
Stochastik I

TU Dortmund, Sommersemester 2014

Dieses Vorlesungsskript ist aus meiner persönlichen Mitschrift der Vorlesung

Stochastik I

bei Professor Dr. M. Voit im Sommersemester 2014 entstanden. Ich habe versucht alles richtig wiederzugeben, kann allerdings keine Garantie darauf geben. Es ist deshalb wahrscheinlich, dass dieses Skript Fehler enthält.

Dieses Skript darf nur umsonst oder zum Selbstkostenpreis weitergegeben werden. Ich untersage jede kommerzielle Nutzung durch Dritte. Dieses Skript ist weder eine offizielle noch eine von Professor Voit autorisierte Version. Deshalb ist bei Fehlern zuerst davon auszugehen, dass diese von mir stammen.

Alexander Schmalstieg

Inhaltsverzeichnis

1	Wahrscheinlichkeitsräume	1
2	Bedingte Wahrscheinlichkeiten und Unabhängigkeit	15
3	Zufallsvariablen und Verteilungen	26
4	Verteilungen auf \mathbb{R} und \mathbb{R}^d	34
5	Momente \mathbb{R}-wertiger Zufallsvariablen	49
6	Verteilungen von Summen unabhängiger Zufallsvariablen	65
7	Konvergenz von Zufallsvariablen	75
8	Markov-Ketten und stochastische Matrizen	92
9	Statistik: Grundlagen des Schätzens	108
10	Konfidenzintervalle	117
11	Optimale Schätzer (Einführung)	127

1 Wahrscheinlichkeitsräume

1.1 Beispiel: (fairer Würfel) Bei einem fairen Würfel tritt jede der Augenzahlen von $1, \dots, 6$ mit der gleichen Wahrscheinlichkeit $\frac{1}{6}$ auf. Dabei werden folgende Sprechweisen festgelegt:

Einmal Würfeln $\hat{=}$ Durchführung eines *Zufallsexperiments*

$1, \dots, 6 \hat{=}$ mögliche Ereignisse („Elementarereignisse“)

$\Omega := \{1, \dots, 6\} \hat{=}$ Menge der möglichen Ereignisse

$\mathcal{P}(\Omega) := \{A \subseteq \Omega\}$ heißt Potenzmenge von Ω mit der Mächtigkeit $|\mathcal{P}(\Omega)| = 2^{|\Omega|} = 2^6$

Die Elemente von $\mathcal{P}(\Omega)$ (die Teilmengen von Ω) heißen Ereignisse und die Potenzmenge $\mathcal{P}(\Omega)$ heißt (in der Stochastik) auch die Ereignismenge.

Beispiele:

$\{1\}$: Das Ereignis, eine 1 zu würfeln.

$\{1, 3, 5\}$: Das Ereignis, eine ungerade Zahl zu würfeln.

$\{5, 6\}$: Das Ereignis, mindestens eine 5 zu würfeln.

Eine sinnvolle Definition von Wahrscheinlichkeiten von Ereignissen $A \subseteq \Omega$ ergibt sich nun durch $P(A) := \frac{|A|}{|\Omega|} = \frac{|A|}{6}$. (P von engl. Probability)

Beispiel: $A = \{1, 3, 5\}$ Das Ereignis, eine ungerade Zahl zu würfeln, besitzt die Wahrscheinlichkeit $P(A) = \frac{3}{6} = \frac{1}{2}$.

Formal betrachtet ist P eine Abbildung von $\mathcal{P}(\Omega)$ nach $[0, 1]$, mit gewissen Eigenschaften.

1.2 Definition: (Laplace'scher Wahrscheinlichkeitsraum)

Sei $\Omega \neq \emptyset$ eine endliche Menge mit zugehöriger Potenzmenge $\mathcal{P}(\Omega)$. Die Abbildung $\mathcal{P}(\Omega) \rightarrow [0, 1]$ mit $P(A) = \frac{|A|}{|\Omega|}$ heißt Gleichverteilung auf Ω . $(\Omega, \mathcal{P}(\Omega), P)$ heißt Laplace'scher Wahrscheinlichkeitsraum.

1.3 Beispiel: (n -faches Würfel) Es sei ein fairer Würfel und $n \in \mathbb{N}$. Dann ist

$$\Omega = \{1, \dots, 6\}^n = \{(i_1, \dots, i_n) \mid i_k \in \{1, \dots, 6\}, k = 1, \dots, n\}$$

Es ist i_k die Augenzahl des k -ten Würfels für $k = 1, \dots, n$ und $|\Omega| = 6^n$.

Betrachte $n = 2$: Sei $A \subseteq \{1, \dots, 6\}^2$ das Ereignis, bei dem die Summe der Augenzahlen beider Würfel 4 ist. Dann ist $A = \{(1, 3), (2, 2), (3, 1)\}$, $|A| = 3$ und $P(A) = \frac{3}{|\Omega|} = \frac{3}{36} = \frac{1}{12}$. Laplace'sche Wahrscheinlichkeitsräume treten in der Praxis oft auf. Berechnung der Wahrscheinlichkeitsräume geschieht im Prinzip durch Abzählen. Das kann aber beliebig schwierig werden. Suche deshalb Regeln für Wahrscheinlichkeiten und untersuche gewisse Klassen von Standardproblemen.

1.4 Regeln in Laplace'schen Wahrscheinlichkeitsräumen:

a) Es gilt stets $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$ und $P(\phi) = \frac{|\phi|}{|\Omega|} = \frac{0}{|\Omega|} = 0$.

b) Seien $A_1, A_2 \subseteq \Omega$ disjunkt, d.h. $A_1 \cap A_2 = \phi$. Dann gilt

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) \quad (1)$$

denn

$$P(A_1 \cup A_2) = \frac{|A_1 \cup A_2|}{|\Omega|} = \frac{|A_1| + |A_2|}{|\Omega|} = \frac{|A_1|}{|\Omega|} + \frac{|A_2|}{|\Omega|} = P(A_1) + P(A_2).$$

c) Für alle $A \subseteq \Omega$ gilt stets

$$P(\Omega \setminus A) = 1 - P(A) \quad (2)$$

denn

$$1 = P(\Omega) = P((\Omega \setminus A) \cup A) \stackrel{(1)}{=} P(\Omega \setminus A) + P(A) \Leftrightarrow P(\Omega \setminus A) = 1 - P(A).$$

Tipp: Ist $P(A)$ schwer zu berechnen, berechne stattdessen $P(\Omega \setminus A)$, falls das leichter ist.

1.5 Beispiel: (Geburtstagsproblem) In einem Raum befinden sich zufällig n Personen. Wie groß ist die Wahrscheinlichkeit p_n , dass mindestens 2 Personen am gleichen Tag Geburtstag haben? Um diese Frage überhaupt beantworten zu können, machen wir die folgenden Annahmen:

(a) Die Geburtstage der Personen sind stochastisch unabhängig (z.B. es sind keine Zwillinge unter den n Personen).

(b) Alle 365 Tage des Jahres seien gleichwahrscheinlich.

(c) Keine Schaltjahre.

Das zu betrachtende mathematische Modell ist der Laplace'sche Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ mit $\Omega = \{1, \dots, 365\}^n$. Es ist offensichtlich, dass gilt $p_1 = 0$ und $p_n = 1$ für $n \geq 366$. Es bleiben also noch die Fälle für $n = 2, 3, 4, \dots, 365$ zu betrachten. Offenbar gilt

$$\begin{aligned} P_n &= P(\text{„Mindestens zwei Personen haben am gleichen Tag Geburtstag“}) \\ &\stackrel{(2)}{=} 1 - P(\text{„Alle } n \text{ Personen haben verschiedene Geburtstage“}) \\ &= 1 - \frac{365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - n + 1)}{365^n} \end{aligned}$$

Viele Abzählbarkeitsprobleme lassen sich auf 4 grundlegende Urnenmodelle reduzieren.

1.6 Standard-Urnenmodelle:

Eine Urne enthalte N Kugeln mit den Nummern $1, \dots, N$. Man ziehe k -mal je eine Kugel und notiere das Ergebnis. Sei wieder Ω die Menge aller möglichen Ziehergebnisse. Man unterscheidet nun 4 grundlegende Fälle:

- Ziehen mit und ohne Zurücklegen.
- Ziehen mit und ohne Rücksicht auf die Reihenfolge.

Man bestimme nun jeweils Ω und $|\Omega|$.

a) k -maliges Ziehen mit Zurücklegen und mit Rücksicht auf die Reihenfolge.

$$\Omega = \{1, \dots, N\}^k, \quad |\Omega| = N^k \quad (3)$$

b) k -maliges Ziehen ohne Zurücklegen aber mit Rücksicht auf die Reihenfolge. $k \leq N$

$$\Omega = \{(a_1, \dots, a_k) \in \{1, \dots, N\}^k \mid a_i \neq a_j \forall i, j \in \{1, \dots, k\}\}, \quad |\Omega| = N \cdot (N-1) \cdot \dots \cdot (N-k+1) \quad (4)$$

c) k -maliges Ziehen ohne Zurücklegen und ohne Rücksicht auf die Reihenfolge. $k \leq N$

$$\Omega = \{A \subseteq \{1, \dots, N\} \mid |A| = k\} = \{(a_1, \dots, a_k) \in \{1, \dots, N\}^k \mid 1 \leq a_1 < \dots < a_k \leq N\} \quad (5)$$

$$|\Omega| = \binom{N}{k} = \frac{N!}{k! \cdot (N-k)!} \quad (6)$$

Beispiel: Lotto „6 aus 49“. $P(\text{„6 aus 49“}) = \frac{1}{\binom{49}{6}} \approx 7 \cdot 10^{-8}$.

d) k -maliges Ziehen mit Zurücklegen aber ohne Rücksicht auf die Reihenfolge.

$$\Omega = \{(a_1, \dots, a_k) \in \{1, \dots, N\}^k \mid 1 \leq a_1 \leq \dots \leq a_k \leq N\}, \quad |\Omega| = \binom{N+k-1}{k} \quad (7)$$

Beweis von d).

Betrachte die Menge

$$W := \{(y_1, \dots, y_k) \in \{1, \dots, N+k-1\}^k \mid 1 \leq y_1 < \dots < y_k \leq N+k-1\}$$

und die Abbildung $f: \Omega \rightarrow W$ mit

$$f((a_1, \dots, a_k)) := (a_1, a_2 + 1, a_3 + 2, \dots, a_k + k - 1)$$

Diese Abbildung ist offenbar für jedes k -Tupel aus Ω wohldefiniert und die offensichtliche Umkehrabbildung

$$f^{-1}((y_1, \dots, y_k)) = (y_1, y_2 - 1, y_3 - 2, \dots, y_k - k + 1)$$

ist für jedes k -Tupel von W wohldefiniert. Die Abbildung f ist also offenbar bijektiv. Also gilt $|W| = |\Omega| \stackrel{c)}{=} \binom{N+k-1}{k}$. \square

1.7 Bemerkung: (Sieb) Zufälliges Tippen einer reellen Zahl aus dem Intervall $[0, 1]$. Hier ist die Menge der möglichen Ergebnisse $\Omega = [0, 1]$ überabzählbar. Dabei wird jedes $x \in [0, 1]$ mit Wahrscheinlichkeit 0 auftreten. Wie definiert man nun sinnvollerweise die Wahrscheinlichkeit auf einer kontinuierlichen Menge wie $[0, 1]$?

Heuristischer Zugang: Bestimme $P(A)$ für solche $A \subseteq \Omega$, die eine einfache „Bauart“ haben.

a) Für $0 \leq a \leq b \leq 1$ setze

$$P([a, b]) := P([a, b)) := P((a, b]) := P((a, b)) := b - a$$

Dann ergibt sich wie erwartet $P(\{a\}) = P([a, a]) = a - a = 0$.

b) Sei A eine endliche disjunkte Vereinigung von Intervallen $I_1, \dots, I_n \subseteq [0, 1]$, also $A = \bigcup_{k=1}^n I_k$. Setze dann

$$P(A) := \sum_{k=1}^n P(I_k)$$

c) Sei A eine abzählbare disjunkte Vereinigung von Intervallen $I_1, I_2, I_3, \dots \subseteq [0, 1]$, also $A = \bigcup_{k \in \mathbb{N}} I_k$. Definiere nun

$$P(A) := \sum_{k=1}^{\infty} P(I_k) \in [0, 1]$$

Der obige Ausdruck ist wohldefiniert, da die Folge der Partialsummen $\sum_{k=1}^m P(I_k) \leq P([0, 1]) = 1$ durch 1 beschränkt ist und wegen $P(I_j) \in [0, 1]$ für alle $j \in \mathbb{N}$ zusätzlich noch monoton wachsend ist. (Monotone beschränkte Folgen sind konvergent.)¹

Es wird sich später herausstellen, dass für eine allgemeine Menge $\Omega \neq \emptyset$ kein sinnvolles Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$, das einer Gleichverteilung entspricht, definiert werden kann. Deshalb definiert man es nur auf „passenden“ Teilmengensystemen von Ω .

(→ Axiomatisierung)

1.8 Definition: (Algebra, σ -Algebra)

Es seien $\Omega \neq \emptyset$ und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine Menge von Teilmengen von Ω . Das Mengensystem \mathcal{A} heißt Algebra, falls

$$(A1) \quad \Omega \in \mathcal{A},$$

$$(A2) \quad A \in \mathcal{A} \Rightarrow \bar{A} := \Omega \setminus A \in \mathcal{A},$$

$$(A3) \quad A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cup A_2 \in \mathcal{A}.$$

Das Mengensystem \mathcal{A} heißt σ -Algebra, falls

$$(\sigma 1) \quad \mathcal{A} \text{ eine Algebra ist und}$$

$$(\sigma 2) \quad (A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}.$$

¹ Ende der ersten Vorlesung vom Mi, 09.04.14

Beispiele: a) Sei $\Omega \neq \emptyset$. Dann ist $\mathcal{A} := \mathcal{P}(\Omega)$ die größtmögliche σ -Algebra von Ω . Die Menge $\mathcal{A} := \{\Omega, \emptyset\}$ ist die kleinstmögliche σ -Algebra von Ω .

b) Sei $\Omega := [0, 1]$ und $\mathcal{A} := \{A \subseteq [0, 1] \mid A \text{ ist endliche Vereinigung von Intervallen}\}$. Dann ist \mathcal{A} eine Algebra, aber keine σ -Algebra, denn A1 bis A3 sind offenbar erfüllt, nur $\sigma 2$ ist nicht erfüllt. Betrachte dazu $A_n := [\frac{1}{n}, \frac{1}{n}]$ offenbar einfache Vereinigung von Intervallen aber

$$\bigcup_{n \in \mathbb{N}} A_n = \{\frac{1}{n} \mid n \in \mathbb{N}\} \notin \mathcal{A}.$$

c) Sei $\Omega := \{1, 2, 3\}$ und $\mathcal{A} := \{\Omega, \emptyset, \{1\}, \{2, 3\}\}$. Dann ist \mathcal{A} eine σ -Algebra.

1.9 Fakten: Sei $\Omega \neq \emptyset$ eine Menge.

a) $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ ist Algebra $\stackrel{A1, A2}{\Rightarrow} \emptyset \in \mathcal{A}$.

b) $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ ist Algebra, $n \in \mathbb{N}$ und $A_1, \dots, A_n \in \mathcal{A} \Rightarrow \bigcup_{k=1}^n A_k \in \mathcal{A}$. Dies folgt induktiv aus A3.

c) Bei der Definition einer σ -Algebra impliziert $\sigma 2$ bereits A3, denn seien $A_1, A_2 \in \mathcal{A}$. Setze nun für $n \geq 3$ $A_n := \emptyset \in \mathcal{A}$, dann folgt mit $\sigma 2$ sofort

$$\bigcup_{n \in \mathbb{N}} A_n = A_1 \cup A_2 \cup \emptyset \cup \emptyset \cup \dots = A_1 \cup A_2 \in \mathcal{A}.$$

d) Sei nun $|\Omega| < \infty$ und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine Algebra. Dann ist \mathcal{A} auch eine σ -Algebra, denn aus $|\Omega| < \infty$ folgt $|\mathcal{P}(\Omega)| = 2^{|\Omega|} < \infty$ und insbesondere $|\mathcal{A}| < \infty$. Dann treten in der Vereinigung $\bigcup_{n \in \mathbb{N}} A_n$ nur endlich viele Mengen auf, sodass die Behauptung sofort aus b) folgt.

e) \mathcal{A} ist Algebra und $n \in \mathbb{N}$, $A_1, \dots, A_n \in \mathcal{A}$. Dann ist $\bigcap_{k=1}^n A_k \in \mathcal{A}$, denn aus A2 folgt

$$\bigcap_{k=1}^n A_k = \bigcup_{k=1}^n \overline{A_k} \in \mathcal{A}.$$

f) \mathcal{A} ist σ -Algebra und $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A}$. Dann ist $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$, denn analog zu e) folgt aus A2 auch

$$\bigcap_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} \overline{A_n} \in \mathcal{A}.$$

Sprechweisen: Sei $\Omega \neq \emptyset$ eine Menge und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ ist σ -Algebra.

Ω heißt Menge aller möglichen Ereignisse.

Jede Menge $A \in \mathcal{A}$ heißt Ereignis.

$\Omega \in \mathcal{A}$ ist das sichere Ereignis.

$\emptyset \in \mathcal{A}$ ist das unmögliche Ereignis.

1.10 Definition: (Wahrscheinlichkeitsmaß)

Es seien $\Omega \neq \emptyset$ eine Menge und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine σ -Algebra. Eine Abbildung $P : \mathcal{A} \rightarrow [0, 1]$, $A \mapsto P(A)$ heißt Wahrscheinlichkeitsmaß (oder Verteilung) auf (Ω, \mathcal{A}) , falls

$$(W1) \quad P(\Omega) = 1 \text{ und}$$

$$(W2) \quad (A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A} \text{ disjunkt} : P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad (\sigma\text{-Additivitat}).$$

Das Tripel (Ω, \mathcal{A}, P) heit Wahrscheinlichkeitsraum.

