k) &= \sum_{j=0}^{n_1} P(X_2 = k - j) P(X_1 = j) \\
&= \sum_{j=0}^{n_1} \binom{n_2}{k-j} p_2^{k-j} (1 - p_2)^{n_2 - (k-j)} \binom{n_1}{j} p_1^j (1 - p_1)^{n_1 - j} \\
(\star) \qquad &= p^k (1-p)^{n_1 + n_2 - k} \underbrace{\sum_{j=0}^{n_1} \binom{n_2}{k-j} \binom{n_1}{j}}_{=\binom{n_1+n_2}{k}} \\
&= \binom{n_1 + n_2}{k} p^k (1-p)^{n_1 + n_2 - k}
\end{aligned}$$

Dabei gilt $(\star)$ für $p_1 = p_2 = p$. Also: Falls X_i jeweils $\mathrm{B}(n_i, p)$ verteilt sind und stochastisch unabhängig sind, so ist $\sum_{i=1}^{l} X_i$ $\mathrm{B}(n_1 + \ldots + n_l, p)$-verteilt.

Bemerkung: Sind $X_1, \ldots, X_m$ stochastisch unabhängig, so bezeichnet man die Verteilung von $\sum_{i=1}^{n} X_i$ auch als *Faltung* der Verteilungen von $X_1, \ldots, X_n$.

Startet man mit Verteilungen aus einer gewissen Familie und bleibt die Faltung in dieser Familie, so spricht man von Faltungsstabilität (z.B. sind $(\mathrm{B}(n,p))_{n\in\mathbb{N}}$ ist faltungsstabil).

10.1.2 Poisson-Verteilung

Für X_1, X_2 jeweils $\mathrm{P}(\lambda_i)$-verteilt und stochastisch unabhängig gilt:

$$\begin{aligned} P(X_1+X_2=k) &= \sum_{j=0}^{\infty} P(X_2=k-j)P(X_1=j) \\ &= \sum_{j=0}^{\infty} e^{-\lambda_1}\frac{\lambda_1^{k-j}}{(k-j)!}e^{-\lambda_2}\frac{\lambda_2^j}{j!} \\ &= e^{-(\lambda_1+\lambda_2)}\sum_{j=0}^{k}\frac{\lambda_1^{k-j}}{(k-j)!}\frac{\lambda_2^j}{j!} \\ &= \frac{(\lambda_1+\lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)} \end{aligned}$$

Also ist X_1+X_2 wieder $\mathrm{P}(\lambda_1+\lambda_2)$-verteilt.

10.2 Satz von Fubini

Bei diskreten Zufallsgrößen erhält man die Verteilung durch Summation wie oben beschrieben. **Frage:** Wie geht dieses allgemein? **Vermutung:** Im diskreten Fall gilt

$$\begin{aligned} E(h(X_1,X_2)) &= \sum_{x_1,x_2} h(x_1,x_2)P(X_1=x_1,X_2=x_2) \\ &= \sum_{x_1,x_2} h(x_1,x_2)P(X_1=x_1)\cdot P(X_2=x_2) \\ &= \sum_{x_2} P(X_2=x_2)E(h(X_1,x_2)) \end{aligned}$$

Warum ist die Formel für $P(X_1+X_2=k)$ ein Spezialfall hiervon? Mit $h(X_1,X_2)=X_1+X_2$ ist $Eh(X_1,X_2)=E(X_1+X_2)=E(X_1)+E(X_2)$!

Beachte: $h=1_A$ bedeutet $1_A\circ(X_1,X_2)=1_{(X_1,X_2)^{-1}(A)}$ ergibt $E(h(X_1,X_2))=P((X_1,X_2)\in A)$, d.h.

$$X_1+X_2=k \Leftrightarrow (X_1,X_2)\in A_k=\{\,(x_1,x_2)\mid x_1+x_2=k\}$$

Es ist

$$\begin{aligned}
P((X_1,X_2)\in A) &= E(1_A(X_1,X_2)) \\
&= \sum_{x_2} E(1_A\circ(X_1,x_2))P(X_2=x_2) \\
&= \sum_{x_2} P((X_1,x_2)\in A)P(X_2=x_2) \\
&= \sum_{x_1} P((x_1,X_2)\in A)P(X_1=x_1)
\end{aligned}$$

Beachte dabei

$$P((X_1,x_2)\in A) = P(\{\omega_i \mid (X_1(\omega),x_2)\in A\})$$

Anders geschrieben mit Erwartungswert als Integral gilt:

$$\int h(X_1,X_2)\,dP = \int_{\mathcal{X}_2}\left(\int_{\mathcal{X}_1} h(x_1,x_2)P^{X_1}(dx_1)\right)P^{X_2}(dx_2)$$

Vermutung: Es gilt allgemein:

$$Eh(X_1,X_2) = \int\int h(x_1,x_2)P^{X_1}(dx_1)P^{X_2}(dx_2) = \int\int h(x_1,x_2)P^{X_2}(dx_2)P^{X_1}(dx_1)$$

Das stimmt! Siehe Satz von Fubini. Dies wird „ständig“ in der Wahrscheinlichkeitstheorie angewandt.

Ziel: Den Satz von Fubini

1. korrekt zu formulieren mit geeigneter Meßbarkeit für h
2. und dann zu beweisen

Definition: $(\mathcal{X}_1,\mathcal{C}_1)$ und $(\mathcal{X}_2,\mathcal{C}_2)$ seien meßbare Räume, definiere *Rechteckmengen*

$$\mathcal{R} := \{\, A_1\times A_2 \mid A_1\in\mathcal{C}_1, A_2\in\mathcal{C}_2\}$$

Beachte: $\mathcal{R}$ ist $\cap$-stabil, denn $(A_1\times A_2)\cap(A_1'\times A_2') = (A_1\cap A_1')\times(A_2\times A_2')$. Die Produkt-$\sigma$-Algebra wird definiert als

$$\mathcal{C}_1\otimes\mathcal{C}_2 := \sigma(\mathcal{R})$$

Bemerkung: Seien $X_i\colon\Omega\to\mathcal{X}_i$ meßbar für $i=1,2$. Dann ist $(X_1,X_2)\colon\Omega\to\mathcal{X}_1\times\mathcal{X}_2$ meßbar bezüglich $\mathcal{C}_1\otimes\mathcal{C}_2$, denn: Für die Menge

$$\mathcal{D} = \left\{\, B\subseteq\mathcal{X}_1\times\mathcal{X}_2 \mid (X_1,X_2)^{-1}(B)\in\mathfrak{A}\right\}$$

gilt:

1. $\mathcal{D}$ ist σ-Algebra

2. $\mathcal{R} \subseteq \mathcal{D}$, denn: $(X_1, X_2)^{-1}(A_1 \times A_2) = X_1^{-1}(A_1) \cap X_2^{-1}(A_2)$

Also ist $\mathcal{C}_1 \otimes \mathcal{C}_2 = \sigma(\mathcal{R}) \subseteq \mathcal{D}$.

Wir können präzisieren: Wie berechnen wir $P((X_1, X_2) \in B)$ für $B \in \mathcal{C}_1 \otimes \mathcal{C}_2$? *Allgemeiner:* Wie berechnen wir $Eh(X_1, X_2)$ für $h\colon \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ meßbar bezüglich $\mathcal{C}_1 \otimes \mathcal{C}_2$? Das beinhaltet die Antwort auf die erste Frage, denn

$$P((X_1, X_2) \in B) = E(1_B(X_1, X_2))$$

10.2.1 Satz von Fubini - Formulierung

SATZ: (Satz von FUBINI) $X_i\colon \Omega \to \mathcal{X}_i$ seien stochastisch unabhängige Zufallsvariablen für $i = 1, 2$. Sei $h\colon \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ meßbar (bezüglich $\mathcal{C}_1 \otimes \mathcal{C}_2$), $h(X_1, X_2)$ sei regulär. Dann gilt:

$$\begin{aligned} Eh(X_1, X_2) &= \int_{\mathcal{X}_1 \times \mathcal{X}2} h(x_1, x_2) P^{(X_1, X_2)}(d(x_1, x_2)) \\ &= \int_{\mathcal{X}_1} \int_{\mathcal{X}_2} h(x_1, x_2) P^{X_2}(d(x_2)) P^{X_1}(d(x_1)) \\ &= \int_{\mathcal{X}_2} \int_{\mathcal{X}_1} h(x_1, x_2) P^{X_1}(d(x_1)) P^{X_2}(d(x_2)) \end{aligned}$$

KOROLLAR: $X_i\colon \Omega \to \mathcal{X}_i$ seien Zufallsvariablen für $i = 1, 2$. Dann gilt für jedes $B \in \mathcal{C}_1 \otimes \mathcal{C}_2$:

$$\begin{aligned} P((X_1, X_2) \in B) &= \int_{\mathcal{X}_2} P^{X_1}(\{x_1 \mid (x_1, x_2) \in B\}) P^{X_2}(dx_2) \\ &= \int_{\mathcal{X}_1} P^{X_2}(\{x_2 \mid (x_1, x_2) \in B\}) P^{X_1}(dx_1) \\ &= \int_{\mathcal{X}_2} P((X_1, x_2) \in B) P^{X_2}(dx_2) \\ &= \int_{\mathcal{X}_1} P((x_1, X_2) \in B) P^{X_1}(dx_1) \end{aligned}$$

Faustregel: Ersetze ein großes X_i durch ein kleines x_i und integriere dann bezüglich P^{X_i}.