Bemerkung: Ist in Definition 1.10 \mathcal{A} nur eine Algebra statt einer σ -Algebra und gilt statt W2 nur die schwachere Bedingung

$$(\widetilde{W2}) \quad n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{A} : P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k),$$

so heit P ein Wahrscheinlichkeitsinhalt (statt -ma).

1.11 Rechenregeln: Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gilt:

a) $P(\emptyset) = 0$, denn $P(\emptyset) = P\left(\bigcup_{n \in \mathbb{N}} \emptyset\right) \stackrel{W2}{=} \sum_{n=1}^{\infty} P(\emptyset)$. Da $P(\emptyset) \in [0, 1]$ per Definition, folgt $P(\emptyset) = 0$.

b) $A_1, \dots, A_n \in \mathcal{A}$ disjunkt $\Rightarrow P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n P(A_k)$, denn setze $A_{n+1} := A_{n+2} = \dots = \emptyset$, dann folgt

$$P\left(\bigcup_{k \in \mathbb{N}} A_k\right) \stackrel{W2}{=} \sum_{k=1}^{\infty} P(A_k) \stackrel{a)}{=} \sum_{k=1}^n P(A_k).$$

c) Fur alle $A \in \mathcal{A}$ gilt $P(\overline{A}) = 1 - P(A)$, denn es gilt

$$A \cup \overline{A} = \Omega \stackrel{b)}{\Rightarrow} P(A) + P(\overline{A}) = P(\Omega) = 1.$$

d) Fur $A, B \in \mathcal{A}$ gilt $P(A) \stackrel{b)}{=} P(A \setminus B) + P(A \cap B)$, denn $A = (A \setminus B) \cup (A \cap B)$.

e) Fur $A, B \in \mathcal{A}$ gilt $P(A \cup B) \stackrel{b)}{=} P(A \setminus B) + P(B)$, denn $A \cup B = (A \setminus B) \cup (B)$.

f) Fur $A, B \in \mathcal{A}$ mit $B \subseteq A$ gilt $P(B) \leq P(A)$, denn es gilt

$$P(A) \stackrel{d)}{=} \underbrace{P(A \setminus B)}_{\in [0, 1]} + \underbrace{P(A \cap B)}_{= B} \geq P(B).$$

g) Fur $A, B \in \mathcal{A}$ gilt $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, denn

$$P(A \cup B) \stackrel{e)}{=} P(A \setminus B) + P(B) \stackrel{d)}{=} P(A) + P(B) - P(A \cap B).$$

h) Aus $A_1, \dots, A_n \in \mathcal{A}$ folgt $P\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n P(A_k)$, denn

$$\begin{aligned} P\left(\bigcup_{k=1}^n A_k\right) &= P(A_1 \cup \dots \cup A_n) = P(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \dots \cup (A_n \setminus (\bigcup_{k=1}^{n-1} A_k))) \\ &= P\left(\bigcup_{k=1}^n (A_k \setminus (\bigcup_{j=1}^{k-1} A_j))\right) \stackrel{\text{b)}}{=} \sum_{k=1}^n P\left(A_k \setminus (\bigcup_{j=1}^{k-1} A_j)\right) \\ &\stackrel{\text{f)}}{\leq} \sum_{k=1}^n P(A_k). \end{aligned}$$

i) Aus $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{A}$ folgt $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n) \in [0, \infty]$ (∞ im Falle der Divergenz), denn

$$\begin{aligned} P\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= P(A_1 \cup A_2 \cup \dots) = P(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \dots) \\ &= P\left(\bigcup_{n \in \mathbb{N}} (A_n \setminus (\bigcup_{k=1}^{n-1} A_k))\right) \stackrel{\text{b)}}{=} \sum_{n=1}^{\infty} P\left(A_n \setminus (\bigcup_{k=1}^{n-1} A_k)\right) \\ &\stackrel{\text{f)}}{\leq} \sum_{n=1}^{\infty} P(A_n) \quad (\text{Im Falle der Konvergenz, sonst } \infty). \end{aligned}$$

1.12 Satz: (Siebformel)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum $\mathbb{N} \ni n \geq 2$ und $A_1, \dots, A_n \in \mathcal{A}$, dann gilt

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n \left[(-1)^{k+1} \cdot \sum_{\substack{J \subseteq \{1, \dots, n\}, \\ |J|=k}} P\left(\bigcap_{j \in J} A_j\right) \right] = \sum_{\substack{J \subseteq \{1, \dots, n\}, \\ |J| \neq \emptyset}} (-1)^{k+1} \cdot P\left(\bigcap_{j \in J} A_j\right).$$

Beweis: mit Induktion nach n . (s. passende Stochastikbücher) □

1.13 Satz: (Ein- und Ausschlussprinzip)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum $\mathbb{N} \ni n \geq 2$, $A_1, \dots, A_n \in \mathcal{A}$ und $k \in \{0, \dots, n\}$. Sei $B_k \in \mathcal{A}$ das Ereignis, dass genau k der Ereignisse A_1, \dots, A_n eintreten und die restlichen nicht, d.h.

$$B_k := \bigcup_{\substack{J \subseteq \{1, \dots, n\}, \\ |J|=k}} \left(\left(\bigcap_{j \in J} A_j \right) \cap \left(\bigcap_{l \in \{1, \dots, n\} \setminus J} \bar{A}_l \right) \right)$$

Dann gilt für die Wahrscheinlichkeit von B_k

$$P(B_k) = \sum_{l=k}^n \left[(-1)^{l-k} \cdot \binom{l}{k} \sum_{\substack{J \subseteq \{1, \dots, n\}, \\ |J|=l}} P\left(\bigcap_{j \in J} A_j\right) \right].$$

Beweis: mit Induktion nach n . (s. passende Stochastikbücher)² □

1.14 Beispiel: (zufällige Permutationen)

Es seien $n \in \mathbb{N}$ und $\Omega := S_n = \{f : \{1, \dots, n\} \rightarrow \{1, \dots, n\} \mid f \text{ bijektiv}\}$ die symmetrische Gruppe, d.h. die Menge aller Permutationen von $\{1, \dots, n\}$. Sie hat die Mächtigkeit $|\Omega| = n!$. Betrachte nun den dazu assoziierten Laplace'schen Wahrscheinlichkeitsraum $(\Omega = S_n, \mathcal{A} = \mathcal{P}(\Omega), P)$. Eine Zahl $l \in \{1, \dots, n\}$ heißt Fixpunkt unter der Permutation $\pi \in S_n$, falls $\pi(l) = l$ gilt. Berechne nun für $k \in \{0, \dots, n\}$ die Wahrscheinlichkeit $P_k(\{\pi \in S_n \mid \pi \text{ hat genau } k \text{ verschiedene Fixpunkte}\})$. Für $k = n$ ist $\{\pi \in S_n \mid \pi(l) = l \forall l \in \{1, \dots, n\}\} = \{id\} \Rightarrow P_{k=n} = \frac{1}{n!}$.

Die allgemeine Lösung ergibt sich aus 1.13. Dazu definiere die Mengen $A_l := \{\pi \in S_n \mid \pi(l) = l\}$ und $B_k := \{\pi \in S_n \mid \pi \text{ hat genau } k \text{ verschiedene Fixpunkte}\}$. Mit diesen Definitionen befindet man sich genau in der Situation aus 1.13² und es ergibt sich für die Wahrscheinlichkeit

$$\begin{aligned} P_k &= P(B_k) = \sum_{l=k}^n \left[(-1)^{l-k} \cdot \binom{l}{k} \sum_{\substack{J \subseteq \{1, \dots, n\}, \\ |J|=l}} P\left(\bigcap_{j \in J} A_j\right) \right] \\ &= \sum_{l=k}^n \left[(-1)^{l-k} \cdot \binom{l}{k} \cdot \binom{n}{l} \cdot \frac{(n-l)!}{n!} \right] = \sum_{l=k}^n \left[(-1)^{l-k} \cdot \frac{l!}{k!(l-k)!} \cdot \frac{n!}{j!(n-j)!} \cdot \frac{(n-l)!}{n!} \right] \\ &\stackrel{j:=l-k}{=} \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \end{aligned}$$

Falls kein Fixpunkt existiert ($k = 0$) erhält man als Wahrscheinlichkeit

$$P_{k=0} = \sum_{j=0}^n \frac{(-1)^j}{j!} \xrightarrow{n \rightarrow \infty} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} = e^{-1} = \frac{1}{e}$$

Bisher war Ω endlich, $\mathcal{A} = \mathcal{P}(\Omega)$, P „Gleichverteilung“ und (Ω, \mathcal{A}, P) ein Laplace'scher Wahrscheinlichkeitsraum. Wie modelliert man nun aber Wahrscheinlichkeiten $P : \mathcal{P}(\Omega) \rightarrow [0, 1]$, falls Ω höchstens abzählbar (unendlich) ist, $\mathcal{A} = \mathcal{P}(\Omega)$ und alle einelementigen Ereignisse nicht mehr gleichwahrscheinlich sind?

1.15 Definition: (Diskrete Wahrscheinlichkeitsräume und Zähldichten)

Sei Ω eine höchstens abzählbare Menge. Ein Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Wahrscheinlichkeitsmaß P heißt diskret, falls $\mathcal{A} = \mathcal{P}(\Omega)$.

Die Funktion

$$f : \Omega \rightarrow [0, 1], \quad f(\omega) := P(\{\omega\})$$

heißt Zähldichte von P .

² Ende der zweiten Vorlesung vom Fr, 11.04.14

1.16 Fakten:

Sei (Ω, \mathcal{A}, P) ein diskreter Wahrscheinlichkeitsraum.

a) Für die Zähldichte f zum Wahrscheinlichkeitsmaß P gilt $\sum_{\omega \in \Omega} f(\omega) = 1$, denn

$$\sum_{\omega \in \Omega} f(\omega) = \sum_{\omega \in \Omega} P(\{\omega\}) \stackrel{W2}{=} P(\Omega) \stackrel{W1}{=} 1 \text{ wegen } \Omega = \bigcup_{\omega \in \Omega} \{\omega\}.$$

b) Sei nun $f: \Omega \rightarrow [0, 1]$ eine Funktion mit $\sum_{\omega \in \Omega} f(\omega) = 1$. Dann wird durch

$$P: \mathcal{P}(\Omega) \rightarrow [0, 1], \quad P(A) := \sum_{\omega \in A} f(\omega)$$

ein Wahrscheinlichkeitsmaß P mit zugehöriger Zähldichte f definiert.

Beweis. Es ist zunächst zu zeigen, dass $P(A) \in [0, 1]$ für alle $A \in \mathcal{A} = \mathcal{P}(\Omega)$ ist:

$$P(A) = \sum_{\omega \in A} \underbrace{f(\omega)}_{\in [0,1]} \geq 0$$

W1: $P(\Omega) = \sum_{\omega \in \Omega} f(\omega) = 1$. Aus W1 folgt also auch $P(A) \leq 1$ für alle $A \in \mathcal{A}$, also insgesamt $P(A) \in [0, 1]$ für alle $A \in \mathcal{A}$.

W2: (σ -Additivität) Seien $(A_n) \subseteq \mathcal{A}$ disjunkt, dann folgt mit dem Umordnungssatz für absolut konvergente Reihen

$$P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{\omega \in \bigcup_{n \in \mathbb{N}} A_n} f(\omega) = \sum_{n \in \mathbb{N}} \underbrace{\left(\sum_{\omega \in A_n} f(\omega)\right)}_{=P(A_n)} = \sum_{n \in \mathbb{N}} P(A_n).$$

□

Beispiel: Sei $|\Omega| < \infty$ und (Ω, \mathcal{A}, P) ein Laplace'scher Wahrscheinlichkeitsraum, dann ist die Zähldichte

$$f(\omega) = P(\{\omega\}) = \frac{|\{\omega\}|}{|\Omega|} = \frac{1}{|\Omega|}.$$

1.17 Binomialverteilung:

a) Idee: Ein Experiment mit zwei möglichen Ausgängen 1 (Wahrscheinlichkeit $p \in (0, 1)$) oder 0 (Wahrscheinlichkeit $1 - p \in (0, 1)$) wird n -mal unabhängig voneinander wiederholt. Die Menge aller möglichen Ergebnisse des Gesamtexperiments ist dann $\Omega = \{0, 1\}^n$. Betrachte die zueinander disjunkten Ereignisse

$$B_k := \{(x_1, \dots, x_n) \in \Omega \mid \sum_{l=1}^n x_l = k\}$$

für $k = 0, \dots, n$. Dann gilt $\bigcup_{k=0}^n B_k = \Omega$ und

$$P(B_k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dies motiviert den Begriff der Binomialverteilung.

b) Definition: Für $n \in \mathbb{N}$ und $p \in (0, 1)$ ist die Binomialverteilung $B_{n,p}$ das eindeutig definierte Wahrscheinlichkeitsmaß auf $\{0, \dots, n\}$ zu der Zähldichte

$$f_{n,p}(k) := \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \dots, n.$$

c) Wohldefiniertheit: $f_{n,p}$ ist tatsächlich eine Zähldichte, denn offensichtlich ist $f_{n,p}(k) \geq 0 \forall k = 0, \dots, n$ und mit dem binomischen Satz folgt

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

Also ist auch $f_{n,p}(k) \leq 1 \forall k = 0, \dots, n$ und insgesamt folgt aus 1.16 b), dass $B_{n,p}$ wirklich ein Wahrscheinlichkeitsmaß mit Zähldichte $f_{n,p}$ ist.

d) Beispiel: Eine faire Münze ($p = \frac{1}{2}$) wird $n = 20$ -mal geworfen. Man sagt, dass die Anzahl von Kopf ($\hat{=}$ 1) $B_{20, \frac{1}{2}}$ -verteilt ist.

1.18 Multinomialverteilung:

a) Idee: Ein Experiment mit $r \geq 2$ möglichen Ausgängen a_1, \dots, a_r und zugehörigen Wahrscheinlichkeiten $p_1, \dots, p_r \in [0, 1]$ mit $\sum_{l=1}^r p_l = 1$ wird n -mal unabhängig voneinander wiederholt. Die Menge aller möglichen Ergebnisse des Gesamtexperiments ist dann $\Omega = \{a_1, \dots, a_r\}^n$. Seien $k_1, \dots, k_r \in \mathbb{N}_0$ mit $\sum_{l=1}^r k_l = n$.

Es gilt dann

$$P(\text{„Bei } n \text{ Versuchen tritt } a_1 \text{ } k_1\text{-mal, } a_2 \text{ } k_2\text{-mal, } \dots, a_r \text{ } k_r\text{-mal auf“}) = \frac{n!}{k_1! k_2! \dots k_r!} \cdot p_1^{k_1} \cdot p_2^{k_2} \dots p_r^{k_r}.$$

Dies motiviert den Begriff der Multinomialverteilung.

b) Definition: Für $n, r \in \mathbb{N}$, $r \geq 2$ und $p_1, \dots, p_r \in [0, 1]$ mit $\sum_{l=1}^r p_l = 1$ ist die Multinomialverteilung $M_{n,r;p_1, \dots, p_r}$ das eindeutig definierte Wahrscheinlichkeitsmaß auf $\Omega = \{(k_1, \dots, k_r) \in \mathbb{N}_0^r \mid \sum_{l=1}^r k_l = n\}$ zu der Zähldichte

$$f_{n,r;p_1, \dots, p_r}((k_1, \dots, k_r)) := \frac{n!}{\underbrace{k_1! k_2! \dots k_r!}_{:= \binom{n}{k_1, \dots, k_r} \text{ Multinomialkoeffizient}}} \cdot p_1^{k_1} \cdot p_2^{k_2} \dots p_r^{k_r}.$$

c) Wohldefiniertheit: $f_{n,r;p_1,\dots,p_r}$ ist tatsächlich eine Zähldichte, denn (s. Übung 3 Aufgabe 5 Beweis analog zur Binomialverteilung mit Multinomialtheorem).

d) Beispiel: Ein fairer Würfel ($r = 6$, $p_1 = \dots = p_6 = \frac{1}{6}$) mit Augenzahlen 1, 2, 3, 4, 5, 6 wird $n = 60$ -mal geworfen. Wie groß ist nun die Wahrscheinlichkeit, dass nach 60 Würfeln jede Augenzahl genau $k_1 = \dots = k_6 = 10$ mal auftritt?³

$$P_{10} := P(\text{„Bei 60 Würfeln tritt jede Augenzahl genau 10 mal auf“}) \\ = f_{60,6;p_1=\frac{1}{6},\dots,p_6=\frac{1}{6}}((k_1 = 10, \dots, k_6 = 10)) = \frac{60!}{(10!)^6} \cdot \left(\frac{1}{6}\right)^{60} \approx 7 \cdot 10^{-5}.$$

1.19 Beispiel: Man werfe einen fairen Würfel mit den Augenzahlen $\{1, \dots, 6\}$ 10 mal unabhängig. Dann ist die Menge aller möglichen Ergebnisse $\Omega = \{1, \dots, 6\}^{10}$. Sei $B \subseteq \Omega$ das Ereignis, dass jede Augenzahl mindestens einmal auftritt. Welchen Wert hat dann $P(B)$?

Lösung 1: Die Multinomialverteilung für $n = 10$, $r = 6$ und $p_1 = \dots = p_6 = \frac{1}{6}$ ergibt den folgenden schwer zu berechnenden Ausdruck

$$P(B) = \sum_{\substack{k_1, \dots, k_6 \geq 1, \\ k_1 + \dots + k_6 = 10}} \frac{10!}{k_1! \dots k_6!} \cdot \left(\frac{1}{6}\right)^{10}.$$

Lösung 2: Das Ein-/Ausschlußprinzip aus 1.13 liefert für $n = 6$ und die Ereignisse $A_l :=$ „Das Ereignis, dass die Zahl l bei 10 Versuchen nie geworfen wird“ ein leichter auswertbares Ergebnis. Bei dieser Wahl der A_l für $l = 1, \dots, 6$ ergibt sich dann $B = B_0$ (dem B_k für $k = 0$ aus 1.13) und man erhält

$$P(B) = P(B_0) = \sum_{l=0}^6 (-1)^l \binom{l}{0} \underbrace{\sum_{\substack{J \subseteq \{1, \dots, 6\}, \\ |J|=l}} P(\bigcap_{j \in J} A_j)}_{= \binom{6}{l} \frac{(6-l)^{10}}{6^{10}}} = \sum_{l=0}^6 (-1)^l \binom{6}{l} \frac{(6-l)^{10}}{6^{10}} \approx 0,27.$$

1.20 Geometrische Verteilung:

a) Idee: Man führe ein Experiment, das die zwei möglichen Ausgänge 0 mit Wahrscheinlichkeit $1 - p$ und 1 mit Wahrscheinlichkeit p besitzt, solange durch bis zum ersten Mal eine 1 auftritt. Zähle dabei die Zahl der Fehlversuche (die Fälle, in denen keine 1, also 0 eintritt). Dann gilt für alle $k \in \mathbb{N}_0$

$$P_k(\text{„}k \text{ Nullen bis zur ersten Eins,“}) = (1 - p)^k \cdot p$$

³ Ende der dritten Vorlesung vom Mi, 16.04.14

b) Definition: Die geometrische Verteilung ist das Wahrscheinlichkeitsmaß auf $\Omega = \mathbb{N}_0$ mit zugehöriger Zähldichte

$$f_p(k) = (1-p)^k \cdot p$$

für eine Zahl $p \in (0, 1)$.

c) Wohldefiniertheit: f_p ist tatsächlich eine Zähldichte, denn offensichtlich ist $f_p(k) \geq 0 \forall k \in \mathbb{N}_0$ und mit der geometrischen Reihe folgt

$$\sum_{k=0}^{\infty} f_p(k) = \sum_{k=0}^{\infty} (1-p)^k \cdot p = p \sum_{k=0}^{\infty} \underbrace{(1-p)^k}_{<1} = p \frac{1}{1-(1-p)} = 1$$

Also ist auch $f_p(k) \leq 1 \forall k \in \mathbb{N}_0$ und insgesamt folgt aus 1.16 b), dass die geometrische Verteilung wirklich ein Wahrscheinlichkeitsmaß mit Zähldichte f_p ist.

d) Warnung: In (50%) der Literatur wird die geometrische Verteilung auf $\Omega = \mathbb{N}$ statt $\Omega = \mathbb{N}_0$ definiert. Die Zähldichte ist dann

$$f_p(k) = (1-p)^{k-1} \cdot p \text{ für } k \geq 1.$$

e) Beispiel: Ein Schalter geht mit der Wahrscheinlichkeit $p = 10^{-3}$ nach einer Betätigung kaputt. Wie groß ist nun die Wahrscheinlichkeit P_{1000} , dass der Schalter frühestens nach der tausendsten Betätigung kaputt geht?

$$P_{1000} = \sum_{k=0}^{1000} (1-p)^k p = 10^{-3} \cdot \frac{1 - (1-10^{-3})^{1001}}{1 - (1-10^{-3})} \approx 0,63$$

f) Merkregel: Geometrische Verteilungen treten oft bei Wartezeiten bis Eintritt eines Ereignisses auf.

Es gibt auch Verteilungen, die als Grenzwerte anderer Verteilungen auftreten. Diese können dann als Näherungen für die anderen verwendet werden.

1.21 Definition: (Poisson-Verteilung)

Die Poissonverteilung Π_λ mit $\lambda \in (0, \infty)$ ist das Wahrscheinlichkeitsmaß auf $\Omega = \mathbb{N}_0$ mit zugehöriger Zähldichte

$$f_\lambda(k) := e^{-\lambda} \cdot \frac{\lambda^k}{k!} \text{ für } k \in \mathbb{N}_0.$$

Wohldefiniertheit: f_λ ist tatsächlich eine Zähldichte, denn offensichtlich ist $f_\lambda(k) \geq 0 \forall k \in \mathbb{N}_0$ und mit der Exponentialreihe folgt

$$\sum_{k=0}^{\infty} f_\lambda(k) = e^{-\lambda} \cdot \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^\lambda = e^{\lambda-\lambda} = e^0 = 1$$

Also ist auch $f_\lambda(k) \leq 1 \quad \forall k \in \mathbb{N}_0$ und insgesamt folgt aus 1.16 b), dass die Poisson-Verteilung Π_λ wirklich ein Wahrscheinlichkeitsmaß mit Zähldichte f_λ ist.

Die Poisson-Verteilung ist bei geeigneten Voraussetzungen der Grenzfall der Binomialverteilung für große n .

1.22 Satz:

Sei $(p_n) \subseteq (0, 1)$ eine Folge mit der Eigenschaft $0 < \lim_{n \rightarrow \infty} n \cdot p_n < \infty$. Dann gilt für alle $k \in \mathbb{N}_0$

$$\lim_{n \rightarrow \infty} B_{n,p_n}(\{k\}) = \Pi_\lambda(\{k\}) \quad (8)$$

mit $\lambda := \lim_{n \rightarrow \infty} n \cdot p_n$.

Beweis.

a) Definiere die Hilfsfunktion $r(n) := np_n - \lambda \rightarrow 0$ für $n \rightarrow \infty$. Durch Umformen erhält man $p_n = \frac{\lambda + r(n)}{n}$.

b) Aus der Definition der Binomialverteilung ergibt sich

$$\begin{aligned} B_{n,p_n}(\{k\}) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot (np_n)^k (1-p_n)^{n-k} \\ &\stackrel{a)}{=} \frac{1}{k!} \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n^k}}_{\rightarrow 1 \text{ für } n \rightarrow \infty} \cdot \underbrace{(np_n)^k}_{\rightarrow \lambda^k \text{ für } n \rightarrow \infty} \cdot \overbrace{\frac{\left(1 - \frac{\lambda+r(n)}{n}\right)^n}{\left(1 - \frac{\lambda+r(n)}{n}\right)^k}}^{\rightarrow e^{-\lambda} \text{ für } n \rightarrow \infty} \\ &\xrightarrow{n \rightarrow \infty} \frac{1}{k!} \lambda^k e^{-\lambda} = \Pi_\lambda(\{k\}). \end{aligned}$$

□

1.23 Beispiel: In einem Brennstab befinden sich $n = 10^8$ Uranatome, von denen jedes pro Sekunde mit einer Wahrscheinlichkeit von $p = 10^{-7}$ unabhängig von den anderen zerfällt. Wie groß ist die Wahrscheinlichkeit P_{10} , dass maximal 10 Atome pro Sekunde zerfallen?

$$P_{10} = B_{10^8, 10^{-7}}(\{0, \dots, 10\}) \approx \Pi_{10}(\{0, \dots, 10\}) = e^{-10} \sum_{k=0}^{10} \frac{10^k}{k!} \approx 0,58$$

1.24 Satz:

Für $n \in \mathbb{N}$ sei P_n gemäß 1.14 die Verteilung der Anzahl von Fixpunkten bei einer zufälligen Permutation $\sigma \in S_n$ der Zahlen $\{1, \dots, n\}$. Dann gilt für alle $k \in \mathbb{N}_0$

$$\lim_{n \rightarrow \infty} P_n(\{k\}) = \Pi_1(\{k\}) \quad (9)$$

Beweis.

$$P_n(\{k\}) \stackrel{1.14}{=} \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \xrightarrow{n \rightarrow \infty} \frac{1}{k!} e^{-1} = \Pi_1(\{k\})$$

□

2 Bedingte Wahrscheinlichkeiten und Unabhängigkeit

2.1 Definition:

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $P(B) > 0$. Dann heißt

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \in [0, 1]$$

die bedingte Wahrscheinlichkeit von A unter der Bedingung B .

2.2 Definition und Satz:

Für festes $B \in \mathcal{A}$ mit $P(B) > 0$ wird durch

$$P_B : \mathcal{A} \rightarrow [0, 1], \quad P_B(A) := P(A|B)$$

wieder ein Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) definiert. Es heißt bedingtes Wahrscheinlichkeitsmaß von P unter der Bedingung B .

Beweis.

(W1) Für $B \in \mathcal{A}$ fest gilt $P_B(\Omega) = P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

(W2) Seien $(A_n) \subseteq \mathcal{A}$ disjunkt, dann gilt

$$P_B\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \frac{P\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\bigcup_{n \in \mathbb{N}} (A_n \cap B)\right)}{P(B)} = \sum_{n \in \mathbb{N}} \frac{P(A_n \cap B)}{P(B)} = \sum_{n \in \mathbb{N}} P(A_n|B).$$

□

2.3 Rechenregeln:

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

a) Seien $A_1, \dots, A_n \in \mathcal{A}$ mit $P\left(\bigcap_{k=1}^{n-1} A_k\right) > 0$. Dann gilt

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1) \cdot \prod_{l=2}^n P(A_l | \bigcap_{k=1}^{l-1} A_k)$$

denn

$$\begin{aligned} P(A_1) \cdot \prod_{l=2}^n P(A_l | \bigcap_{k=1}^{l-1} A_k) &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n | \bigcap_{k=1}^{n-1} A_k) \\ &= P(A_1) \cdot \frac{P(A_1 \cap A_2)}{P(A_1)} \cdots \frac{P\left(\bigcap_{k=1}^n A_k\right)}{P\left(\bigcap_{k=1}^{n-1} A_k\right)} = P\left(\bigcap_{k=1}^n A_k\right). \end{aligned}$$

b) Formel von der totalen Wahrscheinlichkeit: Sei $B_1, \dots, B_n \in \mathcal{A}$ eine Partition (d.h. disjunkte Zerlegung) von Ω . Sei weiter $P(B_k) > 0 \forall k = 1, \dots, n$. Dann gilt für alle $A \in \mathcal{A}$

$$P(A) = \sum_{k=1}^n P(B_k) \cdot P(A|B_k)$$

denn wegen $P(A|B_k) = \frac{P(A \cap B_k)}{P(B_k)} \Leftrightarrow P(A \cap B_k) = P(A|B_k) \cdot P(B_k)$ folgt

$$P(A) = P(A \cap \Omega) = P(A \cap (\bigcup_{k=1}^n B_k)) = P(\bigcup_{k=1}^n (A \cap B_k)) = \sum_{k=1}^n P(A \cap B_k) = \sum_{k=1}^n P(B_k) \cdot P(A|B_k).$$

Die Rechenregeln aus 2.3 sind oft zur Berechnung konkreter Wahrscheinlichkeiten bei zeitlich oder logisch gestuften Experimenten.

2.4 Beispiel: Ein gewöhnliches Skatblatt besteht aus 32 Karten, 4 davon sind Ass. Diese Karten werden nun zufällig auf 4 Spieler verteilt, d.h. jeder erhält 8 Karten. Wie groß ist die Wahrscheinlichkeit P_A , dass jeder Spieler genau ein Ass bekommt?

Lösung: Sei A_k das Ereignis, dass Spieler k genau ein Ass erhält, $k = 1, 2, 3, 4$. Dann folgt⁴

$$\begin{aligned} P_A &= P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot P(A_4|A_1 \cap A_2 \cap A_3) \\ &= \frac{\binom{4}{1} \binom{28}{7}}{\binom{32}{8}} \cdot \frac{\binom{3}{1} \binom{21}{7}}{\binom{24}{8}} \cdot \frac{\binom{2}{1} \binom{14}{7}}{\binom{16}{8}} \cdot \frac{\binom{1}{1} \binom{7}{7}}{\binom{8}{8}} \approx 0,11 \end{aligned}$$

2.5 Satz: (Formel von Bayes)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $B_1, \dots, B_n \in \mathcal{A}$ eine disjunkte Zerlegung von Ω mit $P(B_k) > 0$ für $k = 1, \dots, n$. Dann gilt für $A \in \mathcal{A}$ mit $P(A) > 0$

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{k=1}^n P(A|B_k)P(B_k)} \text{ für } j \in \{1, \dots, n\}. \quad (10)$$

Beweis.

Mit der Formel von der totalen Wahrscheinlichkeit folgt für $j \in \{1, \dots, n\}$

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A)} \stackrel{2.3b)}{=} P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{k=1}^n P(A|B_k)P(B_k)}.$$

□

⁴ Ende der vierten Vorlesung vom Mi, 23.04.14

2.6 Beispiel: (Krankheit) Eine seltene Krankheit komme bei 0,1% der Bevölkerung vor. Ein Test erkenne die Krankheit mit einer Wahrscheinlichkeit von 0,998 bei einer kranken Person richtig und er schlage bei 0,3% der gesunden Personen zu Unrecht an. Wie hoch ist die Wahrscheinlichkeit, dass eine Person wirklich krank ist, wenn der Test ein positives Resultat liefert?

Lösung: Um die Formel von Bayes anwenden zu können, muss man zunächst alle gegebenen Daten korrekt formalisieren. Dazu sei A das Ereignis, dass der Test positiv ausgefallen ist und B_1 das Ereignis, dass die Person krank ist, sowie B_2 das Ereignis, dass die Person gesund ist. In der Situation der Formel von Bayes ist dann $n = 2$, $P(B_1) = 0,001$, $P(B_2) = 0,999$, $P(A|B_1) = 0,998$ und $P(A|B_2) = 0,003$. Für die gesuchte Wahrscheinlichkeit $P(B_1|A)$ ergibt sich dann

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} \approx 25\%.$$

2.7 Definition: (stochastische Unabhängigkeit)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $I \neq \emptyset$ eine beliebige Indexmenge.

a) Zwei Mengen $A, B \in \mathcal{A}$ heißen unabhängig, falls

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

b) Mengen $A_i \in \mathcal{A}$, $i \in I$ heißen paarweise unabhängig, falls für $i \neq j \in I$ gilt

$$P(A_i \cap A_j) = P(A_i)P(A_j).$$

c) Mengen $A_i \in \mathcal{A}$, $i \in I$ heißen unabhängig, falls für alle $n \in \mathbb{N}$, $i_1, \dots, i_n \in I$ verschieden gilt

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

Bemerkung: Aus der Unabhängigkeit folgt die paarweise Unabhängigkeit, die Umkehrung gilt im Allgemeinen nicht.

2.8 Fakten: Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $I \neq \emptyset$ eine beliebige Indexmenge.

a) Sind $A_1, \dots, A_n \in \mathcal{A}$ unabhängig, so folgt

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

Die Rückrichtung ist im Allgemeinen falsch.

b) Sei $B \in \mathcal{A}$ mit $P(B) = 0$ und $A \in \mathcal{A}$ beliebig, dann sind A und B unabhängig, denn

$$0 \leq P(A \cap B) \leq P(B) = 0 \Rightarrow P(A \cap B) = 0 = P(A)P(B).$$

c) Sei $B \in \mathcal{A}$ mit $P(B) > 0$ und $A \in \mathcal{A}$ beliebig, dann gilt

$$A, B \text{ unabhängig} \Leftrightarrow P(A|B) = P(A),$$

denn

$$\begin{aligned} A, B \text{ unabhängig} &\Leftrightarrow P(A \cap B) = P(A)P(B) \\ &\Leftrightarrow P(A) = \frac{P(A \cap B)}{P(B)} = P(A|B). \end{aligned}$$

d) Seien $A, B \in \mathcal{A}$ unabhängig, dann sind auch A und $\bar{B} = \Omega \setminus B$ unabhängig, denn

$$\begin{aligned} P(A \cap \bar{B}) &= P(A \cap (\Omega \setminus B)) = P((A \cap \Omega) \setminus (A \cap B)) = P(A \setminus (A \cap B)) \\ &\stackrel{1.11d)}{=} P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) \\ &= P(A)P(\bar{B}). \end{aligned}$$

2.9 Definition: (Unabhängigkeit von Mengensystemen)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $I \neq \emptyset$ eine beliebige Indexmenge. Mengensysteme $\mathcal{C}_i \subseteq \mathcal{A}$ für $i \in I$ heißen unabhängig, falls $\forall n \in \mathbb{N}$, $i_1, \dots, i_n \in I$ verschieden und $\forall A_{i_1} \in \mathcal{C}_{i_1}, \dots, A_{i_n} \in \mathcal{C}_{i_n}$ gilt

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

Bemerkung: Eine äquivalente Definition der Unabhängigkeit von Mengensystemen ist: Mengensysteme $\mathcal{C}_i \subseteq \mathcal{A}$ für $i \in I$ sind unabhängig, falls bei allen möglichen Auswahlen $A_i \in \mathcal{C}_i$ ($i \in I$) die A_i im Sinne von Definition 2.7 c) unabhängig sind.

2.10 Definition: (Dynkin-Systeme)

Sei $\Omega \neq \emptyset$ eine Menge.

a) Für ein Mengensystem $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ heißt

$$\sigma(\mathcal{C}) := \bigcap_{\substack{\mathcal{F} \text{ ist } \sigma\text{-Algebra} \\ \text{auf } \Omega \text{ mit } \mathcal{C} \subseteq \mathcal{F}}} \mathcal{F}$$

die von \mathcal{C} erzeugte σ -Algebra.

b) Ein Mengensystem $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ heißt durchschnittstabil (durchschnittabgeschlossen), falls für alle $A, B \in \mathcal{C}$ auch stets $A \cap B \in \mathcal{C}$ folgt.

c) Ein Mengensystem $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ heißt Dynkin-System, falls

$$\begin{aligned} \text{(D1)} \quad &\Omega \in \mathcal{D}, \\ \text{(D2)} \quad &D_1, D_2 \in \mathcal{D} \text{ mit } D_1 \subseteq D_2 \Rightarrow D_2 \setminus D_1 \in \mathcal{D}, \\ \text{(D3)} \quad &(D_n)_{n \in \mathbb{N}} \subseteq \mathcal{D} \text{ disjunkt} \Rightarrow \bigcup_{n \in \mathbb{N}} D_n \in \mathcal{D}. \end{aligned}$$

d) Für ein Mengensystem $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ heißt

$$\mathcal{D}(\mathcal{C}) := \bigcap_{\substack{\mathcal{F} \text{ ist Dynkin-System} \\ \text{auf } \Omega \text{ mit } \mathcal{C} \subseteq \mathcal{F}}} \mathcal{F}$$

das von \mathcal{C} erzeugte Dynkin-System.