10.2.2 Summe zweier unabhängiger Zufallsgrößen

Seien $X_i \colon \Omega \to \mathbb{R}$ stochastisch unabhängige Zufallsgrößen für $i = 1, 2$. Dann gilt gemäß dem Korollar:

$$\begin{aligned} P(X_1 + X_2 \in B) &= \int P(X_1 + x_2 \in B) P^{X_2}(dx_2) \\ &= \int P(X_1 \in B - x_2) P^{X_2}(dx_2) \end{aligned}$$

Also ist

$$\begin{aligned} P(X_1 + X_2 \leq t) &= \int_{\mathbb{R}} P(X_2 \leq t - x_1) P^{X_1}(dx_1) \\ &= \int_{\mathbb{R}} P(X_1 \leq t - x_2) P^{X_2}(dx_2) \end{aligned}$$

Bei Vorliegen einer stetigen Dichte f_i für $i = 1, 2$ ist weiter:

$$\begin{aligned} P(X_1 + X_2 \leq t) &= \int_{-\infty}^{\infty} \int_{-\infty}^{t - x_2} f_1(x_1)\, dx_1 f_2(x_2)\, dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{t} f_1(x_1 - x_2)\, dx_1 f_2(x_2)\, dx_2 \\ &= \int_{-\infty}^{t} \underbrace{\int_{-\infty}^{\infty} f_1(z - x_2) f_2(x_2)\, dx_2}_{h(z)}\, dz \end{aligned}$$

Damit ist h die Dichte von $X_1 + X_2$:

$$h(z) = \int_{-\infty}^{\infty} f_1(z - x_2) f_2(x_2)\, dx_2$$

Beispiel: Seien X_1, X_2 stochastisch unabhängig und $\mathrm{N}(a, \sigma^2)$- bzw. $\mathrm{N}(b, \tau^2)$-verteilt. Was ist die Verteilung von $X_1 + X_2$? Es gilt:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-x-a)^2}{\sigma^2}} \cdot \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{(x-b)^2}{\tau^2}}\, dx = \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} e^{-\frac{(z-(a+b))^2}{2(\sigma^2+\tau^2)}}$$

Damit ist $X_1 + X_2$ normalverteilt mit $\mathrm{N}(a + b, \sigma^2 + \tau^2)$

10.2.3 Satz von Fubini - Beweis

Beweis der Aussage zunächst für Indikatorfunktion $h = 1_B$. Definiere dazu:

$$\mathcal{D} = \left\{ B \subseteq \mathcal{X}_1 \times \mathcal{X}_2 \;\middle|\; \begin{array}{l} 1_B(x_1, \cdot) \colon \mathcal{X}_2 \to \mathbb{R} \text{ meßbar} \\ \int 1_B(\cdot, x_2) P^{X_2}(dx_2) \colon \mathcal{X}_1 \to \mathbb{R} \text{ meßbar} \\ E 1_B(X_1, X_2) = \int_{\mathcal{X}_1} \int_{\mathcal{X}_2} 1_B(x_1, x_2) P^{X_2}(dx_2) P^{X_1}(dx_1) \end{array} \right\}$$

Gewünscht ist: $\mathcal{C}_1 \otimes \mathcal{C}_2 \subseteq \mathcal{D}$!

1. $\mathcal{D}$ ist offensichtlich Dynkin-System

2. $\mathcal{R} \subseteq \mathcal{D}$ ebenfalls offensichtlich: Sei $B = A_1 \times A_2$, also

$$1_B(x_1, x_2) = 1_{A_1 \times A_2}(x_1, x_2) = 1_{A_1}(x_1) 1_{A_2}(x_2)$$

   Damit ist

$$1_B(x_1, \cdot) = 1_{A_1}(x_1) 1_{A_2} = \begin{cases} 0 & \text{falls} \quad x_1 \notin A_1 \\ 1_{A_2} & \text{falls} \quad x_1 \in A_1 \end{cases}$$

   Weiter gilt:

$$\int 1_B(x_1, x_2) P^{X_2}(dx_2) = \begin{cases} 0 & \text{falls} \quad x_1 \notin A_1 \\ P(X_2 \in A_2) & \text{falls} \quad x_1 \in A_1 \end{cases}$$

   Und es gilt auch:

$$\begin{array}{rcl} & & \displaystyle\int_{\mathcal{X}_1} \int_{\mathcal{X}_2} 1_B(x_1, x_2) P^{X_2}(dx_2) P^{X_1}(dx_1) \\ & = & \displaystyle\int_{X_1} P(X_2 \in A_2) 1_{A_1} \, dP^{X_1} \\ & = & P(X_2 \in A_2) P(X_1 \in A_1) \\ \text{Unabh.:} & = & P(X_1 \in A_1, X_2 \in A_2) \\ & = & P((X_1, X_2) \in B) \\ & = & E 1_B(X_1, X_2) \end{array}$$

Mit dem Satz von Dynkin folgt aus den Eigenschaften: $\mathcal{C}_1 \otimes \mathcal{C}_2 = \sigma(\mathcal{R}) \subseteq \mathcal{D}$. Damit ist der Satz von FUBINI gültig für alle h der Vorm $h = 1_B$ mit $B \in \mathcal{C}_1 \otimes \mathcal{C}_2$. Mit der üblichen Vorgehensweise erhält man unter Benutzung des Satzes von der monotonen Konvergenz, daß der Satz von FUBINI gültig ist für jedes meßbare $h \geq 0$

10.2.4 Die Varianz der Summen von Zufallsgrößen

Seien $X_1, X_2, \ldots, X_n$ Zufallsgrößen mit endlichen Varianzen. Dann gilt:

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^n X_i\right) &= E\left(\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)^2\right) \\
&= E\left(\left(\sum_{i=1}^n (X_i - E(X_i))\right)^2\right) \\
&= E\left(\sum_{i=1}^n (X_i - E(X_i))^2\right) + E\left(\sum_{i\neq j} (X_i - E(X_i))(X_j - E(X_j))\right) \\
&= \sum_{i=1}^n E\left((X_i - E(X_i))^2\right) + \sum_{i\neq j} E\left((X_i - E(X_i))(X_j - E(X_j))\right) \\
&= \sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i\neq j} \operatorname{Kov}(X_i, X_j)
\end{aligned}$$

Dabei gilt:

DEFINITION: Die *Kovarianz* von X_i und X_j ist

$$\operatorname{Kov}(X_i, X_j) = E\left((X_i - E(X_i))(X_j - E(X_j))\right)$$

Der Korrelationskoeffizient von X_i und X_j ist

$$\operatorname{Korr}(X_i, X_j) = \frac{\operatorname{Kov}(X_i, X_j)}{\sqrt{\operatorname{Var}(X_i)}\sqrt{\operatorname{Var}(X_j)}}$$

Dabei gilt: $|\operatorname{Korr}(X_i, X_j)| \leq 1$ nach der Cauchy-Schwarzschen-Ungleichung[15]. Die Korrelationskoeffizient ist eine Maßzahl für die gegenseitige Beeinflussung von X_i und X_j. Es gilt

$$\begin{aligned}
\operatorname{Kov}(X_i, X_j) &= E\left((X_i - E(X_i))(X_j - E(X_j))\right) \\
&= E\left(X_i X_j - E(X_i)X_j - X_i E(X_j) + E(X_i)E(X_j)\right) \\
&= E(X_i X_j) - E(X_i)E(X_j) - E(X_i)E(X_j) + E(X_i)E(X_j) \\
&= E(X_i X_j) - E(X_i)E(X_j)
\end{aligned}$$

[15] wende die Ungleichung an auf $Z_i = X_i - E(X_i)$

Für X_1, X_2 stochastisch unabhängig und integrierbar ist X_1X_2 integrierbar, und es gilt:

$$E(X_1X_2) = E(X_1)E(X_2) \text{ und } \operatorname{Kov}(X_1, X_2) = 0$$

Beweis: Es gilt:

$$\begin{aligned}
E(|X_1X_2|) &= \int\int |x_1|\,|x_2|\, P^{X_2}(dx_2)P^{X_1}(dx_1) \\
&= \int |x_1| \int |x_2|\, P^{X_2}(dx_2)P^{X_1}(dx_1) \\
&= \int |x_1|\, E(|X_2|)P^{X_1}(dx_1) \\
&= E(|X_2|)E(|X_1|)
\end{aligned}$$

Also ist X_1X_2 integrierbar, damit läßt sich der Satz von Fubini auf X_1X_2 anwenden und liefert entsprechend $E(X_1X_2) = E(X_1)E(X_2)$.