Bemerkungen: a) Es lässt sich leicht zeigen (s. Übungen), dass $\sigma(\mathcal{C})$ wieder eine σ -Algebra ist. Sie ist die kleinste σ -Algebra, die \mathcal{C} enthält.

b) Nach 1.9 e), f) ist jede Algebra und auch jede σ -Algebra durchschnittstabil.

c) Jede σ -Algebra \mathcal{A} ist auch ein Dynkin-System, denn

(D1) $\Omega \in \mathcal{A}$ ist klar,

(D2) $A_1, A_2 \in \mathcal{A}$ mit $A_1 \subseteq A_2 \Rightarrow A_2 \setminus A_1 = \underbrace{A_2}_{\in \mathcal{A}} \cap \underbrace{(\Omega \setminus A_1)}_{\in \mathcal{A}} \in \mathcal{A}$,

(D3) ist klar.

d) Es lässt sich leicht zeigen, dass $\mathcal{D}(\mathcal{C})$ wieder ein Dynkin-System ist. Es ist das kleinste Dynkin-System, das \mathcal{C} enthält.

e) Aus (D1) und (D2) folgt sofort, dass auch $\phi \in \mathcal{D}(\mathcal{C})$ ist.

Beispiel: Zweimaliges unabhängiges Würfeln wird durch den Laplace'schen Wahrscheinlichkeitsraum $(\Omega = \{1, \dots, 6\}^2, \mathcal{A} = \mathcal{P}(\Omega), P)$ beschrieben. Die Menge

$$\mathcal{C}_1 := \{\{i\} \times \{1, \dots, 6\} \in \mathcal{A} \mid i \in \{1, \dots, 6\}\}$$

beschreibt die Ereignisse des ersten Wurfs. Analog beschreibt die Menge

$$\mathcal{C}_2 := \{\{1, \dots, 6\} \times \{j\} \in \mathcal{A} \mid j \in \{1, \dots, 6\}\}$$

die Ereignisse des zweiten Wurfs. In diesem Fall ist die Indexmenge $I = \{1, 2\}$ und die Mengensysteme $\mathcal{C}_1, \mathcal{C}_2 \in \mathcal{A}$ sind unabhängig, denn betrachte zu festem $i \in \{1, \dots, 6\}$ das Ereignis $A_i := \{i\} \times \{1, \dots, 6\} \in \mathcal{C}_1$ und für festes $j \in \{1, \dots, 6\}$ das Ereignis $A_j := \{1, \dots, 6\} \times \{j\} \in \mathcal{C}_2$. Dann gilt für die Einzelwahrscheinlichkeiten natürlich $P(A_i) = \frac{|A_i|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$ und $P(A_j) = \frac{|A_j|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$, also $P(A_i)P(A_j) = \frac{1}{36}$. Wegen $A_i \cap A_j = \{(i, j)\}$ gilt auch

$$P(A_i \cap A_j) = \frac{|A_i \cap A_j|}{|\Omega|} = \frac{|\{(i, j)\}|}{36} = \frac{1}{36} = P(A_i)P(A_j).$$

Die kleinste σ -Algebra, die \mathcal{C}_1 enthält ist

$$\sigma(\mathcal{C}_1) = \{A \times \{1, \dots, 6\} \in \mathcal{A} \mid A \subseteq \{1, \dots, 6\}\}.$$

Die kleinste σ -Algebra, die \mathcal{C}_2 enthält ist

$$\sigma(\mathcal{C}_2) = \{\{1, \dots, 6\} \times B \in \mathcal{A} \mid B \subseteq \{1, \dots, 6\}\}.$$

2.11 Lemma: (Charakterisierung von σ -Algebren)

Sei $\Omega \neq \emptyset$ und \mathcal{C} ein Teilmengensystem von $\mathcal{P}(\Omega)$. Dann gilt:

- a) \mathcal{C} ist σ -Algebra $\Leftrightarrow \mathcal{C}$ ist durchschnittstabiles Dynkinsystem.
 b) Wenn \mathcal{C} durchschnittstabil ist, so gilt $\sigma(\mathcal{C}) = \mathcal{D}(\mathcal{C})$.⁵

Beweis.

a) Für Teil a) sind zwei Implikationen zu zeigen

„ \Rightarrow “: vgl. Bemerkung c) nach Definition 2.10.

„ \Leftarrow “: Sei $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ ein durchschnittstabiles Dynkinsystem, dann erfüllt es die Axiome einer σ -Algebra

(A1): Aus (D1) folgt sofort $\Omega \in \mathcal{C}$.

(A2): Wegen $\Omega \in \mathcal{C}$ folgt mit (D2) aus $A \in \mathcal{C}$ sofort $\bar{A} = \Omega \setminus A \in \mathcal{C}$.

(A3): Seien $A_1, A_2 \in \mathcal{C}$, dann folgt aus (D3) und (D2) sofort

$$A \cup B = A \cup (B \setminus (A \cap B)) \cup \emptyset \cup \dots \in \mathcal{C}.$$

(σ 2): Es sei $(A_n)_{n \in \mathbb{N}} \subseteq \mathcal{C}$, dann folgt induktiv aus (A3), dass auch $B_k := A_1 \cup \dots \cup A_k$, $k \in \mathbb{N}$ in \mathcal{C} liegt und wegen $B_k \subseteq B_{k+1}$ folgt letztlich

$$\bigcup_{n \in \mathbb{N}} A_n = B_1 \cup (B_2 \setminus B_1) \cup (B_3 \setminus B_2) \cup \dots \in \mathcal{C}.$$

Also ist \mathcal{C} eine σ -Algebra.

b) Für Teil b) sind zwei Mengeninklusionen zu zeigen

„ \supseteq “: $\sigma(\mathcal{C})$ ist σ -Algebra $\Rightarrow \sigma(\mathcal{C})$ ist Dynkinsystem mit $\mathcal{C} \subseteq \sigma(\mathcal{C})$. $\mathcal{D}(\mathcal{C})$ ist das kleinste Dynkinsystem, das \mathcal{C} enthält $\Rightarrow \sigma(\mathcal{C}) \supseteq \mathcal{D}(\mathcal{C})$.

„ \subseteq “: Zeige, dass wenn \mathcal{C} durchschnittstabil ist, auch $\mathcal{D}(\mathcal{C})$ durchschnittstabil ist.

Zu $A \in \mathcal{D}(\mathcal{C})$ definiere das Mengensystem

$$\mathcal{D}_A := \{C \subseteq \Omega \mid A \cap C \in \mathcal{D}(\mathcal{C})\} \subseteq \mathcal{D}(\mathcal{C}).$$

\mathcal{D}_A ist Dynkin-System, denn

(D1) $\Omega \in \mathcal{D}_A$, denn $\Omega \cap A = A \in \mathcal{D}(\mathcal{C})$ und $\Omega \in \mathcal{D}(\mathcal{C})$.

(D2) Seien $C_1, C_2 \in \mathcal{D}_A$ mit $C_1 \subseteq C_2$, dann folgt

$$A_1 \cap (C_2 \setminus C_1) = \underbrace{(A \cap C_2)}_{\in \mathcal{D}(\mathcal{C})} \setminus \underbrace{(A \cap C_1)}_{\in \mathcal{D}(\mathcal{C})} \in \mathcal{D}(\mathcal{C}) \Rightarrow C_2 \setminus C_1 \in \mathcal{D}_A.$$

(D3) Seien $(C_n)_{n \in \mathbb{N}} \subseteq \mathcal{D}_A$ disjunkt, dann gilt

$$A \cap \left(\bigcup_{n \in \mathbb{N}} C_n \right) = \bigcup_{n \in \mathbb{N}} (A \cap C_n) \in \mathcal{D}(\mathcal{C}) \Rightarrow \bigcup_{n \in \mathbb{N}} C_n \in \mathcal{D}_A.$$

⁵ Ende der fünften Vorlesung vom Fr, 25.04.14

Sei nun $A \in \mathcal{C}$, dann ist auch $\mathcal{C} \subseteq \mathcal{D}_A$, denn für alle $C \in \mathcal{C}$ gilt $A \cap C \in \mathcal{C} \subseteq \mathcal{D}(\mathcal{C})$, weil \mathcal{C} durchschnittstabil ist. Also ist für alle $A \in \mathcal{C}$ das kleinste Dynkin-System, das \mathcal{C} enthält $\mathcal{D}(\mathcal{C})$ auch in \mathcal{D}_A enthalten. Daraus folgt dann

$$\forall A \in \mathcal{C}, \forall C \in \mathcal{D}(\mathcal{C}) : A \cap C \in \mathcal{D}(\mathcal{C}).$$

Nun ist also auch $\mathcal{C} \subseteq \mathcal{D}_C$ für alle $C \in \mathcal{D}(\mathcal{C})$. Es folgt natürlich dann auch wieder $\mathcal{D}(\mathcal{C}) \subseteq \mathcal{D}_C$ für alle $C \in \mathcal{D}(\mathcal{C})$. Hieraus folgt nun, dass $\mathcal{D}(\mathcal{C})$ durchschnittstabil ist.

Nach Teil a) ist jedes durchschnittstabile Dynkinsystem auch eine σ -Algebra. Also ist $\mathcal{D}(\mathcal{C})$ eine σ -Algebra, die \mathcal{C} enthält $\Rightarrow \sigma(\mathcal{C}) \subseteq \mathcal{D}(\mathcal{C})$. □

2.12 Satz:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(\mathcal{C}_n)_{n \in \mathbb{N}}$ durchschnittstabil und unabhängig bzgl. P . Dann ist auch die Folge $(\sigma(\mathcal{C}_n))$ der von \mathcal{C}_n erzeugten σ -Algebren unabhängig.

Beweis. (s. Übung)

□

Beispiel: Zweimaliges Würfeln sei durch den Laplace'schen Wahrscheinlichkeitsraum $(\Omega = \{1, \dots, 6\}^2, \mathcal{A} = \mathcal{P}(\Omega), P)$ beschrieben. Dann sind

$$\tilde{\mathcal{C}}_1 = \{\{i\} \times \{1, \dots, 6\} | i \in \{1, \dots, 6\}\} \cup \emptyset$$

und

$$\tilde{\mathcal{C}}_2 = \{\{1, \dots, 6\} \times \{j\} | j \in \{1, \dots, 6\}\} \cup \emptyset$$

durchschnittstabile und unabhängige Mengensysteme. Dann sind auch die von ihnen erzeugten σ -Algebren

$$\sigma(\tilde{\mathcal{C}}_1) = \{A \times \{1, \dots, 6\} | A \subseteq \{1, \dots, 6\}\}$$

und

$$\sigma(\tilde{\mathcal{C}}_2) = \{\{1, \dots, 6\} \times B | B \subseteq \{1, \dots, 6\}\}$$

nach Satz 2.12 unabhängig.

2.13 Produktexperimente bei diskreten Wahrscheinlichkeitsräumen:

Für $k = 1, \dots, n$ seien $(\Omega_k, \mathcal{A}_k, P_k)$ diskrete Wahrscheinlichkeitsräume mit zugehörigen Zähldichten f_k . Dann definiere den höchstens abzählbaren Produktergebnisraum

$$\Omega := \Omega_1 \times \dots \times \Omega_n.$$

Definiere nun für $\omega = (\omega_1, \dots, \omega_n) \in \Omega$

$$f(\omega) := \prod_{k=1}^n f_k(\omega_k).$$

Dies ist dann eine Zähldichte auf Ω , denn

$$\sum_{\omega \in \Omega} f(\omega) = \prod_{k=1}^n \underbrace{\sum_{\omega_k \in \Omega_k} f_k(\omega_k)}_{=1, \forall k=1, \dots, n} = 1^n = 1.$$

Das zu f assoziierte Wahrscheinlichkeitsmaß auf $\mathcal{P}(\Omega)$ heißt Produktwahrscheinlichkeitsmaß und der zugehörige Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ heißt Produktwahrscheinlichkeitsraum. Für $k = 1, \dots, n$ sei

$$\mathcal{A}_k := \{\Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n \mid A_k \in \mathcal{P}(\Omega_k)\}$$

und

$$P = P_1 \otimes \dots \otimes P_n.$$

Die $\mathcal{A}_1, \dots, \mathcal{A}_n$ bilden eine σ -Algebra (leichtes Nachrechnen der Axiome) und sind unabhängig, denn

seien $A_1 \subseteq \Omega_1, \dots, A_n \subseteq \Omega_n$, dann gilt

$$\begin{aligned} P\left(\bigcap_{k=1}^n \mathcal{A}_k\right) &= P\left(\bigcap_{k=1}^n (\Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n)\right) \\ &= P(A_1 \times \dots \times A_n) = \prod_{k=1}^n \sum_{\omega_k \in \Omega_k} f_k(\omega_k) = \prod_{k=1}^n P_k(A_k) \\ &= \prod_{k=1}^n P(\Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n). \end{aligned}$$

Deutung: Die Mengen $\Omega_1 \times \dots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \dots \times \Omega_n$ beschreiben die Ereignisse der k -ten Komponente. In Produkträumen sind die Ergebnisse einzelner Komponenten also stochastisch unabhängig. Es ist sinnvoll mehrere voneinander unabhängige Experimente durch Produkträume zu modellieren.

Häufig hat man es auch mit gestuften Experimenten zu tun, bei denen Abhängigkeiten zwischen den jeweiligen Stufen bestehen. Solche Experimente sollen nun durch Rückgriffe auf 2.3 a) modelliert werden.

2.14 Modellierung gestufter Experimente:

Ein Experiment bestehe aus n gestuften Einzelexperimenten mit den höchstens abzählbaren Ergebnismengen $\Omega_1, \dots, \Omega_n$. Bekannt seien:

1. Die Verteilung P_1 der Ergebnisse aus Ω_1 mit Zähldichte f_1 (1. Experiment)
2. Die sogenannten bedingten Zähldichten für $k = 2, \dots, n$ und $\omega_k \in \Omega_k$

$f_k(\omega_k | \bigcap_{l=1}^{k-1} \omega_l) :=$ Wahrscheinlichkeit für Ereignis $\omega_k \in \Omega_k$ unter der Bedingung, dass zuvor $\omega_1, \dots, \omega_{k-1}$ eingetreten sind

Man fordert nun für alle $\omega_k \in \Omega_k$ ($k = 1, \dots, n$), dass gilt:

$$f_k(\omega_k | \bigcap_{l=1}^{k-1} \omega_l) \geq 0 \text{ und } \sum_{\omega_k \in \Omega_k} f_k(\omega_k | \bigcap_{l=1}^{k-1} \omega_l) = 1$$

Die Zähldichten f_k aus 2. ergeben sich direkt aus der Modellbeschreibung. Das Gesamtexperiment wird durch den Wahrscheinlichkeitsraum

$$(\Omega = \Omega_1 \times \dots \times \Omega_n, \mathcal{A} = \mathcal{P}(\Omega), P)$$

modelliert.

Wegen 2.3 a) hat P die Zähldichten

$$f((\omega_1, \dots, \omega_n)) := f_1(\omega_1) \cdot f_2(\omega_2 | \omega_1) \cdots f_n(\omega_n | \bigcap_{k=1}^{n-1} \omega_k). \tag{11}$$

Dann ist f tatsächlich eine Zähldichte, denn $f((\omega_1, \dots, \omega_n)) \geq 0$ und

$$\sum_{\omega \in \Omega} f(\omega) = \sum_{\omega_1 \in \Omega_1} \cdots \underbrace{\sum_{\omega_n \in \Omega_n} f_n(\omega_n | \bigcap_{k=1}^{n-1} \omega_k) \cdots f_1(\omega_1)}_{=1} = 1.$$

Bei gestuften Experimenten ist eine graphische Darstellung durch sogenannte Entscheidungsbäume (Übergangsgraphen, Wahrscheinlichkeitsbäume, ...) möglich. Dabei werden die Kanten mit den Wahrscheinlichkeiten versehen. Die Elemente aus Ω sind gerade die „Pfade“ im Entscheidungsbaum von Start bis Ende.

2.15 Beispiel: (Polyasches Urnenmodell)

Seien $c \in \{-1, 0, 1, 2, \dots\}$ fest und $w, s \in \mathbb{N}$. Eine Urne beinhalte w weiße und s schwarze Kugeln. Bei jeder Ziehung aus der Urne zieht man zufällig eine Kugel, notiere deren Farbe und ersetze sie durch $(c + 1) \in \mathbb{N}_0$ Kugeln gleicher Farbe, also

$c = -1$: Ziehen ohne Zurücklegen

$c = 0$: Ziehen mit Zurücklegen

$c \geq 1$: Vermehrung der bereits gezogenen Kugeln.

[Wird häufig auf Populationsmodelle mit 2 Arten o.Ä. angewendet.]

Modell für n Ziehungen: Sei für $k = 1, \dots, n$ jede Einzelziehung durch $\Omega_k := \{\text{schwarz, weiß}\}$ und das Gesamtexperiment durch $\Omega := \{\text{schwarz, weiß}\}^n$ modelliert. Die bedingten Zähldichten gemäß Gleichung (11) sind:

1. $f_1(\text{weiß}) = \frac{w}{w+s}, f_1(\text{schwarz}) = \frac{s}{w+s}$

2. Für $k = 2, \dots, n$: Sei $(\omega_1, \dots, \omega_{k-1}) \in \{\text{schwarz, weiß}\}^{k-1}$, sodass dort l -mal ($l = 0, \dots, k-1$) weiß auftritt. Dann ist

$$f_k(\text{weiß} | \bigcap_{j=1}^{k-1} \omega_j) = \frac{w + l \cdot c}{w + s + (k-1) \cdot c}$$

und⁶

$$f_k(\text{schwarz} | \bigcap_{j=1}^{k-1} \omega_j) = 1 - f_k(\text{weiß} | \bigcap_{j=1}^{k-1} \omega_j).$$

2.16 Satz:

Sei $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ und $k := |\{i \in \{1, \dots, n\} | \omega_i = \text{weiß}\}| \in \{0, \dots, n\}$. Dann ist die zu f_1, \dots, f_n gehörende Zähldichte des Gesamtexperiments auf $\Omega = \Omega_1 \times \dots \times \Omega_n$ gemäß Gleichung (11) gegeben durch

$$f^{(n)}(\omega) = \frac{\prod_{i=0}^{k-1} (w + ic) \cdot \prod_{j=0}^{n-k-1} (s + jc)}{\prod_{l=0}^{n-1} (w + s + lc)}. \tag{12}$$

Beweis: (mit Induktion nach n).

Induktionsanfang $\boxed{n = 1}$: $f^{(n)}$ entspricht der Zähldichte f_1 .

Induktionsvoraussetzung (I.V.): Die Aussage gelte bereits für beliebiges aber festes n .

Induktionsschluss $\boxed{n \rightarrow n + 1}$: Sei die $n + 1$ -te Kugel schwarz, dann ist

$$\begin{aligned} f^{(n+1)}(\omega) &= f^{(n+1)}(\omega) \cdot f^{(n+1)}(\text{schwarz} | \omega_1, \dots, \omega_n) \stackrel{\text{I.V.}}{=} \frac{\prod_{i=0}^{k-1} (w + ic) \cdot \prod_{j=0}^{n-k-1} (s + jc)}{\prod_{l=0}^{n-1} (w + s + lc)} \cdot \frac{s + (n-k)c}{w + s + nc} \\ &= \frac{\prod_{i=0}^{k-1} (w + ic) \cdot \prod_{j=0}^{n-k} (s + jc)}{\prod_{l=0}^n (w + s + lc)}. \end{aligned}$$

Für den Fall, dass die Kugel weiß ist, geht der Beweis völlig analog. □

⁶ Ende der sechsten Vorlesung vom Mi, 30.04.14

2.17 Korollar:

Für alle $s, w, n \in \mathbb{N}$ und $-1 \leq c \in \mathbb{Z}$ gilt

$$\sum_{k=0}^n \binom{n}{k} \frac{\prod_{i=0}^{k-1} (w + ic) \cdot \prod_{j=0}^{n-k-1} (s + jc)}{\prod_{l=0}^{n-1} (w + s + lc)} = 1. \tag{13}$$

2.18 Beispiel:

$c = 0$: Ziehen mit Zurücklegen mit der Wahrscheinlichkeit $p := \frac{w}{w+s}$ eine weiße Kugel zu ziehen.

$$\stackrel{2.17}{\Rightarrow} \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = 1 \text{ für } n \leq s+w$$

$c = -1$: Ziehen ohne Zurücklegen. Dann gilt für $n \leq s+w$

$$\begin{aligned} P(\text{„weiß wird } k \text{ mal gezogen“}) &= \binom{n}{k} \frac{w(w-1)\cdots(w-k+1) \cdot s \cdot (s-1)\cdots(s-n+k+1)}{(w+s)(w+s-1)\cdots(w+s-n+1)} \\ &= \frac{n!}{k!(n-k)!} \cdot \frac{(w+s-n)! \cdot w! \cdot s!}{(w+s)! \cdot (w-k)! \cdot (s-n+k)!} \\ &= \frac{\binom{w}{k} \cdot \binom{s}{n-k}}{\binom{s+w}{n}}. \end{aligned}$$

$$\stackrel{2.17}{\Rightarrow} \sum_{k=0}^n \binom{w}{k} \binom{s}{n-k} = \binom{s+w}{n} \quad \forall s, w \in \mathbb{N}.$$

Diese Formel gilt sogar für alle $s, w \in \mathbb{R}$. (s. Übungen)

2.19 Hypergeometrische Verteilung:

Für $s, w, n \in \mathbb{N}$ mit $n \leq s+w$ ist die Hypergeometrische Verteilung $H_{s,w,n}$ die Verteilung auf $\Omega = \{0, \dots, n\}$ mit zugehöriger Zähldichte

$$f_{s,w,n}(k) = \frac{\binom{w}{k} \cdot \binom{s}{n-k}}{\binom{s+w}{n}} \text{ für } k = 0, \dots, n.$$

3 Zufallsvariablen und Verteilungen

3.1 Definition:

- a) Sei $\Omega \neq \emptyset$ eine Menge und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine σ -Algebra auf Ω . Dann heißen alle Mengen $A \in \mathcal{A}$ messbar und (Ω, \mathcal{A}) Messraum.
 b) Seien $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ Messräume. Eine Abbildung

$$X : \Omega_1 \rightarrow \Omega_2$$

heißt messbar bzgl \mathcal{A}_1 und \mathcal{A}_2 , falls für alle $A_2 \in \mathcal{A}_2$ gilt

$$X^{-1}(A_2) \in \mathcal{A}_1 \text{ mit } X^{-1}(A_2) := \{\omega_1 \in \Omega_1 | X(\omega_1) \in A_2\}.$$

Das bedeutet, dass die Abbildung X genau dann messbar ist, falls die Urbilder messbarer Mengen wieder messbar sind.

- c) Sei nun (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(\tilde{\Omega}, \tilde{\mathcal{A}})$ ein Messraum. Dann wird die messbare Abbildung

$$X : \Omega \rightarrow \tilde{\Omega}$$

Zufallsvariable (ZV) genannt.

3.2 Beispiel:

- a) n -facher Münzwurf mit $(\Omega = \{0, 1\}^n, \mathcal{A} = \mathcal{P}(\Omega), P)$. Sei P die Gleichverteilung und $(\tilde{\Omega} = \{0, \dots, n\}, \tilde{\mathcal{A}} = \mathcal{P}(\tilde{\Omega}))$ ein Messraum. Dann ist die Abbildung

$$Z : \Omega \rightarrow \tilde{\Omega} \text{ mit } Z((\omega_1, \dots, \omega_n)) := w_1 + \dots + w_n$$

messbar, also eine Zufallsvariable. Die Anzahl der Einsen ist $B_{n, \frac{1}{2}}$ -verteilt, also gilt für $k = 0, \dots, n$

$$B_{n, \frac{1}{2}}(k) = \frac{\binom{n}{k}}{2^n} = \frac{|\{\omega \in \Omega | \omega_1 + \dots + \omega_n = k\}|}{|\Omega|} = P(\omega \in \Omega | Z(\omega) = k) = P(Z^{-1}(k)).$$

- b) Ist (Ω, \mathcal{A}, P) ein diskreter Wahrscheinlichkeitsraum mit $\mathcal{A} = \mathcal{P}(\Omega)$, so ist für jeden Messraum $(\tilde{\Omega}, \tilde{\mathcal{A}} = \mathcal{P}(\tilde{\Omega}))$ jede Abbildung

$$X : \Omega \rightarrow \tilde{\Omega}$$

eine Zufallsvariable. In diskreten Wahrscheinlichkeitsräumen ist Messbarkeit also immer gegeben, falls $\mathcal{A} = \mathcal{P}(\Omega)$.

3.3 Definition und Satz:

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\tilde{\Omega}, \tilde{\mathcal{A}})$ ein Messraum und $X : \Omega \rightarrow \tilde{\Omega}$ eine Zufallsvariable.

Dann wird durch

$$X(P)(A) := P(X \in A) := P_X(A) := P(X^{-1}(A)) = P(\{\omega \in \Omega | X(\omega) \in A\}) \quad (14)$$

ein Wahrscheinlichkeitsmaß auf $(\tilde{\Omega}, \tilde{\mathcal{A}})$ definiert. P_X heißt das *Bildmaß* von P unter X oder die *Verteilung* der Zufallsvariablen X .

Beweis. Zeige, dass P_X ein Wahrscheinlichkeitsmaß ist:

(W1) $P_X(\tilde{\Omega}) = P(X^{-1}(\tilde{\Omega})) = P(\Omega) = 1$

(W2) Seien $(A_n)_{n \in \mathbb{N}} \subseteq \tilde{\mathcal{A}}$ disjunkt, dann sind auch die Urbilder $(X^{-1}(A_n))_{n \in \mathbb{N}} \subseteq \mathcal{A}$ disjunkt. Also gilt

$$P_X\left(\bigcup_{n \in \mathbb{N}} A_n\right) = P\left(X^{-1}\left(\bigcup_{n \in \mathbb{N}} A_n\right)\right) = P\left(\bigcup_{n \in \mathbb{N}} X^{-1}(A_n)\right) = \sum_{n \in \mathbb{N}} P(X^{-1}(A_n)) = \sum_{n \in \mathbb{N}} P_X(A_n).$$

□

Beispiel: Eine Zufallsvariable beschreibe die Anzahl der Einsen bei n Münzwürfen. Da

$$P(Z = k) = P(Z^{-1}(\{k\})) = B_{n, \frac{1}{2}}(\{k\})$$

und eine diskrete Verteilung durch ihre Zähldichte eindeutig festgelegt ist, folgt

$$P_Z = B_{n, \frac{1}{2}}.$$

Die Zufallsvariable Z („Anzahl der Einsen“) ist $B_{n, \frac{1}{2}}$ -verteilt.

3.4 Definition und Satz:

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\tilde{\Omega}, \tilde{\mathcal{A}})$ ein Messraum und $X : \Omega \rightarrow \tilde{\Omega}$ eine Zufallsvariable. Dann ist

$$\sigma(X) := \{X^{-1}(A) | A \in \tilde{\mathcal{A}}\} \subseteq \mathcal{A}$$

eine σ -Algebra. Sie heißt die von X erzeugte σ -Algebra.

[Bemerkung: Sie besteht aus Ereignissen der Form: „Die Zufallsvariable X nimmt Ereignisse in $A \in \tilde{\mathcal{A}}$ an“.]

Beweis. Zeige, dass $\sigma(X)$ eine σ -Algebra ist:

(A1) $\Omega = X^{-1}(\tilde{\Omega}) \in \sigma(X)$

(A2) Für $X^{-1}(A) \in \sigma(X)$ folgt

$$\Omega \setminus X^{-1}(A) = X^{-1}\left(\underbrace{\Omega \setminus A}_{\in \tilde{\mathcal{A}}}\right) \in \sigma(X).$$

($\sigma 2$) Für $(X^{-1}(A_n))_{n \in \mathbb{N}} \subseteq \sigma(X)$ folgt ⁷

$$\bigcup_{n \in \mathbb{N}} X^{-1}(A_n) = X^{-1}\left(\underbrace{\bigcup_{n \in \mathbb{N}} A_n}_{\in \tilde{\mathcal{A}}}\right) \in \sigma(X).$$

□

3.5 Definition:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $I \neq \emptyset$ eine beliebige Indexmenge, $(\Omega_i, \mathcal{A}_i)$ Messräume und $X_i : \Omega \rightarrow \Omega_i$ Zufallsvariablen für alle $i \in I$. Die Zufallsvariablen $(X_i)_{i \in I}$ heißen unabhängig, falls die davon erzeugten σ -Algebren $(\sigma(X_i))_{i \in I} \subseteq \mathcal{A}$ unabhängig sind.

Beispiel: a) Ein n -facher Münzwurf sei durch den Laplace'schen Wahrscheinlichkeitsraum

$$(\Omega = \{0, 1\}^n, \mathcal{A} = \mathcal{P}(\Omega), P)$$

beschrieben, wobei P die Gleichverteilung auf Ω sei. Sei nun die Indexmenge $I := \{1, \dots, n\}$. Wähle für $i \in I$ die Messräume $(\Omega_i = \{0, 1\}, \mathcal{A}_i = \mathcal{P}(\Omega_i))$ und die Zufallsvariablen

$$X_i : \Omega \rightarrow \Omega_i, (\omega_1, \dots, \omega_n) \mapsto \omega_i.$$

Die Zufallsvariable X_i beschreibt das Ereignis des i -ten Wurfs. Die Zufallsvariablen X_1, \dots, X_n sind unabhängig, denn die von ihnen erzeugten σ -Algebren sind unabhängig. Um dies einzusehen betrachte

$$\sigma(X_i) = \{X^{-1}(A) \mid A \subseteq \{0, 1\}\} = \{\{0, 1\} \times \dots \times \underbrace{\{0, 1\}}_{i\text{-te Komponente}} \times \{0, 1\} \times \dots \times \{0, 1\} \mid A \subseteq \{0, 1\}\}.$$

Nach 2.13 sind also $\sigma(X_1), \dots, \sigma(X_n)$ unabhängig.

b) Man betrachte das Polyasche Urnenmodell 2.15 für $s = w = 1$, $c = -1$ und $n = 2$. Das heißt man betrachtet das Modell zweimaliges Ziehen ohne Zurücklegen. Man hat dabei den Wahrscheinlichkeitsraum

$$(\Omega = \{\text{schwarz, weiß}\}^2, \mathcal{A} = \mathcal{P}(\Omega), P),$$

wobei P das Wahrscheinlichkeitsmaß auf Ω mit der Zähldichte

$$f((\text{schwarz, schwarz})) = f((\text{weiß, weiß})) = 0$$

⁷ Ende der siebten Vorlesung vom Fr, 02.05.14

und

$$f((\text{weiß}, \text{schwarz})) = f((\text{schwarz}, \text{weiß})) = \frac{1}{2}$$

ist. Für $i \in \{1, 2\}$ sind

$$X_i : \Omega \rightarrow \{\text{schwarz}, \text{weiß}\}, (\omega_1, \omega_2) \mapsto \omega_i$$

nicht unabhängige Zufallsvariablen. Betrachte dazu $A_1 = A_2 := \{\text{schwarz}\} \subseteq \{\text{schwarz}, \text{weiß}\}$. Dann gilt

$$X_1^{-1}(A_1) = \{\text{im ersten Versuch schwarz}\} = \{(\text{schwarz}, \text{weiß}), (\text{schwarz}, \text{schwarz})\}$$

und

$$X_2^{-1}(A_2) = \{\text{im zweiten Versuch schwarz}\} = \{(\text{weiß}, \text{schwarz}), (\text{schwarz}, \text{schwarz})\}.$$

$$P(X_1^{-1}(A_1) \cap X_2^{-1}(A_2)) = P(\{(\text{schwarz}, \text{schwarz})\}) = 0 \neq \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(X_1^{-1}(A_1)) \cdot P(X_2^{-1}(A_2)).$$

3.6 Lemma:

Seien $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$, $(\Omega_3, \mathcal{A}_3)$ Messräume und $X_1 : \Omega_1 \rightarrow \Omega_2$, $X_2 : \Omega_2 \rightarrow \Omega_3$ Abbildungen. Dann gilt

a) Sind X_1 und X_2 messbar, so ist auch $X_2 \circ X_1$ wieder messbar.

b) Sei $\mathcal{F} \subseteq \mathcal{A}_2$ mit $\sigma(\mathcal{F}) = \mathcal{A}_2$, dann ist X_1 messbar, falls für alle $A \in \mathcal{F}$ stets $X_1^{-1}(A) \in \mathcal{A}_1$ gilt.

Beweis.

a) Für alle $A \in \mathcal{A}_3$ gilt

$$(X_2 \circ X_1)^{-1}(A) = (X_1^{-1} \circ X_2^{-1})(A) = X_1^{-1}(\underbrace{X_2^{-1}(A)}_{\in \mathcal{A}_2}) \in \mathcal{A}_1 \Rightarrow X_2 \circ X_1 \text{ ist messbar.}$$

b) Man zeigt leicht, dass $\tilde{\mathcal{A}} := \{A \subseteq \Omega_2 \mid X_1^{-1}(A) \in \mathcal{A}_1\}$ eine σ -Algebra ist. Per Annahme ist dann $\mathcal{F} \subseteq \tilde{\mathcal{A}}$. Dann folgt aber sofort, da $\tilde{\mathcal{A}}$ σ -Algebra ist, dass auch $\sigma(\mathcal{F}) \subseteq \tilde{\mathcal{A}}$ ist. Nach Voraussetzung ist dann auch $\mathcal{A}_2 \subseteq \tilde{\mathcal{A}}$ und es gilt für alle $A \in \mathcal{A}_2$, dass $X_1^{-1}(A) \in \mathcal{A}_1$ ist. Also ist X_1 messbar. \square

3.7 Definition:

Seien $(\Omega_i, \mathcal{A}_i)$ Messräume für $i = 1, \dots, n$. Dann wird durch

$$\mathcal{A}_P := \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n := \sigma(\{A_1 \times \dots \times A_n \mid A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n\})$$

die sogenannte *Produkt σ -Algebra* auf $\Omega_P := \Omega_1 \times \dots \times \Omega_n$ definiert.

Beispiel: Seien Ω_i höchstens abzählbar und $\mathcal{A}_i = \mathcal{P}(\Omega_i)$ für $i = 1, \dots, n$. Dann ist auch $\Omega_P := \Omega_1 \times \dots \times \Omega_n$ höchstens abzählbar und es gilt

$$\mathcal{A}_P = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n = \mathcal{P}(\Omega_1 \times \dots \times \Omega_n) = \mathcal{P}(\Omega_P), \text{ denn}$$

„ \subseteq “: klar per Definition.

„ \supseteq “: Seien $\omega_1 \in \Omega_1, \dots, \omega_n \in \Omega_n$, dann folgt

$$\{(\omega_1, \dots, \omega_n)\} = \underbrace{\{\omega_1\}}_{\in \mathcal{A}_1} \times \dots \times \underbrace{\{\omega_n\}}_{\in \mathcal{A}_n} \in \mathcal{A}_P$$

Also gehört jede einelementige Teilmenge von Ω_P zu \mathcal{A}_P . Damit folgt die Behauptung.
Konvention: Wenn nicht anders vermerkt sei Ω_P stets mit der Produkt σ -Algebra \mathcal{A}_P versehen.

3.8 Lemma:

Seien (Ω, \mathcal{A}) ein Messraum und $(\Omega_1, \mathcal{A}_1), \dots, (\Omega_n, \mathcal{A}_n)$ Messräume. Dann gilt:

a) Für $i = 1, \dots, n$ ist die Abbildung

$$p_i : \Omega_1 \times \dots \times \Omega_n \rightarrow \Omega_i, (\omega_1, \dots, \omega_n) \mapsto \omega_i$$

messbar.

b) Die Abbildungen $X_i : \Omega \rightarrow \Omega_i$ für $i = 1, \dots, n$ sind genau dann messbar, wenn

$$X := (X_1, \dots, X_n) : \Omega \rightarrow \Omega_1 \times \dots \times \Omega_n, X(\omega) := (X_1, \dots, X_n)(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

messbar ist.

Beweis.

a) Für $A_i \in \mathcal{A}_i$ und $i = 1, \dots, n$ gilt

$$p_i^{-1}(A_i) = \Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n \in \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n.$$

b) Es sind zwei Implikationen zu zeigen:

„ \Leftarrow “: Sei X messbar, dann folgt sofort wegen Teil a), dass auch $X_i = p_i \circ X$ messbar ist für $i = 1, \dots, n$.