Folgerung: Sind $X_1, \ldots, X_n$ stochastisch unabhängig mit endlichen Varianzen, dann gilt:

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i)$$

11 Gesetze der großen Zahlen

11.1 Motivation zum Gesetz der großen Zahlen

Daraus läßt sich etwas folgern, was unseren Erfahrungsschatz innerhalb des Begriffsapparats der Wahrscheinlichtkeitstheorie abbildet. Dieses „Etwas“ wird als *Gesetz der großen Zahlen* bezeichnet:

Es wird gewürfelt: einmal, zweimal, ..., n-mal, ... und der relative Anteil der 1 unter den ersten n Würfen wird registriert. Die Erfahrung besagt: dieser relative Anteil sollte nahe bei $\frac{1}{6}$ liegen. Wahrscheinlichkeitstheoretische Modellierung gegeben durch stochastisch unabhängige $Y_1, \ldots$ mit Y_i als Ergebnis des i-ten Wurfes. Relativer Anteil der 1 ist gegeben durch

$$\frac{|\{i \leq n \mid Y_i = 1\}|}{n}$$

Vermutung für jedes $\varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{|\{i \leq n \mid Y_i = 1\}|}{n} - \frac{1}{6}\right| \geq \varepsilon\right) = 0$$

Sei nun $X_i = 1_{\{1\}}(Y_i)$ (d.h. nur gleich 1, falls $Y_i = 1$ ist, sonst 0), also ist $|\{ i \le n \mid Y_i = 1\}| = \sum_{i=1}^n X_i$.

11.1.1 Versuchswiderholungen und relative Häufigkeiten

Seien $Y_1, Y_2, \ldots$ Zufallsvariablen mit Werten in $\mathcal{Y}$. Für $A \subseteq \mathcal{Y}$ sei

$$h(A,n) = \frac{|\{ i \le n \mid Y_i \in A\}|}{n} = \frac{1}{n}\sum_{i=1}^n 1_A(Y_i)$$

Wir sprechen von einer *Versuchswiederholung*, falls $Y_1, Y_2, \ldots$ stochastisch unabhängig sind und $P^{Y_i} = P^{Y_1} = W$ für alle i gilt (beim Würfelwurf: $W(\{l\}) = \frac{1}{6}$ für $l = 1, \ldots, 6$). Beachte:

$$1_A(Y_i) = \begin{cases} 1 & \text{falls} \quad Y_i \in A \\ 0 & \text{falls} \quad Y_i \notin A \end{cases}$$

, also ist für $X_i = 1_A(Y_i)$ (mit $X_i^2 = X_i$):

$$\begin{aligned} E(X_i) &= P(Y_i \in A) = W(A) \\ \operatorname{Var}(X_i) &= E(X_i^2) - (E(X_i)^2) = W(A)(1 - W(A)) \end{aligned}$$

Damit folgt:

$$\begin{aligned} E(|\{ i \le n \mid Y_i \in A\}|) &= E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = nW(A) \\ \operatorname{Var}(|\{ i \le n \mid Y_i \in A\}|) &= \operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var}(X_i) \\ &= nW(A)(1 - W(A)) \le \frac{n}{4} \end{aligned}$$

Damit ist $E(h(A,n)) = W(A)$ und $\operatorname{Var}(h(A,n)) = \frac{1}{n}W(A)(1 - W(A))$.

Exkurs: Seien nun $X_1, \ldots, X_n$ nicht notwendig stochastisch unabhängig mit $EX_i = a$ und $\operatorname{Var}(X_i) = b > 0$ und $\operatorname{Kov}(X_i, X_j) = c \neq 0$, dann ist

$$\begin{aligned} \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) &= \frac{1}{n^2}\left(\sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{i \neq j} \operatorname{Kov}(X_i, X_j)\right) \\ &= \frac{b}{n} + \frac{n(n-1)}{n^2}c = \frac{b + (n-1)c}{n} \end{aligned}$$

Es folgt mit Чебышев-Ungleichung:

$$P(|h(A,n) - W(A)| \geq \varepsilon) \leq \frac{\operatorname{Var} h(A,n)}{\varepsilon^2} = \frac{1}{n} \frac{W(A)(1-W(A))}{\varepsilon^2} \longrightarrow 0$$

Satz: Seien $Y_1, Y_2, \dots$ stochastisch unabhängig mit $P^{Y_i} = P^{Y_1} = W$ für alle i. Dann gilt:

$$\lim_{n\to\infty} P(|h(A,n) - W(A)| \geq \varepsilon) = 0 \;\; \forall\, \varepsilon > 0$$

Beweis: Wie oben mit $E(h(A,n)) = W(A)$ und $\operatorname{Var}(h(A,n)) = \frac{W(A)(1-W(A))}{n}$ sowie der Чебышев-Ungleichung.

11.2 Schwaches Gesetz der großen Zahlen

Satz: Seien $X_1, X_2, \dots$ stochastisch unabhängige Zufallsgrößen mit endlichen Varianzen, für die gelte: $\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^n \operatorname{Var}(X_i) = 0$. Dann folgt

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i) \right| \geq \varepsilon \right) = 0 \;\; \forall\, \varepsilon > 0$$

Beweis: Es gilt mit Чебышев-Ungleichung[16]:

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i) \right| \geq \varepsilon \right) \leq \frac{1}{n^2} \sum_{i=1}^n \operatorname{Var}(X_r)$$

Die im schwachen Gesetz der großen Zahlen auftretende Konvergenzart wird als *schwache Konvergenz* oder *Konvergenz in Wahrscheinlichkeit* bezeichnet:

Definition: Sind $Z, Z_1, Z_2, \dots$ Zufallsgrößen, so wird definiert: Z_n *konvergiert schwach* oder *konvergiert in Wahrscheinlichkeit* gegen Z, wenn

$$\lim_{n\to\infty} P(|Z_n - Z| \geq \varepsilon) = 0 \;\; \forall\, \varepsilon > 0$$

Das schwache Gesetz der großen Zahlen besagt also:

$$\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i) \longrightarrow 0 \;\; \text{in Wahrscheinlichkeit}$$

[16] „Warum heißt es *schwaches* Gesetz? Weil der Beweis so einfach ist, daß schon ein *schwacher* Mathematiker es beweisen kann!“

Und für relative Häufigkeiten bei Versuchswiederholungen:

$$h(A,n)) \longrightarrow W(A) \text{ in Wahrscheinlichkeit}$$

11.3 Fast sichere Konvergenz

Schwache Gesetze in der Wahrscheinlichkeitstheorie behandeln *schwache* Konvergenz, d.h. Konvergenz in Wahrscheinlichkeit, *starke* Gesetze behandeln eine stärkere Konvergenzart, die sogenannte *fast sichere Konvergenz*.

$Z, Z_1, Z_2, \ldots$ seien Zufallsgrößen. Dann wird definiert: Z_n *konvergiert fast sicher*, falls $P(\{\omega \mid \lim Z_n(\omega) = Z(\omega)\}) = 1$.

Danke an unendliches Würfeln $\Omega = \{1, \ldots, 6\}^{\mathbb{N}}$, sei $h(\{1\}, n)(\omega) = |\{i \leq n \mid \omega_i = 1\}|$, Frage: Gilt

$$P\left(\left\{\omega \mid \lim_{n\to\infty} h(\{1\}, n)(\omega) = \frac{1}{6}\right\}\right) = 1?$$

11.3.1 Umformulierung der fast sicheren Konvergenz

$$\begin{aligned}
\lim_{n\to\infty} Z_n(\omega) = Z(\omega) &\iff \forall\, l \in \mathbb{N}\colon \exists\, n_l\colon \forall\, m \geq n_l\colon\ |Z_m(\omega) - Z(\omega)| \leq \frac{1}{l} \\
&\iff \omega \in \bigcap_{l\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \left\{\omega' \;\middle|\; |Z_m(\omega') - Z(\omega')| \leq \frac{1}{l}\right\}
\end{aligned}$$

Zum Nachweis von $Z_n \longrightarrow Z$ fast sicher, d.h. zum Nachweis von

$$P(\{\omega \mid \lim Z_n(\omega) = Z(\omega)\}) = 1$$

ist zu zeigen:

$$P\left(\bigcap_{l\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \left\{\omega' \;\middle|\; |Z_m(\omega') - Z(\omega')| \leq \frac{1}{l}\right\}\right) = 1$$

Es ist aber

$$\begin{aligned}
& P\left(\bigcap_{l\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \left\{\omega' \;\middle|\; |Z_m(\omega') - Z(\omega')| \leq \frac{1}{l}\right\}\right) \\
= \ & \lim_{l\to\infty} P\left(\bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \left\{\omega' \;\middle|\; |Z_m(\omega') - Z(\omega')| \leq \frac{1}{l}\right\}\right)
\end{aligned}$$

Somit ist

$$P\left(\bigcap_{l\in\mathbb{N}}[\ldots]\right) = 1 \Longleftrightarrow P([\ldots]) = 1 \Longleftrightarrow P([\ldots]^c) = 0$$

Es gilt also: Z_n konvergiert fast sicher gegen Z genau dann, wenn für alle $\varepsilon > 0$ gilt:

$$P\left(\left(\bigcup_{n\in\mathbb{N}}\bigcap_{m\geq n}\{\omega' \mid |Z_m(\omega') - Z(\omega')| \leq \varepsilon\}\right)^c\right) = 0$$

D.h. es ist für jedes $\varepsilon > 0$:

$$\begin{aligned} 0 &= P\left(\bigcap_{n\in\mathbb{N}}\bigcup_{m\geq n}\{\omega' \mid |Z_m(\omega') - Z(\omega')| > \varepsilon\}\right) \\ &= \lim_{n\to\infty} P\left(\bigcup_{m\geq n}\{\omega' \mid |Z_m(\omega') - Z(\omega')| > \varepsilon\}\right) \\ &= P\left(\limsup\{\omega' \mid |Z_m(\omega') - Z(\omega')| > \varepsilon\}\right) \end{aligned}$$

Dies wird benutzt, um die fast sichere Konvergenz zu untersuchen.