„ \Rightarrow “: Für $i = 1, \dots, n$ seien die X_i messbar. Definiere das Mengensystem

$$\mathcal{F} := \{A_1 \times \dots \times A_n \mid A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n\}.$$

Dann ist per Definition $\sigma(\mathcal{F}) = \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$. Ferner gilt für $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$

$$X^{-1}(A_1 \times \dots \times A_n) = X_1^{-1}(A_1) \cap \dots \cap X_n^{-1}(A_n) \in \mathcal{A}.$$

Es gilt also $\forall A \in \mathcal{F}$, dass $X^{-1}(A) \in \mathcal{A}$ ist. Also ist X nach Lemma 3.6 messbar. \square

3.9 Definition: (Gemeinsame Verteilung)

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\Omega_i, \mathcal{A}_i)$ Messräume und $X_i : \Omega \rightarrow \Omega_i$ Zufallsvariablen für alle $i = 1, \dots, n$.

Nach Lemma 3.8 b) ist dann

$$X := (X_1, \dots, X_n) : \Omega \rightarrow \Omega_1 \times \dots \times \Omega_n, \quad X(\omega) := (X_1, \dots, X_n)(\omega) = (X_1(\omega), \dots, X_n(\omega))$$

eine Zufallsvariable. (X heißt auch Zufallsvektor, ist aber kein Vektor im Sinne der linearen Algebra).

Die Verteilung P_X von X ist ein Wahrscheinlichkeitsmaß auf

$$(\Omega_P := \Omega_1 \times \dots \times \Omega_n, \mathcal{A}_P := \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n).$$

Diese Verteilung heißt auch die *gemeinsame Verteilung* von X_1, \dots, X_n .

Beispiel: Drei Spielkarten mit den Ziffern 1, 2, 3 werden in zufälliger Reihenfolge auf die Plätze 1, 2, 3 verteilt. Die Menge aller möglichen Anordnungen ist dann

$$\Omega := \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}.$$

Das Experiment wird also durch den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A} = \mathcal{P}(\Omega), P)$ beschrieben, wobei P die Gleichverteilung auf Ω ist.

Für $i = 1, 2, 3$ beschreibe die Zufallsvariable $X_i : \Omega \rightarrow \{1, 2, 3\}$ die Ziffer der Karte auf Position i .

Die Zufallsvariablen X_1, X_2, X_3 sind offenbar auf $\{1, 2, 3\}$ gleichverteilt. Das heißt: $P_{X_1} = P_{X_2} = P_{X_3}$ ist die Gleichverteilung auf $\{1, 2, 3\}$ mit der Zähldichte $f(k) = \frac{1}{3}$ für $k = 1, 2, 3$.

Was ist nun die gemeinsame Verteilung von X_1 und X_2 ?

Die gemeinsame Verteilung von X_1 und X_2

$$X := (X_1, X_2)$$

nimmt Werte in $\{1, 2, 3\}^2$ an. P_X ist also das Wahrscheinlichkeitsmaß auf $\{1, 2, 3\}^2$ mit der Zähldichte

$$\begin{aligned} f((1, 2)) &= f((1, 3)) = f((2, 1)) = f((2, 3)) = f((3, 1)) = f((3, 2)) = \frac{1}{6} \\ f((1, 1)) &= f((2, 2)) = f((3, 3)) = 0. \end{aligned}$$

P_x ist also offensichtlich nicht die Gleichverteilung auf $\{1, 2, 3\}^2$.

(Die Ursache dafür ist, dass X_1 und X_2 abhängig sind.)

3.10 Satz:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\Omega_i, \mathcal{A}_i)$ Messräume und $X_i : \Omega \rightarrow \Omega_i$ Zufallsvariablen für alle $i = 1, \dots, n$.

Dann sind die Zufallsvariablen X_i für $i = 1, \dots, n$ genau dann unabhängig, wenn für alle $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$ stets

$$P_X(A_1 \times \dots \times A_n) = P_{X_1}(A_1) \cdots P_{X_n}(A_n) \quad (15)$$

gilt.

Beweis.

Es sind zwei Implikationen zu zeigen:

„ \Rightarrow “: Für X_1, \dots, X_n unabhängig gilt

$$P_X(A_1 \times \dots \times A_n) = P(X^{-1}(A_1 \times \dots \times A_n)) = P(\underbrace{X_1^{-1}(A_1)}_{\in \sigma(X_1)} \cap \dots \cap \underbrace{X_n^{-1}(A_n)}_{\in \sigma(X_n)}) = P_{X_1}(A_1) \cdots P_{X_n}(A_n).$$

„ \Leftarrow “: Die Rechnung aus „ \Rightarrow “ zeigt für alle $A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n$

$$P(\underbrace{X_1^{-1}(A_1)}_{\in \sigma(X_1)} \cap \dots \cap \underbrace{X_n^{-1}(A_n)}_{\in \sigma(X_n)}) = P_{X_1}(A_1) \cdots P_{X_n}(A_n).$$

Also sind $\sigma(X_1), \dots, \sigma(X_n)$ unabhängig. Das bedeutet, dass auch X_1, \dots, X_n unabhängig sind.⁸ \square

3.11 Definition und Satz:

Seien $(\Omega_k, \mathcal{A}_k, P_k)$ Wahrscheinlichkeitsräume $k = 1, \dots, n$.

Dann existiert genau ein Wahrscheinlichkeitsmaß

$$P_P := P_1 \times \dots \times P_n \text{ auf } (\Omega_P := \Omega_1 \times \dots \times \Omega_n, \mathcal{A}_P := \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n)$$

mir der Eigenschaft

$$\forall A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n : P_P(A_1 \times \dots \times A_n) = P_1(A_1) \cdots P_n(A_n).$$

Dieses Wahrscheinlichkeitsmaß P_P heißt *Produktwahrscheinlichkeitsmaß* der Wahrscheinlichkeitsmaße P_1, \dots, P_n und $(\Omega_P, \mathcal{A}_P, P_P)$ heißt *Produktwahrscheinlichkeitsraum*.

⁸ Ende der achten Vorlesung vom Mi, 07.05.14

Beweis.

a) Betrachte die Menge $\mathcal{F} := \{A_1 \times \dots \times A_n \mid A_1 \in \mathcal{A}_1, \dots, A_n \in \mathcal{A}_n\}$. Dann ist offenbar $\sigma(\mathcal{F}) = \mathcal{A}_P$ und da σ -Algebren Durchschnittstabil sind, folgt für $A, B \in \mathcal{F}$

$$A \cap B = (A_1 \times \dots \times A_n) \cap (B_1 \times \dots \times B_n) = \underbrace{(A_1 \cap B_1)}_{\in \mathcal{A}_1} \times \dots \times \underbrace{(A_n \cap B_n)}_{\in \mathcal{A}_n} \in \mathcal{F}$$

Das Mengensystem \mathcal{F} ist also Durchschnittstabil. Nach Voraussetzung ist nun P_P eindeutig auf \mathcal{F} festgelegt und die Behauptung folgt nun aus dem folgenden Lemma 3.12.

b) Die Existenz folgt aus dem Maßerweiterungssatz von Carathéodory (Inhalt der Stochastik 2 Vorlesung). \square

3.12 Lemma:

Seien (Ω, \mathcal{A}) ein Messraum, $\mathcal{F} \subseteq \mathcal{A}$ eine Durchschnittstabile Teilmenge von \mathcal{A} mit der Eigenschaft $\sigma(\mathcal{F}) = \mathcal{A}$ und P_1, P_2 Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) , sodass für alle $A \in \mathcal{F}$ stets $P_1(A) = P_2(A)$ gilt, dann folgt, dass sogar für alle $A \in \mathcal{A}$ stets $P_1(A) = P_2(A)$ gilt, also $P_1 = P_2$ ist.

Beweis.

Man zeigt leicht, dass $\mathcal{D} := \{A \in \mathcal{A} \mid P_1(A) = P_2(A)\} \subseteq \mathcal{A}$ ein Dynkinsystem ist. Nach Voraussetzung ist $\mathcal{F} \subseteq \mathcal{D}$, also auch $\mathcal{D}(\mathcal{F}) \subseteq \mathcal{D}$. Da \mathcal{F} Durchschnittstabil ist, folgt aus Lemma 2.11 sofort $\mathcal{D}(\mathcal{F}) = \sigma(\mathcal{F}) \stackrel{\text{Voraussetzung}}{=} \mathcal{A}$. Es folgt also auch $\mathcal{A} \subseteq \mathcal{D}$. Damit ergibt sich insgesamt $\mathcal{A} = \mathcal{D}$. \square

Mit der Begriffsbildung aus 3.11 lässt sich der Satz 3.10 umformulieren:

3.13 Satz:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\Omega_i, \mathcal{A}_i)$ Messräume und $X_i : \Omega \rightarrow \Omega_i$ Zufallsvariablen für alle $i = 1, \dots, n$.

Dann sind die Zufallsvariablen X_i für $i = 1, \dots, n$ genau dann unabhängig, wenn $P_X = P_P$ ist.

4 Verteilungen auf \mathbb{R} und \mathbb{R}^d

4.1 Satz:

Die Mengensystemen

- a) $\mathcal{O} := \{A \subseteq \mathbb{R} \mid A \text{ offen}\}$
- b) $\mathcal{A} := \{A \subseteq \mathbb{R} \mid A \text{ abgeschlossen}\}$
- c) $\mathcal{K} := \{A \subseteq \mathbb{R} \mid A \text{ kompakt}\}$
- d) $\mathcal{J} := \{A \subseteq \mathbb{R} \mid A \text{ Intervall}\}$
- e) $\mathcal{E} := \{(-\infty, x] \subseteq \mathbb{R} \mid x \in \mathbb{R}\}$

erzeugen die selbe σ -Algebra.

Beweis.

Man verwendet im folgenden für $\mathcal{C}_1, \mathcal{C}_2 \in \{\mathcal{O}, \mathcal{A}, \mathcal{K}, \mathcal{J}, \mathcal{E}\}$ das Argument

$$\mathcal{C}_i \subseteq \sigma(\mathcal{C}_j) \Rightarrow \sigma(\mathcal{C}_i) \subseteq \sigma(\mathcal{C}_j), \quad i, j \in \{1, 2\}. \quad (16)$$

a) „ \subseteq “: Ist $A \in \mathcal{O}$ offen, dann ist das Komplement $\mathbb{R} \setminus A \in \mathcal{A}$ abgeschlossen. Also folgt aus (A2) sofort $A \in \sigma(\mathcal{A})$, also $\mathcal{O} \subseteq \sigma(\mathcal{A})$. Mit Argument (16) folgt dann $\sigma(\mathcal{O}) \subseteq \sigma(\mathcal{A})$.

„ \supseteq “: Ist $A \in \mathcal{A}$ abgeschlossen, dann ist das Komplement $\mathbb{R} \setminus A \in \mathcal{O}$ offen. Also folgt aus (A2) sofort $A \in \sigma(\mathcal{O})$, also $\mathcal{A} \subseteq \sigma(\mathcal{O})$. Mit Argument (16) folgt dann $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{O})$.

Insgesamt folgt also $\sigma(\mathcal{A}) = \sigma(\mathcal{O})$.

b) „ \subseteq “: Offensichtlich ist $\mathcal{K} \subseteq \mathcal{A}$, also auch $\sigma(\mathcal{K}) \subseteq \sigma(\mathcal{A})$.

„ \supseteq “: Ist $A \in \mathcal{A}$ abgeschlossen, dann ist

$$A = A \cap \mathbb{R} = A \cap \bigcup_{n \in \mathbb{N}} (A \cap [-n, n]) = \bigcup_{n \in \mathbb{N}} \underbrace{(A \cap [-n, n])}_{\in \mathcal{K}} \in \sigma(\mathcal{K}).$$

Also folgt aus (σ 2) sofort $A \in \sigma(\mathcal{K})$, also $\mathcal{A} \subseteq \sigma(\mathcal{K})$. Mit Argument (16) folgt dann $\sigma(\mathcal{A}) \subseteq \sigma(\mathcal{K})$.

Insgesamt folgt also $\sigma(\mathcal{A}) = \sigma(\mathcal{K})$.

c) „ \subseteq “: Sei $A \in \mathcal{O}$, dann ist

$$A = \bigcup_{\substack{a < b \in \mathbb{Q} \\ (a, b) \in A}} (a, b) \in \sigma(\mathcal{J}).$$

Also folgt aus (σ 2) sofort $A \in \sigma(\mathcal{J})$, also $\mathcal{O} \subseteq \sigma(\mathcal{J})$. Mit Argument (16) folgt dann $\sigma(\mathcal{O}) \subseteq \sigma(\mathcal{J})$.

„ \supseteq “: Ist $A \in \mathcal{J}$ ein Intervall. Falls A offen oder abgeschlossen ist, ist nach a) und b) sofort klar $A \in \sigma(\mathcal{A}) = \sigma(\mathcal{O})$. Für halboffene Intervalle ergibt sich

$$A = [a, b) = \bigcup_{n \in \mathbb{N}} \left[a, b - \frac{1}{n} \right) \in \sigma(\mathcal{A}) = \sigma(\mathcal{O})$$

und

$$A = (a, b] = \bigcup_{n \in \mathbb{N}} \left[a - \frac{1}{n}, b \right] \in \sigma(\mathcal{A}) = \sigma(\mathcal{O}).$$

Also folgt aus ($\sigma 2$) sofort $A \in \sigma(\mathcal{O})$, also $\mathcal{J} \subseteq \sigma(\mathcal{O})$. Mit Argument (16) folgt dann $\sigma(\mathcal{J}) \subseteq \sigma(\mathcal{O})$.

Insgesamt folgt also $\sigma(\mathcal{J}) = \sigma(\mathcal{O})$.

d) „ \subseteq “: Offensichtlich ist $\mathcal{E} \subseteq \mathcal{J}$, also auch $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{J})$.

„ \supseteq “: Ist $A \in \mathcal{J}$ und $a < b$, dann ist

$$A = (a, b) = (-\infty, b] \setminus (-\infty, a] \in \sigma(\mathcal{E}),$$

$$A = [a, b) = (-\infty, b) \setminus (-\infty, a) \in \sigma(\mathcal{E}),$$

$$A = [a, b] = (-\infty, b] \setminus (-\infty, a) \in \sigma(\mathcal{E}),$$

und

$$A = (a, b) = (-\infty, b) \setminus (-\infty, a] \in \sigma(\mathcal{E}),$$

Also folgt aus (A2) sofort $A \in \sigma(\mathcal{E})$, also $\mathcal{J} \subseteq \sigma(\mathcal{E})$. Mit Argument (16) folgt dann $\sigma(\mathcal{J}) \subseteq \sigma(\mathcal{E})$.

Insgesamt folgt also $\sigma(\mathcal{J}) = \sigma(\mathcal{E})$. □

4.2 Definition:

a) Sei $d \in \mathbb{N}$. Die σ -Algebra

$$\mathcal{B}(\mathbb{R}^d) := \sigma(\{A \subseteq \mathbb{R}^d \mid A \text{ offen}\})$$

heißt *Borel σ -Algebra* auf \mathbb{R}^d . Die Mengen $A \in \mathcal{B}(\mathbb{R}^d)$ heißen *Borelmengen* in \mathbb{R}^d .

b) Sei $B \in \mathcal{B}(\mathbb{R}^d)$ eine Borelmenge. Die Menge

$$\mathcal{B}(B) := \{A \cap B \mid A \in \mathcal{B}(\mathbb{R}^d)\}$$

heißt die *Borel σ -Algebra* auf B .

Bemerkungen: a) Nach Satz 4.1 sind alle offenen bzw. abgeschlossenen Mengen und alle Intervalle Borelmengen.

b) Man verwendet auf \mathbb{R} bzw. \mathbb{R}^d nicht die Potenzmenge als σ -Algebra, denn darauf lässt sich kein Wahrscheinlichkeitsmaß mit den gewünschten Eigenschaften definieren.

c) Man findet keine „einfache“ Menge, die keine Borelmenge ist. Trotzdem gibt es solche Mengen.

4.3 Beispiel: (Vitali-Menge)

a) Man möchte auf $[0, 1]$ gerne ein translationsinvariantes Wahrscheinlichkeitsmaß haben, d.h. ein Wahrscheinlichkeitsmaß P , das

$$\forall A \subseteq [0, 1] \forall x \in [0, 1] \text{ mit } A+x := \{a+x | a \in A\} \subseteq [0, 1] : P(A+x) = P(A) \quad (17)$$

erfüllt. (Das Lebesgue-Maß λ besitzt diese Eigenschaft.)

b) Verwendet man als σ -Algebra die Potenzmenge $\mathcal{P}([0, 1])$, so findet man kein solches Wahrscheinlichkeitsmaß. Betrachte dazu auf \mathbb{R} die Äquivalenzrelation

$$x \sim y : \Leftrightarrow x - y \in \mathbb{Q}.$$

Sei ferner $E \subseteq [\frac{1}{3}, \frac{2}{3}]$ ein Repräsentantensystem der Äquivalenzklassen von \sim . (Diese Definition ist wegen des *Auswahlaxioms der Mengenlehre* möglich.)

Dann gilt

$$[0, 1] \supseteq \bigcup_{r \in \mathbb{Q} \cap [0, \frac{1}{3}]} (E+r) \Rightarrow 1 = P([0, 1]) \geq P\left(\bigcup_{r \in \mathbb{Q} \cap [0, \frac{1}{3}]} (E+r)\right) = \sum_{r \in \mathbb{Q} \cap [0, \frac{1}{3}]} P(E+r) \stackrel{(17)}{=} \sum_{r \in \mathbb{Q} \cap [0, \frac{1}{3}]} P(E)$$

Damit $\sum_{r \in \mathbb{Q} \cap [0, \frac{1}{3}]} P(E)$ endlich ist, muss $P(E) = 0$ gelten. Dann folgt wieder mit (17), dass

auch $P(E+r) = 0$ sein muss für alle $r \in \mathbb{Q} \cap [-\frac{1}{3}, \frac{1}{3}]$.