11.3.2 Fast sichere und schwache Konvergenz

Satz: Seien $Z, Z_1, Z_2, \ldots$ Zufallsgrößen Dann gilt:

1. Wenn $Z_n \longrightarrow Z$ fast sicher, so gilt $Z_n \longrightarrow Z$ in Wahrscheinlichkeit.
2. Falls $\sum_{m\in\mathbb{N}} P(|Z_m - Z| > \varepsilon) < \infty$ für alle $\varepsilon > 0$, so ist $Z_n \longrightarrow Z$ fast sicher

Beweis:

1. Falls $Z_n \longrightarrow Z$ fast sicher, so ist für alle $\varepsilon > 0$: $\lim_{m\to\infty} P(|Z_m - Z| > \varepsilon) = 0$. Nun folgt die Aussage aus obiger Umformulierung der fast sicheren Konvergenz wegen

$$P(|Z_m - Z| > \varepsilon) \leq P\left(\bigcup_{m'\geq m}\{|Z_{m'} - Z| > \varepsilon\}\right)$$

2. Mit dem Borel-Cantelli-Lemma ($\sum_n P(A_n) < \infty \Rightarrow P(\limsup A_n) = 0$) gilt:

$$\sum_{m\in\mathbb{N}} P(|Z_m - Z| > \varepsilon) < \infty \Longrightarrow P\left(\limsup\{|Z_m(\omega') - Z(\omega')| > \varepsilon\}\right) = 0$$

Dies ist wieder eine Umformulierung der fast sicheren Konvergenz von oben.

11.4 Starkes Gesetz der großen Zahlen

11.4.1 Versuch, zu einem Starken Gesetz zu gelangen

Seien $X_1, X_2, \ldots$ stochastisch unabhängige Zufallsgrößen mit $E(X_i^2) < \infty$. Dann ist

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))$$

Gehe über zu $X_i' = X_i - E(X_i)$, wobei $E(X_i') = 0$ und $\mathrm{Var}(X_i') = \mathrm{Var}(X_i)$, bzw. nehme gleich o.B.d.A. an, daß $E(X_i) = 0$ vorliegt.

Erster Versuch zum Nachweis von $\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow 0$ fast sicher: Sei $\varepsilon > 0$, betrachte mit der Annahme $\mathrm{Var}(X_1) = \mathrm{Var}(X_i) = b > 0$

$$\begin{aligned}
P\left(\bigcup_{m\geq n}\left\{\left|\frac{1}{m}\sum_{i=1}^{m} X_i\right| \geq \varepsilon\right\}\right) &\leq \sum_{m=n}^{\infty} P\left(\left\{\left|\frac{1}{m}\sum_{i=1}^{m} X_i\right| \geq \varepsilon\right\}\right) \\
&\leq \sum_{m=n}^{\infty} \frac{1}{m^2\varepsilon^2}\sum_{i=1}^{m}\mathrm{Var}(X_i) \\
&\leq \sum_{m=n}^{\infty} \frac{mb}{m^2\varepsilon^2} \\
&\leq \frac{b}{\varepsilon^2}\sum_{m=n}^{\infty}\frac{1}{m} = \infty
\end{aligned}$$

Das klappt so nicht! **Zweiter Versuch:** Beachte $E(X_i) = 0$, daher $\mathrm{Var}(X_i) = E(X_i^2)$; es ist:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \varepsilon\right) \leq \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n} E(X_i^2) = \frac{b}{\varepsilon \cdot n}$$

Möglichkeit, um die Summierung über $\frac{1}{n}$ zu vermeiden: Denke an

$$P(|Z| \geq \varepsilon) \leq \frac{E(|Z|^p)}{\varepsilon^p}$$

Versuch: $p = 3$, dann gilt:

$$\begin{aligned}
P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \varepsilon\right) &\leq \frac{1}{n^3\varepsilon^3}\sum_{i=1}^{n} E(|X_i|^3) \\
&\leq \frac{1}{n^3\varepsilon^3} E\left(\left(\sum_{i=1}^{n} |X_i|\right)^3\right) \\
\text{dabei ist } \left(\sum_{i=1}^{n} |X_i|\right)^3 &= \sum_{i,j,k} |X_i|\,|X_j|\,|X_k| \\
&= \sum_{i=1}^{n} |X_i|^3 + \sum_{i\neq j} |X_i|^2\,|X_j| + \sum_{i\neq j\neq k} |X_i|\,|X_j|\,|X_k|
\end{aligned}$$

Auch das ist unschön, betrachte deswegen eine höhere Potenz, die die Beträge überflüssig macht:

Dritter Versuch: Setze $p = 4$, es gilt:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \varepsilon\right) \leq \frac{1}{n^4\varepsilon^4} E(\left(\sum_{i=1}^{n} X_i\right)^4)$$

Dabei gilt:

$$\begin{aligned}
& E\left(\left(\sum_{i=1}^{n} X_i\right)^4\right) = E\left(\sum_{i,j,k,l} X_i X_j X_k X_l\right) \\
= \;& E\left(\sum_{i=1}^{n} X_i^4 + \sum_{i\neq j} X_i^3 X_j + \sum_{i\neq j} X_i^2 X_j^2 + \sum_{i\neq j\neq k} X_i^2 X_j X_k + \sum_{i\neq j\neq k\neq l} X_i X_j X_k X_l\right) \\
= \;& \sum_{i=1}^{n} E(X_i^4) + \sum_{i\neq j} E(X_i^2)E(X_j^2)
\end{aligned}$$

11.4.2 Starkes Gesetz der großen Zahlen

SATZ: Seien $X_1, X_2, \ldots$ stochastisch unabhängige Zufallsgrößen mit $E(X_i) = 0$ und $\sup_{i\in\mathbb{N}} E(X_i^4) = b < \infty$. Dann folgt:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = 0 \text{ fast sicher}$$

Beweis: Beachte zunächst mit der Jensenschen Ungleichung: $(E(X_i^2))^2 \leq E(X_i^4)$. Es gilt:

$$\begin{aligned}
P\left(\bigcup_{m\geq n}\left\{\left|\frac{1}{m}\sum_{i=1}^m X_i\right| \geq \varepsilon\right\}\right) &\leq \sum_{m=n}^{\infty} P\left(\left|\frac{1}{m}\sum_{i=1}^m X_i\right| \geq \varepsilon\right) \\
&\leq \sum_{m=n}^{\infty}\frac{1}{m^4\varepsilon^4}E\left(\left(\sum_{i=1}^m X_i\right)^4\right) \\
\text{siehe dritter Versuch} &\leq \sum_{m=n}^{\infty}\frac{1}{m^4\varepsilon^4}\left(\sum_{i=1}^m E(X_i^4) + \sum_{i\neq j} E(X_i^2)E(X_j^2)\right) \\
&\leq \sum_{m=n}^{\infty}\frac{1}{m^4\varepsilon^4}\left(mb + m(m-1)b\right) \\
&\leq \frac{2b}{\varepsilon^2}\sum_{m=n}^{\infty}\frac{1}{m^2}
\end{aligned}$$

Dies strebt gegen 0 für $n \longrightarrow \infty$, da $\sum_{m=1}^{\infty}\frac{1}{m^2} < \infty$.

11.4.3 Starkes Gesetz der großen Zahlen für relative Häufigkeiten

Satz: Seien $Y_1, Y_2, \ldots$ stochastisch unabhängige, identisch verteilte Zufallsvariablen (d.h. $P^{Y_i} = P^{Y_1} = W$) mit $Y_i \colon \Omega \to \mathcal{Y}$. Dann gilt für jedes Ereignis $A \subseteq \mathcal{Y}$:

$$\lim_{n\to\infty} h(A,n) = W(A) \text{ fast sicher}$$

Beweis: Es gilt mit $X_i = 1_A(Y_i)$ (und damit $E(X_i^4) \leq 1$): $h(A,n) = \frac{1}{n}\sum_{i=1}^n X_i$, also $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n (X_i - W(A)) = 0$ fast sicher; und damit $\lim_{n\to\infty} h(A,n) = W(A)$.

11.5 Empirische Verteilungsfunktion

In der Statistik hat man häufig folgende Situation vorliegen: Beobachtet werden Zuvallsvariablen $Y_1, Y_2, \ldots$, die stochastisch unabhängig sind und jeweils die Verteilung $W = P^{Y_i}$ besitzen mögen. **Situation der Statistik:** W ist nicht oder nur unvollständig bekannt, Aufgabe ist es, aus den Beobachtungen Aussagen über W zu treffen.

Seien $Y_1, Y_2, \ldots$ stochastisch unabhängig und identisch verteilt ($P^{Y_i} = P^{Y_1} = W$), sie mögen Werte in $\mathbb{R}$annehmen. Das Wahrscheinlichkeitsmaß W ist dann eindeutig gegeben durch seine Verteilungsfunktion F mit $F(t) = W((-\infty, t])$. Die empirische Verteilungsfunktion F_n ist definiert durch

$$F_n(t) = h((-\infty, t], n) = \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(Y_i)$$

Als Abbildung ist $F_n\colon \Omega \to \{\text{Verteilungsfunktionen auf } \mathbb{R}\}$

$$F_n(\omega)\colon t \mapsto \frac{1}{n}\sum_{i=1}^{n} 1_{(-\infty,t]}(Y_i(\omega))$$

Das starke Gesetz der großen Zahlen für relative Häufigkeiten liefert: $\lim_{n\to\infty} F_n(t) = \lim_{n\to\infty} h((-\infty, t], n) = W((-\infty, t]) = F(t)$ fast sicher für jedes t.

11.6 Das starke Gesetz der Großen Zahlen von Колмогоров

11.6.1 vereinfachte Form des Lemmas von Kronecker

Lemma: Sei $(a_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}$. Sei $s_n = \sum_{i=1}^{n} \frac{a_i}{i}$ für alle $n \in \mathbb{N}$. Falls $(s_n)_{n\in\mathbb{N}}$ konvergiert, so folgt

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} a_i = 0$$

Beweis: Da $(s_n)_{n\in\mathbb{N}}$ konvergiert, existiert ein $s \in \mathbb{R}$ mit $s = \lim_{n\to\infty} s_n$. Es gilt

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} &= \underbrace{\frac{a_n}{n}}_{s_n - s_{n-1}} \cdot \frac{n}{n} + \underbrace{\frac{a_{n-1}}{n-1}}_{s_{n-1}-s_{n-2}} \cdot \frac{n-1}{n} + \ldots + \frac{a_2}{2}\cdot\frac{2}{n} + \frac{a_1}{1}\cdot\frac{1}{n} \\
&= s_n - \frac{1}{n}(n-(n-1))s_{n-1} - \frac{1}{n}((n-1)-(n-2))s_{n-2} - \frac{1}{n}(2-1)s_1 \\
&= s_n - \frac{1}{n}s_{n-1} - \frac{1}{n}s_{n-2} - \ldots - \frac{1}{n}s_1 \\
&= s_n - \frac{1}{n}\sum_{i=1}^{n-1} s_i
\end{aligned}$$

Nun gilt mit dem Grenzwertsatz von Cauchy[17]:

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} &= \left(\lim_{n\to\infty} s_n\right) - \left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1} s_i\right)\\ &= s - s = 0\end{aligned}$$

11.6.2 Ungleichung von Колмогоров

Satz: Seien $X_1, X_2, \dots, X_n$ stochastisch unabhängige, quadratintegrierbare Zuvallsvariablen mit $E[X_k] = 0, k = 1, \dots, n$. Sei $S_k := \sum_{i=1}^{k} X_i$ für jedes $k = 1, \dots, n$. Dann gilt für jedes $\varepsilon > 0$:

$$P\left(\max\{\,|S_k| \mid k = 1, \dots, n\} \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\sum_{k=1}^{n}\mathrm{Var}(X_k)$$

Beweis: Setze

$$\begin{aligned}A_1 &= \{|S_1| \geq \varepsilon\}\\ A_k &= \{|S_1| < \varepsilon, \dots, |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon\} \quad \text{für } k = 2, \dots, m\end{aligned}$$

Dann sind diese Ereignisse paarweise disjunkt und es gilt

$$\sum_{k=1}^{n} A_k = \{\max\{\,|S_k| \mid k = 1, \dots, n\} \geq \varepsilon\}$$

Beachte nun, dass für jedes k die Zufallsgröße $1_{A_k} S_k$ und $S_n - S_k = \sum_{i=k+1}^{n} X_i$ stochastisch unabhängig sind, denn $1_{A_k} S_k$ hängt nur von $X_1, \dots, X_k$ ab und $S_n - S_k$ nur von $X_{k+1}, \dots, X_n$. Damit gilt für alle k:

$$E[1_{A_k} S_k (S_n - S_k)] = E[1_A S_k]\underbrace{(S_n - S_k)}_{=0} = 0$$

[17] aus $\lim_{n\to\infty} x_n = x$ folgt: $\lim_{n\to\infty}\frac{x_1 + \ldots + x_n}{n} = x$

Folglich gilt (wobei $(\star)$ gilt wegen $\sum_{k=1}^{n} A_k \subseteq \Omega$):

$$\begin{aligned}
\sum_{k=1}^{n} \mathrm{Var}(X_k) &= \mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) \\
&= \mathrm{Var}(S_n) = E(S_n^2) - \underbrace{E(S_n)^2}_{=0} \\
(\star) \quad &\geq E\left(\sum_{k=1}^{n} 1_{A_k} S_n^2\right) = \sum_{k=1}^{n} E\left(1_{A_k} S_n^2\right) \\
&= \sum_{k=1}^{n} E(1_{A_k}(S_k + (S_n - S_k))^2) \\
&\geq \sum_{k=1}^{n} \left(E(1_{A_k} S_k^2) + 2\,\underbrace{E(1_{A_k} S_k (S_n - S_k))}_{=0}\right) \\
&\geq \sum_{k=1}^{n} E(1_{A_k} \varepsilon^2) \\
&= \varepsilon^2 \sum_{k=1}^{n} P(A_k) = \varepsilon^2 P\left(\sum_{k=1}^{n} A_k\right) \\
&= \varepsilon^2 P\left(\max\{|S_k| \mid k = 1, \ldots, n\} \geq \varepsilon\right)
\end{aligned}$$

11.6.3 Konvergenzkriterium

SATZ: Es seien $X_1, X_2, \ldots$ stochastisch unabhängige, quadratintegrierbare Zufallsgrößen mit $E[X_n] = 0$ für $n \in \mathbb{N}$. Sei $S_n = \sum_{i=1}^{n} X_i$ für $n \in \mathbb{N}$. Es gelte $\sum_{n=1}^{\infty} \mathrm{Var}(X_n) < \infty$. Dann folgt:

$$P((S_n)_{n\in\mathbb{N}} \text{ konvergiert}) = 1$$

Einschub: Konvergenzkriterium für reelle Zahlen. Seien $a_1, a_2, \ldots \in \mathbb{R}$. Dann gilt:

$$(a_n)_{n\in\mathbb{N}} \text{ konvergiert in } \mathbb{R} \iff \forall\, j \in \mathbb{N}\; \exists\, m \in \mathbb{N}\colon \;\; \{\,|a_n - a_m| \mid n \geq m\} \leq \frac{1}{j}$$

Als Mengen geschrieben ergibt sich bezüglich Zufallsgrößen $Z_1, Z_2, \ldots$: Die Menge aller ω, für die $(Z_n(\omega))_{n\in\mathbb{N}}$ konvergiert, ist gleich

$$\bigcap_{j\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \left\{\omega \;\middle|\; \sup\{\,|Z_n(\omega) - Z_m(\omega)| \mid n \geq m\} < \frac{1}{j}\right\}$$

Kurz geschrieben:

$$\{(Z_n)_{n\in\mathbb{N}} \text{ konvergiert}\} = \bigcap_{j\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \left\{\sup\{\,|Z_n(\omega) - Z_m(\omega)| \mid n \geq m\} < \frac{1}{j}\right\}$$

Die fast sichere Konvergenz bedeutet:

$$(Z_n)_{n\in\mathbb{N}} \text{ konvergiert fast sicher} \iff P((Z_n)_{n\in\mathbb{N}} \text{ konvergiert}) = 1$$