Andererseits hat man $[\frac{1}{3}, \frac{2}{3}] \subseteq \bigcup_{r \in \mathbb{Q} \cap [-\frac{1}{3}, \frac{1}{3}]} (E+r)$ und deshalb

$$0 \leq P\left([\frac{1}{3}, \frac{2}{3}]\right) \leq P\left(\bigcup_{r \in \mathbb{Q} \cap [-\frac{1}{3}, \frac{1}{3}]} (E+r)\right) = 0.$$

Dann ergibt sich wieder mit (17)

$$P\left([0, \frac{1}{3}]\right) = P\left([\frac{1}{3}, \frac{2}{3}]\right) = P\left([\frac{2}{3}, 1]\right) = 0$$

also auch

$$P([0, 1]) = 0. \not\checkmark$$

Man findet also auf $([0, 1], \mathcal{P}([0, 1]))$ kein Wahrscheinlichkeitsmaß mit Eigenschaft (17). Die hier konstruierte Menge E heißt *Vitali-Menge*. Sie ist insbesondere keine Borelmenge.

4.4 Lemma:

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Dann gilt

a) für eine Folge $(A_n) \subseteq \mathcal{A}$ mit $A_{n+1} \subseteq A_n$ für alle $n \in \mathbb{N}$ folgt

$$P\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

b) für eine Folge $(A_n) \subseteq \mathcal{A}$ mit $A_{n+1} \supseteq A_n$ für alle $n \in \mathbb{N}$ folgt

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

Beweis.

b) Setze $A := \bigcup_{n \in \mathbb{N}} A_n = A_n \cup \left(\bigcup_{k \geq n} A_{k+1} \setminus A_k \right)$. Dann erhält man für alle $n \in \mathbb{N}$

$$P(A) = P(A_n) + P\left(\bigcup_{k \geq n} (A_{k+1} \setminus A_k)\right) = P(A_n) + \sum_{k \geq n} P(A_{k+1} \setminus A_k)$$

Die Folge der Zahlen $\sum_{k \geq n} P(A_{k+1} \setminus A_k)$ ist offenbar monoton fallend und beschränkt in $[0, 1]$ und konvergiert dort gegen 0. Insbesondere folgt also

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

a) Es folgt sofort aus Teil b) ⁹

$$\begin{aligned} P\left(\bigcap_{n \in \mathbb{N}} A_n\right) &= 1 - P\left(\Omega \setminus \bigcup_{n \in \mathbb{N}} A_n\right) = 1 - P\left(\bigcup_{n \in \mathbb{N}} (\Omega \setminus A_n)\right) = 1 - \lim_{n \rightarrow \infty} P(\Omega \setminus A_n) = 1 - (1 - \lim_{n \rightarrow \infty} P(A_n)) \\ &= \lim_{n \rightarrow \infty} P(A_n). \end{aligned}$$

□

4.5 Definition und Satz: (Verteilungsfunktion)

Sei P ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

a) Die Funktion $F_P : \mathbb{R} \rightarrow [0, 1]$, $F_P(x) := P((-\infty, x])$ besitzt die folgenden Eigenschaften

$$F_P \text{ ist monoton wachsend, rechtsstetig und } \lim_{x \rightarrow \infty} F_P(x) = 1, \lim_{x \rightarrow -\infty} F_P(x) = 0. \quad (18)$$

Die Funktion F_P heißt *Verteilungsfunktion* vom Wahrscheinlichkeitsmaß P .

b) Umgekehrt gilt: Ist $F : \mathbb{R} \rightarrow [0, 1]$ eine Funktion mit den Eigenschaften aus Gleichung (18), dann existiert ein eindeutiges Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $F = F_P$.

Beweis.

a) Sei P ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $F_P : \mathbb{R} \rightarrow [0, 1]$, $F_P(x) := P((-\infty, x])$ die zugehörige Verteilungsfunktion. Dann ist:

(1) F_P monoton wachsend, denn für $x_1 \leq x_2 \in \mathbb{R}$ gilt

$$F_P(x_1) = P((-\infty, x_1]) \leq P((-\infty, x_2]) = F_P(x_2).$$

(2) F_P rechtsstetig, denn sei $(x_n) \subseteq \mathbb{R}$ mit $x_n \rightarrow x \in \mathbb{R}$ von oben für $n \rightarrow \infty$, dann gilt

$$(-\infty, x_1] \supseteq (-\infty, x_2] \supseteq (-\infty, x_3] \supseteq \dots \text{ und } (-\infty, x] = \bigcap_{n \in \mathbb{N}} (-\infty, x_n].$$

⁹ Ende der neunten Vorlesung vom Fr, 09.05.14

Deshalb gilt

$$\lim_{n \rightarrow \infty} F_P(x_n) = \lim_{n \rightarrow \infty} P((-\infty, x_n]) \stackrel{4.4^a)}{=} P((-\infty, x]) = F_P(x).$$

$$(3) \lim_{x \rightarrow \infty} F_P(x) = \lim_{x \rightarrow \infty} P((-\infty, x]) \stackrel{4.4^b)}{=} P(\mathbb{R}) = 1.$$

$$(4) \lim_{x \rightarrow -\infty} F_P(x) = \lim_{x \rightarrow -\infty} P((-\infty, x]) \stackrel{4.4^a)}{=} P(\emptyset) = 0.$$

b) Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine Funktion mit den Eigenschaften aus Gleichung (18). Zeige zunächst die Eindeutigkeit:

Seien P_1, P_2 zwei Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit $F_{P_1} = F_{P_2}$. Dann folgt für alle $x \in \mathbb{R}$

$$P_1((-\infty, x]) = F_{P_1}(x) = F_{P_2}(x) = P_2((-\infty, x]).$$

Also folgt $P_1 = P_2$ auf $\mathcal{E} = \{(-\infty, x] | x \in \mathbb{R}\}$. Offenbar ist \mathcal{E} durchschnittstabil und wegen Satz 4.1 gilt $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R})$. Mit Lemma 3.12 folgt dann $P_1 = P_2$ auf ganz $\mathcal{B}(\mathbb{R})$.

Die Existenz folgt aus dem Maßerweiterungssatz von Carathéodory (Inhalt der Stochastik 2 Vorlesung). \square

Bemerkung: 4.5 bedeutet, dass es eine Bijektion zwischen der Menge

$$\mathcal{M}^1(\mathbb{R}) := \{P : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1] | P \text{ ist Wahrscheinlichkeitsmaß auf } (\mathbb{R}, \mathcal{B}(\mathbb{R}))\}$$

und der Menge aller Funktionen, die die Eigenschaften aus Gleichung (18) besitzen, gibt.

4.6 Beispiel: (Verteilungsfunktionen von diskreten Verteilungen)

Sei $\phi \neq \Omega \subseteq \mathbb{R}$ höchstens abzählbar. Dann ist

$$\mathcal{B}(\Omega) = \{\Omega \cap A | A \in \mathcal{B}(\mathbb{R})\} = \mathcal{P}(\Omega).$$

Jedes Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{P}(\Omega))$ ist nach Kapitel 1 durch seine Zähldichte $f_P : \Omega \rightarrow [0, 1]$ eindeutig bestimmt, d.h.

$$\forall A \subseteq \Omega : P(A) = \sum_{\omega \in A} f_P(\omega).$$

Das Wahrscheinlichkeitsmaß P auf Ω definiert durch

$$\tilde{P}(A) := P(A \cap \Omega) \quad \forall A \in \mathcal{B}(\mathbb{R})$$

sofort ein Wahrscheinlichkeitsmaß \tilde{P} auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Man identifiziert nun $\tilde{P} \in \mathcal{M}^1(\mathbb{R})$ mit P auf $(\Omega, \mathcal{P}(\Omega))$, indem man $P = \tilde{P}$ schreibt. Betrachte nun die zu $P \in \mathcal{M}^1(\mathbb{R})$ assoziierte Verteilungsfunktion F_P . Die Funktion F_P ist dann eine monotone rechtsstetige Treppenfunktion mit Sprungstellen in $\omega \in \Omega$ und Sprunghöhe $f_P(\omega)$.

Bsp: Sei $P = B_{3, \frac{1}{2}}$ die Binomialverteilung auf $\Omega = \{0, 1, 2, 3\}$ mit Zähldichte $f_P(0) = \frac{1}{8} = f_P(3)$ und $f_P(1) = \frac{3}{8} = f_P(2)$.

4.7 Darstellung allgemeiner Wahrscheinlichkeitsmaße auf diskreten Wahrscheinlichkeitsräumen:

Sei $\phi \neq \Omega \subseteq \mathbb{R}$ höchstens abzählbar mit $\mathcal{P}(\Omega)$ als σ -Algebra. Sei ferner P ein Wahrscheinlichkeitsmaß auf Ω , das eindeutig durch die Zähldichte $f : \Omega \rightarrow [0, 1]$ bestimmt ist, dann ist

$$P = \sum_{\omega \in \Omega} \delta_{\omega} f(\omega).$$

Dabei ist δ_{ω} das Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{P}(\Omega))$ mit

$$\delta_{\omega}(A) := \begin{cases} 1 & , \omega \in A \\ 0 & , \omega \notin A \end{cases}.$$

δ_{ω} heißt Punktmaß oder Dirac-Maß in $\omega \in \Omega$.

Bsp: $\Omega = \{0, 1, \dots, n\}$, dann ist

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

4.8 Kontinuierliche Wahrscheinlichkeitsmaße auf \mathbb{R} mit Dichtefunktionen:

Sei $f : \mathbb{R} \rightarrow [0, \infty)$ eine integrierbare Funktion mit $\int_{\mathbb{R}} f(x) dx = 1$. Dann wird für alle $t \in \mathbb{R}$ durch

$$F(t) := \int_{(-\infty, t]} f(x) dx$$

eine monotone und stetige Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit $\lim_{t \rightarrow \infty} F(t) = 1$ und $\lim_{t \rightarrow -\infty} F(t) = 0$ definiert. Die Funktion F besitzt also insbesondere die Eigenschaften aus Gleichung (18) und daher existiert nach 4.5 genau ein zu F assoziiertes Wahrscheinlichkeitsmaß P auf \mathbb{R} mit $F = F_P$. P heißt das Wahrscheinlichkeitsmaß auf \mathbb{R} mit der Dichtefunktion f . Für eine stückweise stetige Funktion $f \in \mathcal{C}_{St}(\mathbb{R})$ liefert der Hauptsatz der Differential- und Integralrechnung, dass F fast überall (f.ü.), d.h. außerhalb einer Nullmenge (den Unstetigkeitsstellen von f), eine Stammfunktion von f ist, also gilt $F' = f$ f.ü.

4.9 Beispiel: (Gleichverteilung)

Sei $a < b \in \mathbb{R}$. Betrachte die Gleichverteilung auf $[a, b]$ und definiere die Dichtefunktion

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

Dann ergibt sich sofort

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

Dabei ist es völlig egal, ob die Intervallgrenzen a, b dazu gehören oder nicht, da einzelne Punkte (sogar abzählbare Mengen) Nullmengen sind, die bei der Integration nicht beitragen.

4.10 Exponentialverteilung:

Sei $\lambda > 0$. Definiere die Dichtefunktion

$$f_\lambda(x) := \begin{cases} 0 & , x \leq 0 \\ \lambda \cdot e^{-\lambda x} & , x > 0 \end{cases}.$$

Dann ergibt sich sofort

$$F_\lambda(t) = \int_{(-\infty, t]} f(x) dx = \begin{cases} 0 & , t \leq 0 \\ 1 - e^{-\lambda t} & , t > 0 \end{cases}.$$

Bemerkung: Das zu F_λ gehörende eindeutige Wahrscheinlichkeitsmaß heißt *Exponentialverteilung* mit Parameter λ . Man schreibt häufig $\exp_\lambda \in \mathcal{M}^1(\mathbb{R})$. Die Verteilung \exp_λ ist das kontinuierliche Analogon zur diskreten geometrischen Verteilung auf \mathbb{N}_0 . Beide Verteilungen sind insbesondere gedächtnislos (vgl. Übungen). Die Exponentialverteilung findet häufig Anwendung bei der Modellierung von Prozessen mit kontinuierlichen Wartezeiten.

4.11 Satz: (Grenzwertsatz für geometrische Verteilungen)

Sei $(p_n) \subseteq (0, 1)$ eine Nullfolge und $(X_n) \subseteq \mathbb{R}$ eine Folge von Zufallsvariablen, sodass jedes X_n mit zugehörigem p_n geometrisch verteilt ist für $n \in \mathbb{N}$. Die X_n haben Werte in \mathbb{N}_0 . Dann hat die Zufallsvariable $p_n \cdot X_n$ Werte in $p_n \cdot \mathbb{N}_0 := \{0, p_n, 2p_n, 3p_n, \dots\}$. Sei nun F_n die Verteilungsfunktion der Verteilung von $p_n \cdot X_n$ für $n \in \mathbb{N}$. Also ist für alle $x \in \mathbb{R}$ und $n \in \mathbb{N}$

$$F_n(x) = P_{p_n X_n}((-\infty, x]) = P(p_n X_n \leq x) = P(X_n \leq \frac{x}{p_n}).$$

Dann gilt für alle $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

wobei $F = F_{\lambda=1}$ die Verteilungsfunktion einer \exp_1 -Verteilung ist.

[Kurz gesagt: Die Verteilungsfunktionen von $p_n X_n$ konvergieren also punktweise gegen die Verteilungsfunktion einer \exp_1 -Verteilung.]

Beweis.

a) Für $x < 0$ ist $F_n(x) = 0 = F(x)$ ✓

b) Für $x \geq 0$ ist

$$\begin{aligned} F_n(x) &= P(X_n \leq \frac{x}{p_n}) = \sum_{k=0}^{\lfloor \frac{x}{p_n} \rfloor} p_n (1-p_n)^k = p_n \cdot \sum_{k=0}^{\lfloor \frac{x}{p_n} \rfloor} (1-p_n)^k = p_n \cdot \frac{1 - (1-p_n)^{\lfloor \frac{x}{p_n} \rfloor + 1}}{1 - (1-p_n)} = 1 - (1-p_n)^{\lfloor \frac{x}{p_n} \rfloor + 1} \\ &= 1 - [(1-p_n)^{\frac{1}{p_n}}]^{p_n(\lfloor \frac{x}{p_n} \rfloor + 1)} \rightarrow 1 - e^{-x} = F(x) \text{ für } n \rightarrow \infty. \end{aligned}$$

Dabei bezeichne $\lfloor \cdot \rfloor$ die Gaußklammer, die auf die nächste ganze Zahl abrundet.¹⁰ \square

4.12 Normalverteilungen:

Seien $\mu \in \mathbb{R}$, $\sigma^2 > 0$ Konstanten. Die Normalverteilung $N_{\mu, \sigma^2} \in \mathcal{M}^1(\mathbb{R})$ mit Mittelwert μ und Varianz σ^2 ist das Wahrscheinlichkeitsmaß mit der Dichtefunktion

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

Für $\mu = 0$ und $\sigma^2 = 1$ ergibt sich der wichtige Spezialfall $N_{0,1}$, die so genannte *Standardnormalverteilung* mit der Dichtefunktion

$$f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Wohldefiniertheit: Sind die f_{μ, σ^2} tatsächlich Dichtefunktionen?

a) Offenbar gilt $f_{\mu, \sigma^2}(x) > 0$ für alle $x \in \mathbb{R}$.

b) Betrachte $I := \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$, dann folgt mit der Transformationsformel für $x = r \cos \varphi$ und $y = r \sin \varphi$ sofort

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy = \int_{\mathbb{R}} e^{-\frac{x^2+y^2}{2}} d^2(x, y) \\ &= \int_0^{\infty} \int_{-\pi}^{\pi} e^{-\frac{r^2}{2}} r dr d\varphi = 2\pi [-e^{-\frac{r^2}{2}}]_0^{\infty} = 2\pi. \end{aligned}$$

Also ist $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$. Daraus folgt weiter $\int_{-\infty}^{\infty} f_{0,1}(x) dx = 1$ und für $\mu \in \mathbb{R}$, $\sigma^2 > 0$ ergibt sich ferner mit der Substitution $y = \frac{x-\mu}{\sqrt{\sigma^2}}$

$$\int_{-\infty}^{\infty} f_{\mu, \sigma^2}(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = 1.$$

Bemerkungen:

- Viele Probleme lassen sich mit der Substitution $y = \frac{x-\mu}{\sqrt{\sigma^2}}$ von einer allgemeinen Normalverteilung N_{μ, σ^2} auf die Standardnormalverteilung $N_{0,1}$ zurückführen.
- Die Standardnormalverteilung $N_{0,1}$ hat die zugehörige Verteilungsfunktion

$$\Phi(x) := \int_{-\infty}^x f_{0,1}(t) dt, \quad x \in \mathbb{R}.$$

¹⁰ Ende der zehnten Vorlesung vom Mi, 14.05.14

- Man kann beweisen, dass Φ nicht durch elementare Funktionen (Polynome, sin, cos, exp, ln, ...) ausgedrückt werden kann.
- Φ ist tabelliert für $x \geq 0$, denn aus der Achsensymmetrie (bzgl. der y -Achse) von $f_{0,1}$ folgt

$$\Phi(-x) = 1 - \Phi(x).$$

- Normalverteilungen erhalten ihre besondere Bedeutung durch den so genannten *zentralen Grenzwertsatz*.

Galton-Brett:

An jeder Stelle dieses Bretts sei die Wahrscheinlichkeit für links oder rechts bei einer Kugel $p = \frac{1}{2}$ unabhängig von links bzw. rechts in den vorherigen Schritten. Die Füllung des Auffangbehälters ist dann $B_{n, \frac{1}{2}}$ -verteilt. Was passiert für große n , bzw. $n \rightarrow \infty$?