Also ist

$$\begin{aligned}
P((Z_n)_{n\in\mathbb{N}} \text{ konvergiert}) &= 1 \\
\iff P\left(\bigcap_{j\in\mathbb{N}} \bigcup_{m\in\mathbb{N}} \left\{\sup\{\,|Z_n(\omega) - Z_m(\omega)| \mid n \geq m\} < \frac{1}{j}\right\}\right) &= 1 \\
\iff \lim_{j\to\infty} P\left(\bigcup_{m\in\mathbb{N}} \left\{\sup\{\,|Z_n(\omega) - Z_m(\omega)| \mid n \geq m\} < \frac{1}{j}\right\}\right) &= 1 \\
\iff P\left(\bigcup_{m\in\mathbb{N}} \{\sup\{\,|Z_n(\omega) - Z_m(\omega)| \mid n \geq m\} < \varepsilon\}\right) &= 1\; \forall\, \varepsilon > 0 \\
\iff \lim_{m\to\infty} P\left(\{\sup\{\,|Z_n(\omega) - Z_m(\omega)| \mid n \geq m\} < \varepsilon\}\right) &= 1\; \forall\, \varepsilon > 0 \\
\iff \lim_{m\to\infty} P\left(\{\sup\{\,|Z_n(\omega) - Z_m(\omega)| \mid n \geq m\} \geq \varepsilon\}\right) &= 0\; \forall\, \varepsilon > 0
\end{aligned}$$

Zu zeigen: $(s_n)_{n\in\mathbb{N}}$ konvergiert fast sicher. Seien $\varepsilon > 0$ und $m \in \mathbb{N}$. Mit der Колмогоров-Ungleichung gilt:

$$\begin{aligned}
P\left(\sup_{n\geq m} |S_n - S_m| \geq \varepsilon\right) &= P\left(\sup_{n\geq m+1} \left|\sum_{i=m+1}^{n} X_i\right| \geq \varepsilon\right) \\
&= P\left(\bigcup_{k\geq n}\left\{\sup_{k\geq n\geq m+1} \left|\sum_{i=m+1}^{n} X_i\right| \geq \varepsilon\right\}\right) \\
&= \lim_{k\to\infty} P\left(\sup_{k\geq n\geq m+1} \left|\sum_{i=m+1}^{n} X_i\right| \geq \varepsilon\right) \\
\text{Колмогоров-Ungleichung} \quad &\leq \lim_{k\to\infty} \frac{1}{\varepsilon^2} \sum_{i=m+1}^{k} \operatorname{Var}(X_i) \\
&= \frac{1}{\varepsilon^2} \sum_{i=m+1}^{\infty} \operatorname{Var}(X_i)
\end{aligned}$$

Da $\sum_{i=1}^{\infty} \operatorname{Var}(X_i) < \infty$ ist, folgt: $\lim_{m\to\infty} \sum_{i=m+1}^{\infty} \operatorname{Var}(X_i) = 0$ und damit

$$0 \leq \lim_{m\to\infty} P\left(\sup_{n\geq m} |S_n - S_m| \geq \varepsilon\right) \leq 0$$

Somit konvergiert $(S_n)_{n\in\mathbb{N}}$ fast sicher.

11.6.4 Das starke Gesetz der Großen Zahlen von Колмогоров

Satz: Es seien $X_1, X_2, \ldots$ stochastisch unabhängige quadratintegrierbare Zufallsgrößen. Es gelte das *Kolmogorov-Kriterium*:

$$\sum_{n=1}^{\infty} \frac{\operatorname{Var}(X_n)}{n^2} < \infty$$

Dann folgt:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (X_i - E(X_i)) = 0 \quad \text{fast sicher}$$

Beweis: Ohne Einschränkung sei $E(X_i) = 0$ für $i \in \mathbb{N}$. Setze $Y_i = \frac{X_i}{i}$ für $i \in \mathbb{N}$. Dann gilt

$$\sum_{n=1}^{\infty} \operatorname{Var}(Y_n) = \sum_{n=1}^{\infty} \frac{\operatorname{Var}(X_n)}{n^2} < \infty$$

Setze weiter $S_n = \sum_{i=1}^n Y_i$. Dann gilt mit obigem Konvergenzkriterium:

$$P((S_n)_{n\in\mathbb{N}} \text{ konvergiert}) = 1$$

also $(S_n)_{n\in\mathbb{N}}$ konvergiert fast sicher. Damit $S_n = \sum_{i=1}^n Y_i = \sum_{i=1}^n \frac{X_i}{i}$ und somit folgt mit dem Lemma von Kronecker:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n X_i = 0 \quad \text{fast sicher}$$

Bemerkung: Es seien $X_1, X_2, \ldots$ stochastisch unabhängig integrierbare und identisch verteilte Zufallsgrößen mit $EX_i = \mu$. Dann gilt:

$$\frac{1}{n} \sum_{i=1}^n X_i \longrightarrow \mu \text{ fast sicher}$$

12 Verteilungskonvergenz und der zentrale Grenzwertsatz

12.1 Verteilungskonvergenz

Bisher kennengelernt: $Z, Z_1, Z_2, \ldots$ mit $Z_n \longrightarrow Z$ fast sicher, d.h. $P\left(\lim_{n\to\infty} Z_n = Z\right) = 1$, dann folgt $Z_n \longrightarrow Z$ in Wahrscheinlichkeit (d.h. $P(|Z_n - Z| \geq \varepsilon) \longrightarrow 0$ für jedes $\varepsilon > 0$).

Frage: Kann man aus $Z_n \longrightarrow Z$ in Wahrscheinlichkeit (bzw. fast sicher) Aussagen vom Typ $P(Z_n \leq t) \to P(Z \leq t)$ folgern? Nicht unbedingt, falls beispielsweise $Z_n = \frac{1}{n}$ und $Z = 0$ ist, so ist $P(Z_n \leq 0) = 0$, aber $P(Z \leq 0) = 1$! Dies zeigt, daß dieses nicht für jedes t gefolgert werden kann.

Satz: Seien $Z, Z_1, Z_2, \ldots$ Zufallsgrößen mit $Z_n \longrightarrow Z$ in Wahrscheinlichkeit. Dann gilt für alle t mit $P(Z = t) = 0$ (d.h. für alle Stetigkeitspunkte der Verteilungsfunktion):

$$P(Z_n \leq t) \longrightarrow P(Z \leq t)$$

Beweis: Es gilt[18] für ein beliebiges $\varepsilon > 0$:

$$\begin{aligned}
& |P(Z_n \leq t) - P(Z \leq t)| \\
\leq\; & P(Z_n \leq t, Z > t) + P(Z_n > t, Z \leq t) \\
=\; & P(Z_n \leq t, Z > t, |Z_n - z| < \varepsilon) + P(Z_n \leq t, Z > t, |Z_n - Z| \geq \varepsilon) + \\
& P(Z_n > t, Z \leq t, |Z_n - z| < \varepsilon) + P(Z_n > t, Z \leq t, |Z_n - Z| \geq \varepsilon) \\
\leq\; & P(Z_n \leq t, Z > t, |Z_n - Z| < \varepsilon) + P(|Z_n - Z| \geq \varepsilon) + \\
& P(Z_n > t, Z \leq t, |Z_n - Z| < \varepsilon) + P(|Z_n - Z| \geq \varepsilon) \\
\leq\; & P(t < Z < t + \varepsilon) + P(|Z_n - Z| \geq \varepsilon) + \\
& P(t - \varepsilon < Z_n \leq t) + P(|Z_n - Z| \geq \varepsilon)
\end{aligned}$$

Sei $\delta > 0$. Wegen $P(t < Z < t + \varepsilon) \longrightarrow 0$ und $P(z - \varepsilon < Z \leq t) \longrightarrow P(Z = t) = 0$ (beide für $\varepsilon \longrightarrow 0$) existiert ein $\hat{\varepsilon}$ mit $P(t < Z < t + \hat{\varepsilon}) + P(t - \hat{\varepsilon} < Z \leq t) \leq \frac{\delta}{2}$. Wegen $Z_n \longrightarrow Z$ in Wahrscheinlichkeit existiert ein $\hat{n}$ mit $P(|Z_n - Z| \geq \hat{\varepsilon}) \leq \frac{\delta}{2}$ für alle $n \geq \hat{n}$. Also folgt für $n \geq \hat{n}$: $|P(Z_n \leq t) - P(Z \leq t)| \leq \delta$.

Motiviert durch diesen Satz wird daher definiert:

[18] mit $|P(A) - P(B)| = |P(A \cap B) + P(A \cap B^c) - P(A \cap B) - P(B \cap A^c)| \leq P(A \cap B^c) + P(B \cap A^c)$

DEFINITION: Seien $Z, Z_1, Z_2, \ldots$ Zufallsgrößen. Es gilt $Z_n \longrightarrow Z$ *in Verteilung* genau dann, wenn

$$P(Z_n \leq t) \longrightarrow P(Z \leq t) \;\; \forall\, t \text{ mit } P(Z = t) = 0$$

Entsprechend: $W, W_1, W_2, \ldots$ seien Wahrscheinlichkeitsmaße auf $\mathbb{R}$. Dann gilt: $W_n \longrightarrow W$ *in Verteilung* genau dann, wenn

$$W_n((-\infty, t]) \longrightarrow W((-\infty, t]) \;\; \forall\, t \text{ mit } W(\{t\}) = 0$$

SATZ: $W, W_1, W_2, \ldots$ seien Wahrscheinlichkeitsmaße auf $\mathbb{R}$. Dann sind äquivalent:

(i) $W_n \longrightarrow W$ in Verteilung

(ii) $\int g\, dW_n \longrightarrow \int g\, dW$ für jedes stetige beschränkte $g\colon \mathbb{R} \to \mathbb{R}$.

(iii) $\int g\, dW_n \longrightarrow \int g\, dW$ für jedes stetige beschränkte $g\colon \mathbb{R} \to \mathbb{R}$ mit der Eigenschaft, daß $\lim g(x)$ existiert für $x \longrightarrow \pm\infty$

BEWEIS:

(ii) $\Rightarrow$ (iii) klar

(iii) $\Rightarrow$ (i) Für $g = 1_{(-\infty,t]}$ soll gelten (falls $W(\{t\}) = 0$):

$$W_n((-\infty, t]) = \int g\, dW_n \overset{!}{\longrightarrow} \int g\, dW = W((-\infty, t])$$

Die Eigenschaft (iii) ist auf g nicht direkt anwendbar, da g nicht stetig ist, allerdings läßt g einfach durch stetige Funktionen g_k approximieren, auf welche dann (iii) anwendbar ist.

Approximation „von oben": Dann konvertiert g_k gegen $1_{(-\infty,]} = g$. Es gilt: $\int g_k\, dW_n \longrightarrow \int g_k\, dW$ für jedes k mit $n \longrightarrow \infty$; genauso ist $\int g_k\, dW_n \longrightarrow \int g dW_n$ für jedes n mit $k \longrightarrow \infty$. Also folgt $\int g\, dW_n \leq \int g_k\, dW_n \longrightarrow \int g_k\, dW$, damit ist $\limsup_n \int g\, dW_n \leq \int g_k\, dW$ für alle k. Somit gilt insgesamt:

$$\limsup_n \int g\, dW_n \leq \lim_{k \to \infty} \int g_k\, dW = \int g\, dW$$

Entsprechend „von unten": Dann konvergiert g_k gegen $q_{(-\infty,t)} = \hat{g}$. Entsprechend zu eben folgt $\int g\, dW_k \geq \int \hat{g}_k\, dW_n \longrightarrow \int \hat{g}_k\, dW$, also ist

$\liminf_n \int g\,dW_n \geq \int \hat{g}_k\,dW$ für alle k. Somit gilt insgesamt:

$$\begin{aligned}\liminf_n \int g\,dW_n &\geq \lim_{k\to\infty}\int \hat{g}_k\,dW = \int \hat{g}\,dW \\ &= W((-\infty,t)) = W((-\infty,t]) = \int g\,dW\end{aligned}$$

Somit folgt für g:

$$\limsup_n \int g\,dW_n \leq \int g\,dW \leq \liminf_n \int g\,dW_n$$

Daraus folgt

$$\limsup_n \int g\,dW_n = \int g\,dW = \liminf_n \int g\,dW_n$$

(i) $\Rightarrow$ (ii) Wahrscheinlichkeitstheoretische Methode:

1. Suche Zufallsgrößen $Y, Y_1, Y_2, \ldots$ mit $P^Y = W$, $P^{Y_n} = W_n$ Dann gilt: $\int g\,dW = \int g\,dP^Y = Eg(Y)$ und $\int g\,dW_n = \int g\,dP^{Y_n} = Eg(Y_n)$.

   Suche diese derart, daß $Eg(Y_n) \longrightarrow Eg(Y)$ einfach nachzuweisen ist.
2. Kandidaten für $Y, Y_1, Y_2, \ldots$ sind: Starte mit U, die R$(0,1)$-verteilt ist, d.h. $P(U \leq t) = t$ mit $t \in (0,1)$. Seien $F, F_1, F_2, \ldots$ die Verteilungsfunktionen zu $W, W_1, W_2, \ldots$. Dann besitzen $Y = F^{-1}(U)$ und $Y_n = F_n^{-1}(U)$ die gewünschte Eigenschaft.
3. Zu zeigen ist also:

   $$\lim_{n\to\infty} EF_n^{-1}(U) = EF^{-1}(U)$$

   Nach Voraussetzung gilt: $F_n(t) \longrightarrow F(t)$ für alle t mit F stetig in t. Daraus folgt „leicht“: $F_n^{-1}(t) \longrightarrow F^{-1}(t)$ für alle t mit F stetig in t. Sei $Z = \{t \mid F \text{ unstetig in } t\}$. Dann ist Z abzählbar, denn $Z = \{t \mid W(\{t\}) > 0\}$. Sei g stetig und beschränkt. Dann ist

   $$\begin{aligned}\int g\,dW_n &= Eg(F_n^{-1}(U)) \\ &= E(1_{\{U\in Z^c\}} \cdot g(F_n^{-1}(U))) + \underbrace{E(1_{\{U\in Z\}} \cdot g(F_n^{-1}(U)))}_{=0 \text{ wg. } P(U\in Z)=0 \text{ da } Z \text{ abzählbar}}\end{aligned}$$

Es gilt

$$1_{\{U \in Z^c\}} \cdot g(F_n^{-1}(U)) \longrightarrow 1_{\{U \in Z^c\}} g(F^{-1}(U))$$

Damit ist (da g beschränkt ist und somit der Satz der dominanten Konvergenz anwendbar ist):

$$\begin{aligned}
\int g\, dW_n &= Eg(F_n^{-1}(U)) \\
&= E(1_{\{U \in Z^c\}} \cdot g(F_n^{-1}(U))) \\
&\longrightarrow E(1_{\{U \in Z^c\}} g(F^{-1}(U))) \\
&= Eg(F^{-1}(U)) = \int g\, dW
\end{aligned}$$

Achtung! Hier fehlt eine Vorlesung!

$S_n^* = \frac{\sum_{i=1}^n (X_i - E(X_i))}{\sqrt{\sum_{i=1}^n \operatorname{Var}(X_i)}}$
$P^{S_n^*} \longrightarrow N(0,1)$
$W_n \longrightarrow W$: $F_n(t) \longrightarrow F(t)$ für alle t mit F stetig in t genau dann, wenn $\int g\, dW_n \longrightarrow \int g\, dW$ für alle stetigen, beschränkten g
Fouriertranformierte $\varphi_W(t) = \int e^{itx} W(dx)$ $\varphi_X(t) = \int e^{itx} P^X(dx) = Ee^{itX}$

SATZ: Eindeutigkeitssatz: Seien W, W' mit $\varphi_W = \varphi_{W'}$. Dann gilt: $W = W'$.

SATZ: Stetigkeitssatz: Seien $W, W_1, W_2, \ldots$. Dann gilt

$$W_n \longrightarrow W \Longleftrightarrow \varphi_{W_n}(t) \longrightarrow \varphi_W(t)\ \forall\, t$$

BEWEIS: Die Richtung „⇒“ ist nach Definition klar, andere Richtung in mehreren Schritten:

1. $W_n \longrightarrow W$ genau dann, wenn für jede Teilfolge $(n_k)_{k \in \mathbb{N}}$ existiert eine Teilfolge $(m_l)_{l \in \mathbb{N}}$ mit $W_{m_l} \longrightarrow W$

2. *Satz von Helly:* Seien $W_1, W_2, \ldots$ mit der Eigenschaft der *Straffheit*, d.h. Für alle $\varepsilon > 0$ existieren a, b so, daß gilt: $W_n([a,b]) \geq 1 - \varepsilon\ \forall\, n$

   Beweis:

   „⇒“ Es existiert $(n_k)_{k \in \mathbb{N}}$ mit $W_{n_k} \longrightarrow W$. $F_n(t)$ mit $t \in \mathbb{Q}$, $n \in \mathbb{N}$; Diagonalprinzip: Existiert Teilfolge $(n_k)_{k \in \mathbb{N}}$ so, daß $(F_{n_k}(t))_{k \in \mathbb{N}}$ konvergent ist für jedes $t \in \mathbb{Q}$.

Sei $H(t) = \lim_{k \to \infty} F_{n_k}(t)$, $t \in \mathbb{Q}$. Dann gilt:

(a) H monoton wachsend

(b) $\lim_{t\to-\infty} H(t) = 0$ und $\lim_{t\to+\infty} H(t) = 1$

Fortsetzung von H auf $\mathbb{R}$: Definiere

$$F^*\colon \mathbb{R} \to [0,1] \text{ mit } F^*(r) = \inf\{\, F(t) \mid t \in \mathbb{Q}, t > r\}$$

Offensichtlich gilt: F^* ist monoton wachsend, rechtsseitig stetig, es ist $\lim_{r\to-\infty} F^*(r) = 0$ und $\lim_{r\to+\infty} F^*(r) = 1$. Weiter gilt $F^*(t) = H(t)$ mit $t \in \mathbb{Q}$ und zudem $(\star)$ $F_{n_k}(r) \longrightarrow F^*(r)$ für r mit F^* stetig in r.

Nach dem Existenz- und Eindeutigkeitssatz für ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$ existiert genau ein Wahrscheinlichkeitsmaß W^* mit W^* hat Verteilungsfunktion F^*. Gemäß $(\star)$ gilt $W_{n_k} \longrightarrow W^*$.

3. Seien wieder $W, W_1, W_2, \ldots$ gegeben mit $\varphi_n = \varphi_{W_n} \longrightarrow \varphi = \varphi_W$. Zunächst zu zeigen: „Straffheit“ von $(W_n)_{n\in\mathbb{N}}$.

   Betrachte die Funktion $\varphi(t) = \int e^{itx} W(dx)$. Dann ist φ stetig und $\varphi(0) = 1$. Sei nun $\varepsilon > 0$. Dann existiert δ mit

   $$\frac{1}{\delta}\int_{-\delta}^{\delta} |1 - \varphi(t)|\, dt \le \frac{\varepsilon}{2}$$

   Wir schätzen ab: Wegen $\varphi_n \longrightarrow \varphi$ existiert n_0 so, daß für alle $n \ge n_0$ gilt: $\frac{1}{\delta}\int_{-\delta}^{\delta} |1 - \varphi_n(t)|\, dt \le \varepsilon$. Weiter ist:

   $$\begin{aligned} \frac{1}{\delta}\int_{-\delta}^{\delta} (1 - \varphi_n(t))\, dt &= \frac{1}{\delta}\int_{-\delta}^{\delta} \left((\int 1 - e^{itx}) W_n(dx) \right) dt \\ &= \frac{1}{\delta}\int \int_{-\delta}^{\delta} (\int 1 - e^{itx})\, dt\, W_n(dx) \\ &= \int 2(1 - \frac{\sin(\delta x)}{\delta x})\, W_n(dx) \\ (\star) \quad &\ge 2\int_{\{x \mid |\delta x| > 2\}} (1 - \frac{1}{|x\delta|})\, W_n(dx) \\ &\ge W_n(\{x \mid |\delta x| > 2\}) \end{aligned}$$

   Dabei folgt $(\star)$ mit $\frac{\sin r}{r} \le \frac{1}{|r|}$ für $|r| > 2$. Somit gilt insgesamt

   $$W_n\left(\left[-\frac{2}{\delta}, \frac{2}{\delta}\right]\right) \ge 1 - \varepsilon \ \forall\, n \ge n_0$$

   Damit gilt die Straffheit von $(W_n)_{n\ge n_0}$ und somit auch die Straffheit von $(W_n)_{n\in\mathbb{N}}$.

4. Zusammenfügen der ersten drei Schritte: Sei $(W_{n_k})_{k\in\mathbb{N}}$ Teilfolge von $(W_n)_{n\in\mathbb{N}}$. Diese Teilfolge erfüllt gemäßt Schritt 3 die Straffheitsbedingung. Gemäß Schritt 2 existiert W^* und $(W_{m_l})_{l\in\mathbb{N}}$ Teilfolge von $(W_{n_k})_{k\in\mathbb{N}}$ mit $W_{m_l} \longrightarrow W^*$. Gemäß Schritt 1 gilt die Behauptung, falls $W^* = W$ nachgewiesen wird.

   Aus $W_{m_l} \longrightarrow W^*$ folgt $\varphi_{m_l} \longrightarrow \varphi^*$, gemäß Voraussetzung gilt $\varphi_{m_l} \longrightarrow \varphi$, also ist
$$\varphi_W = \varphi = \varphi^* = \varphi_{W^*}$$
   Mit dem Eindeutigkeittsatz folgt $W = W^*$.

   Achtung: *Der Eindeutigkeitssatz folgt aber erst aus dem Satz, den wir hier beweisen wollen. Wir benötigen also noch einen vom Stetigkeitssatz „unabhängigen" Beweis des Eindeutigkeitssatzes! Dieser kann gefolgert werden aus dem Satz von* Stone-Weierstrass *für trigonometrische Polynome oder aus der Umkehrformel von* Levy.

12.1.1 Anwendung auf Summen unabhängiger Zufallsgrößen

1. Seien $X_1, \ldots, X_n$ stochastisch unabhängig. Dann ist
$$\begin{aligned}\varphi_{X_1+\ldots+X_n}(t) &= E\left(e^{it\sum_{i=1}^n X_i}\right) = E\left(\prod_{i=1}^n e^{itX_i}\right)\\ &= \prod_{i=1}^n E(e^{itX_i}) = \prod_{i=1}^n \varphi_{X_i}(t)\end{aligned}$$
   Die liefert eine einfache Möglichkeit, die Verteilung von Summen von Zufallsgrößen zu bestimmen.

2. Betrachte die standadisierte Summe für $X_1, \ldots, X_n$ stochastisch unabhängig, identisch verteilt mit Erwartungswert a und Varianz $\sigma^2 > 0$. Dann ist für Variablen $Y_i = \frac{X_i - a}{\sigma}$ (mit $E(Y_i) = 0$ und $\text{Var}(Y_i) = 1$):
$$S_n^* = \frac{\sum_{i=1}^n (X_i - a)}{\sqrt{n\sigma^2}} = \frac{\sum_{i=1}^n \left(\frac{X_i-a}{\sigma}\right)}{\sqrt{n}} = \frac{\sum_{i=1}^n Y_i}{\sqrt{n}}$$
   Dann ist
$$\varphi_{S_n^*}(t) = Ee^{i\frac{t}{\sqrt{n}}\sum_{i=1}^n Y_i} = \left(\hat{\varphi}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$
   wobei
$$\hat{\varphi}(t) = Ee^{it\left(\frac{X_i-a}{\sigma}\right)} = e^{-\frac{ita}{\sigma}}\varphi\left(\frac{t}{\sigma}\right)$$

12.1.2 Methodik zum Beweis des zentralen Grenzwertsatzes

Nach Stetigkeitssatz wissen wir: Es genügt zu zeigen: $\varphi_{S_n^*} \longrightarrow \varphi_{N(0,1)}$. Dabei ist

$$\varphi_{N(0,1)}(t) = \int e^{itx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = e^{-\frac{t^2}{2}}$$

Damit ist zu zeigen:

$$\left(\hat{\varphi}\left(\frac{t}{\sqrt{n}}\right)\right)^n \longrightarrow e^{-\frac{t^2}{2}} \text{ für alle } t$$

wobei $\hat{\varphi}(t) = E(e^{itY_1})$ mit $EY_1 = 0$ und $\operatorname{Var} Y_1 = 1$. „Kanonisches" Vorgehen: Entwicklung um 0:[19]

$$\left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \varphi'(0)\frac{t}{\sqrt{n}} + \frac{\varphi''(0)}{2}\frac{t^2}{n} + r\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Erinnerung: Es ist $\lim_{n\to\infty}\left(1 + \frac{a}{n} + \frac{b_n}{n}\right)^n = e^a$ für $b_n \longrightarrow 0$. Es verbleibt, eine geeignete Entwicklung von φ zu finden:

$$\frac{d}{dt}\left(E\, e^{itX_1}\right) \overset{(\star)}{=} E\left(\frac{d}{dt} e^{itX_1}\right) = E\left(iX_1 \cdot e^{itX_1}\right)$$

Mit dem Satz von der dominierenden Konvergenz ist auch die Vertauschung von E und $\frac{d}{dt}$ in $(\star)$ möglich. Damit ist $\varphi'(0) = E(iX_1) = 0$. Entsprechend ist

$$\begin{aligned}
\frac{d}{dt}\left(E\left(iX_1 e^{itX_1}\right)\right) &\overset{(\star)}{=} E\left(iX_1 \cdot \frac{d}{dt} e^{itX_1}\right) \\
&= E\left(i^2 X_1^2 \cdot e^{itX_1}\right) \\
&= -E(X^2 e^{itX}) \\
\text{für } t = 0 \quad &= -E(X^2) = -1
\end{aligned}$$

Das Restglied ergibt sich als $r\left(\frac{t}{\sqrt{n}}\right) = \frac{h(t,n)}{n}$ mit $\lim_{n\to\infty} h(t,n) = 0$. Es liegt damit — bei $E(X_1) = 0$ und $\operatorname{Var}(X_1) = 1$ — folgende Entwicklung vor:

$$\left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + \frac{h(t,n)}{n}\right)^n \longrightarrow e^{-\frac{t^2}{2}}$$

Dies ergibt den Beweis des zentralen Grenzwertsatzes im hier betrachteten Fall.

[19] $f(h) = f(0) + f'(0)\cdot h + f''(0)\frac{h^2}{2} + r(h)$

Achtung! Hier fehlt noch das Ende der Vorlesung!