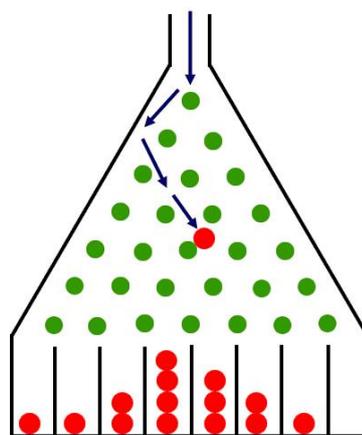


Abbildung 1: Darstellung eines Galton-Bretts

4.13 Stirlingsche Formel:

Die Stirlingsche Formel besagt

$$n! \sim \sqrt{2\pi n} \cdot n^n \cdot e^{-n}, \tag{19}$$

wobei \sim bedeutet, dass

$$\frac{n!}{\sqrt{2\pi n} \cdot n^n \cdot e^{-n}} \rightarrow 1 \text{ für } n \rightarrow \infty.$$

4.14 Beziehung zwischen Binomialverteilung und Normalverteilung:

Sei $p \in (0, 1)$, $n \in \mathbb{N}$, $k \in \{0, \dots, n\}$ mit $n \rightarrow \infty, k \rightarrow \infty, n - k \rightarrow \infty$. Dann gilt

$$\begin{aligned} B_{n,p}(\{k\}) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &\sim \sqrt{\frac{2\pi n}{2\pi k \cdot 2\pi(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{n(1-p)}{n-k}\right)^{n-k} \cdot e^{-n+k+n-k} \\ &= \frac{1}{\sqrt{2\pi n \cdot \frac{k}{n} \cdot \left(\frac{n-k}{n}\right)}} \cdot e^{-n \cdot g\left(\frac{k}{n}\right)}. \end{aligned}$$

Dabei ist

$$\begin{aligned} g(t) &:= t \ln\left(\frac{t}{p}\right) + (1-t) \ln\left(\frac{1-t}{1-p}\right) \text{ und} \\ g'(t) &= \ln\left(\frac{t}{p}\right) - \ln\left(\frac{1-t}{1-p}\right) \text{ und} \\ g''(t) &= \frac{1}{t} + \frac{1}{1-t}. \end{aligned}$$

Also gilt $g(p) = g'(p) = 0$ und $g''(p) = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$. Eine qualitative Taylorentwicklung um p liefert

$$\begin{aligned} g\left(\frac{k}{n}\right) &= 0 + 0 + \frac{1}{2}g''(p)\left(\frac{k}{n} - p\right)^2 + \mathcal{O}\left(\left(\frac{k}{n} - p\right)^3\right) \text{ für } \frac{k}{n} \rightarrow p \\ &= \frac{1}{2p(1-p)}\left(\frac{k}{n} - p\right)^2 + \mathcal{O}\left(\left(\frac{k}{n} - p\right)^3\right) \text{ für } \frac{k}{n} \rightarrow p. \end{aligned}$$

Wähle Parameter nun so, dass $\left(\frac{k}{n} - p\right)^3 \cdot n \rightarrow 0$ für $n \rightarrow \infty$ gefordert werden kann, d.h. es gilt $\frac{k}{n} - p = o\left(\frac{1}{n^{\frac{1}{3}}}\right)$ für $n \rightarrow \infty$. Dann folgt

$$\begin{aligned} B_{n,p}(\{k\}) &= \frac{1}{\sqrt{2\pi np(1-p)}} \cdot e^{-\frac{n\left(\frac{k}{n}-p\right)^2}{2p(1-p)}} \cdot \underbrace{e^{-n\mathcal{O}\left(\left(\frac{k}{n}-p\right)^3\right)}}_{\rightarrow \text{für } n \rightarrow \infty} \\ &\sim \frac{1}{\sqrt{2\pi np(1-p)}} \cdot e^{-\frac{(k-np)^2}{2np(1-p)}} \text{ setze } \sigma^2 := np(1-p) \\ &\sim \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(k-np)^2}{2\sigma^2}} = f_{np, np(1-p)}(k). \end{aligned}$$

4.15 Verteilungen auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$:

Die gemeinsame Verteilung von d \mathbb{R} -wertigen Zufallsvariablen X_1, \dots, X_d ist oft wichtig für die Praxis. Im Folgenden werden diskrete Verteilungen und ausgewählte kontinuierliche Verteilungen vorgestellt.

a) Diskrete Verteilungen auf \mathbb{R}^d : Sei $\phi \neq \Omega \subseteq \mathbb{R}^d$ höchstens abzählbar. Dann ist $\mathcal{B}(\Omega) = \mathcal{P}(\Omega)$. Betrachte nun eine Zähldichte $f : \Omega \rightarrow [0, 1]$ und definiere das zu f gehörende Wahrscheinlichkeitsmaß P auf $(\Omega, \mathcal{P}(\Omega))$; wie im eindimensionalen Fall kann man dann P auch als Wahrscheinlichkeitsmaß auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ auffassen:

$$\forall A \in \mathcal{B}(\mathbb{R}^d) : P(A) = \sum_{\omega \in \Omega \cap A} f(\omega).$$

b) Verteilungen mit (Lebesgue-) Dichtefunktionen: Sei λ^d das Lebesgue-Maß auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Man kann zeigen (Analysis III), dass das Lebesgue-Maß eindeutig dadurch charakterisiert ist, dass es die folgenden Eigenschaften erfüllt:

- Das Lebesgue-Maß $\lambda^d : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, \infty]$ ist σ -additiv, d.h. für eine disjunkte Folge $(A_n) \subseteq \mathcal{B}(\mathbb{R}^d)$ gilt stets

$$\lambda^d\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \lambda^d(A_n).$$

- Für $a_1 \leq b_1, \dots, a_d \leq b_d \in \mathbb{R}$ gilt

$$\lambda^d([a_1, b_1] \times \dots \times [a_d, b_d]) = \prod_{k=1}^d (b_k - a_k).$$

Sei nun $f : \mathbb{R}^d \rightarrow [0, \infty)$ eine Lebesgue-integrierbare Funktion $f \in \mathcal{L}_1(\mathbb{R}^d) := \{f : \mathbb{R}^d \rightarrow \mathbb{R} \mid f \text{ Lebesgue-integrierbar}\}$ mit

$$\int_{\mathbb{R}^d} f d\lambda^d = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_1 \dots dx_d = 1.$$

Solche Funktionen heißen Dichtefunktionen oder Dichten. Zu einer Dichte f existiert ein Wahrscheinlichkeitsmaß $P \in \mathcal{M}^n(\mathbb{R}^d)$ mit

$$P(A) := \int_A f d\lambda^d = \int_{\mathbb{R}^d} \mathbb{1}_A(x) f(x) d\lambda^d(x) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \mathbb{1}_A(x_1, \dots, x_d) f(x_1, \dots, x_d) dx_1 \dots dx_d$$

mit der *charakteristischen Funktion*

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

c) Beispiel: (Direkte Produkte von kontinuierlichen eindimensionalen Verteilungen) Seien $d \in \mathbb{N}$, und $f_1, \dots, f_d : \mathbb{R} \rightarrow [0, \infty)$ uneigentlich-integrierbare \mathcal{C}_{St} -Dichtefunktionen mit

$$\int_{-\infty}^{\infty} f_k(x) dx = 1, \quad k = 1, \dots, d.$$

Definiere nun eine Dichtefunktion $f : \mathbb{R}^d \rightarrow [0, \infty)$ durch:

$$f(x_1, \dots, x_d) := f(x_1) \dots f(x_d).$$

Dann ist $f \in \mathcal{L}_1(\mathbb{R}^d)$ mit

$$\int_{\mathbb{R}^d} f d\lambda^d = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_1 \dots dx_d = \int_{-\infty}^{\infty} f_1(x_1) dx_1 \dots \int_{-\infty}^{\infty} f_d(x_d) dx_d = 1^d = 1.$$

d) Fazit: Die Funktion f ist also die Dichte eines Wahrscheinlichkeitsmaßes $P \in \mathcal{M}^1(\mathbb{R}^d)$. Seien P_1, \dots, P_d die Wahrscheinlichkeitsmaße auf \mathbb{R} , die zu den Zähldichten f_1, \dots, f_d gehören. Dann gilt für

$$A := a_1 \times \dots \times a_d \in \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$$

stets

$$P(A) = \int_A f(x) d\lambda^d(x) = \int_{A_1} f_1(x_1) dx_1 \dots \int_{A_d} f_d(x_d) dx_d = P_1(A_1) \dots P_d(A_d)$$

$$\stackrel{\text{Produktmaß}}{=} P_P(A) = (P_1 \times \dots \times P_d)(A_1 \times \dots \times A_d) = P(A_1 \times \dots \times A_d)$$

auf $\mathcal{E} := \{A_1 \times \dots \times A_d \mid A_1, \dots, A_d \in \mathcal{B}(\mathbb{R})\}$. Das Mengensystem \mathcal{E} ist durchschnittstabil mit $\sigma(\mathcal{E}) = \mathcal{B}(\mathbb{R}^d)$. Damit ergibt sich abschließend $P = P_1 \times \dots \times P_d$.¹¹

4.16 Beispiel: (Verteilungen auf \mathbb{R}^d)

a) Gleichverteilungen auf Teilmengen von \mathbb{R}^d (geometrische Wahrscheinlichkeiten). Sei $A \subseteq \mathbb{R}^d$, $A \in \mathcal{B}(\mathbb{R}^d)$ mit $\lambda^d(A) \in (0, \infty)$. Die sogenannte Gleichverteilung auf A ist definiert als das Wahrscheinlichkeitsmaß P_A auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}^d))$ mit der Dichtefunktion

$$f_A(x) := \frac{1}{\lambda^d(A)} \mathbb{1}_A(x).$$

Dann ist $f_A \in \mathcal{L}_1(\mathbb{R}^d)$ mit

$$\int_A f_A(x) dx = \frac{1}{\lambda^d(A)} \int_A 1 dx = \frac{\lambda^d(A)}{\lambda^d(A)} = 1.$$

Man definiert nun das assoziierte Wahrscheinlichkeitsmaß für alle $B \in \mathcal{B}(\mathbb{R}^d)$ durch

$$P_A(B) := \int_B f_A(x) dx = \frac{1}{\lambda^d(A)} \int_{\mathbb{R}^d} \underbrace{\mathbb{1}_B(x) \mathbb{1}_A(x)}_{=\mathbb{1}_{A \cap B}(x)} dx = \frac{\lambda^d(A \cap B)}{\lambda^d(A)}.$$

Dieser Ausdruck ist analog zur Formel $P = \frac{\text{„Gewicht günstiger Ereignisse“}}{\text{„Gewicht aller möglichen Ereignisse“}}$ auf Laplace'schen Wahrscheinlichkeitsräumen.

b) Die d -dimensionale Standardnormalverteilung: Wähle

$$f_1(x) = \dots = f_d(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

¹¹ Ende der elften Vorlesung vom Fr, 16.05.14

Dann sind die assoziierten Wahrscheinlichkeitsmaße $P_1, \dots, P_d \in \mathcal{M}^1(\mathbb{R})$ die eindimensionalen Standardnormalverteilungen $N_{0,1}$. Das assoziierte Produktmaß

$$P := N_{0,1} \times \dots \times N_{0,1} \in \mathcal{M}^1(\mathbb{R}^d)$$

heißt d -dimensionale Standardnormalverteilung. Nach 4.15 d) ist die zu P gehörende Dichtefunktion

$$f(x_1, \dots, x_d) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{x_1^2 + \dots + x_d^2}{2}}.$$

4.17 Beispiel: (Buffon'sches Nadelproblem)

Auf dem Boden seien parallele Linien im Abstand 1 gezeichnet. Man werfe „zufällig“ eine Nadel der Länge $N \in (0, 1]$ auf den Boden. Wie groß ist nun die Wahrscheinlichkeit, dass die Nadel mindestens eine Linie trifft?

Entscheidend ist der Abstand des Nadelmittelpunktes von der nächsten Linie. Aus Symmetriegründen ist dieser Abstand auf $[0, \frac{1}{2}]$ gleichverteilt. Ferner sei der Winkel α der Nadel zu den Linien gleichverteilt auf $[0, \pi]$.

Lösung: (für dieses Modell) Wähle $A = [0, \frac{1}{2}] \times [0, \pi] \subseteq \mathbb{R}^2$ und P_A als Gleichverteilung auf A . Sei B das Ereignis, das dem Treffen der Nadel einer Linie entspricht. Dann ist

$$P_A(A \setminus B) = \frac{N}{2} \int_0^\pi \sin(t) dt = \frac{2N}{\pi}$$

und

$$P_A(B) = 1 - \frac{2N}{\pi}.$$

4.18 Beispiel: (Paradoxon von Bertrand)

Wähle im Einheitskreis $S := \{x \in \mathbb{R}^2 \mid \|x\|_2 \leq 1\}$ eine zufällige Kreissehne. Wie groß ist nun die Wahrscheinlichkeit, dass die Sehnenlänge mindestens der Länge eines gleichseitigen Dreiecks mit Eckpunkten auf dem Rand von S entspricht? Diese Länge ist offenbar $\sqrt{3}$, denn für die dritten komplexen Einheitswurzeln gilt $e^{\frac{2\pi i}{3}} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$.

Mathematisches Modell 1: Halte einen Randpunkt der Sehne fest, z.B. $1 \in \mathbb{C}$ und wähle den zweiten Randpunkt $z \in \mathbb{C}$ mit $|z| = 1$, sodass $\arg(z)$ auf $[0, 2\pi]$ gleichverteilt ist. Dann liefert die Geometrie, dass die gesuchte Wahrscheinlichkeit $\frac{1}{3}$ ist.

Mathematisches Modell 2: Wähle den Sehnenmittelpunkt gleichverteilt auf der Einheitskreisscheibe S . Dieser Sehnenmittelpunkt bestimmt dann die Sehne (orthogonal zum Durchmesser) eindeutig. Wähle $P = P_S \in \mathcal{M}^1(\mathbb{R}^2)$ als die Gleichverteilung auf S . Aus der Geometrie folgt dann für die „günstigen Ereignisse“ $B := \{z \in \mathbb{C} \mid |z| \leq \frac{1}{2}\}$ sofort

$$P_S(B) = \frac{\lambda^2(B)}{\lambda^2(S)} = \frac{\frac{\pi}{4}}{\pi} = \frac{1}{4}.$$

Es ergeben sich also verschiedene Ergebnisse bei beiden mathematischen Modellen. Dies ist aber nicht paradox, denn das ursprüngliche umgangssprachliche Problem ist zu unpräzise formuliert.

4.19 Definition:

Sei $A \subseteq \mathbb{R}^d$ eine Borelmenge, also $A \in \mathcal{B}(\mathbb{R}^d)$. Dann definiert man

$$\mathcal{M}^1(A) := \{P : A \rightarrow [0, 1] \mid P \text{ ist Wahrscheinlichkeitsmaß auf } (A, \mathcal{B}(A))\},$$

wobei $\mathcal{B}(A) = \{A \cap B \mid b \in \mathcal{B}(\mathbb{R}^d)\}$ ist.

Wie verhalten sich Verteilungen P_X von Zufallsvariablen X bei Transformationen?

4.20 Beispiel: (Verteilungen eindimensionaler transformierter Zufallsvariablen)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit Verteilung $P_X \in \mathcal{M}^1(\mathbb{R})$.

a) Sei F_X die Verteilungsfunktion zu P_X und $\varphi \in \mathcal{C}(\mathbb{R})$ streng monoton wachsend. Dann existiert auch $\varphi^{-1} \in \mathcal{C}(\mathbb{R})$ ebenfalls streng monoton wachsend. Die Transformationen φ, φ^{-1} sind insbesondere messbar (, da sie stetig sind), also ist

$$\varphi \circ X := \varphi(X) : \Omega \rightarrow \mathbb{R}$$

eine Zufallsvariable und $\varphi \circ X$ hat die Verteilungsfunktion

$$F_{\varphi \circ X}(x) = P(\varphi(X) \leq x) = P(X \leq \varphi^{-1}(x)) = F_X(\varphi^{-1}(x)), \quad x \in \mathbb{R}.$$

b) Sei $f_X : \mathbb{R} \rightarrow [0, \infty)$ die Dichtefunktion zu P_X und $\varphi \in \mathcal{C}^1(\mathbb{R})$ streng monoton wachsend. Dann existiert auch $\varphi^{-1} \in \mathcal{C}^1(\mathbb{R})$ ebenfalls streng monoton wachsend. Dann ergibt sich für die Dichtefunktion der transformierten Verteilung für $x \in \mathbb{R}$

$$f_{\varphi \circ X}(x) = \frac{d}{dx} F_{\varphi \circ X}(x) = \frac{d}{dx} F_X(\varphi^{-1}(x)) = (F_X)'(\varphi^{-1}(x)) \cdot (\varphi^{-1})'(x) = f_X(\varphi^{-1}(x)) \cdot (\varphi^{-1})'(x).$$

Man erhält also die nützliche Gleichung

$$f_{\varphi \circ X}(x) = f_X(\varphi^{-1}(x)) \cdot (\varphi^{-1})'(x). \quad (20)$$

c) Beispiel: Sei X eine standardnormalverteilte Zufallsvariable, d.h.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Sei ferner $\sigma > 0$ und $\varphi(x) := \sigma \cdot x$. Dann ist offenbar $\varphi \in \mathcal{C}^\infty(\mathbb{R})$ streng monoton wachsend und auch $\varphi^{-1} \in \mathcal{C}^\infty(\mathbb{R})$ ebenfalls streng monoton wachsend mit $\varphi^{-1}(x) = \frac{1}{\sigma}x$. Dann hat $P_{\varphi \circ X} = P_{\sigma X}$ die Dichtefunktion

$$f_{\varphi \circ X}(x) = f_{\sigma X}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\frac{x}{\sigma})^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}.$$

Die Verteilung von σX ist also $P_{\sigma X} = N_{0, \sigma^2}$.

d) Die Transformationsformeln aus Teil c) lassen sich auch durch individuelle Argumente auf nicht monotone Transformationen φ erweitern.

Beispiel: Sei die Zufallsvariable X $N_{0,1}$ -verteilt und $\varphi(x) := x^2$. Dann ist für $x < 0$

$$F_{\varphi \circ X}(x) = F_{X^2}(x) = P(X^2 \leq x) = 0$$

und für $x \geq 0$

$$F_{X^2}(x) = P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) = N_{0,1}([- \sqrt{x}, \sqrt{x}]) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1.$$

Also ist

$$F_{X^2}(x) = \begin{cases} 0 & , x < 0 \\ 2\Phi(\sqrt{x}) - 1 & , x \geq 0 \end{cases}$$

die Verteilungsfunktion von P_{X^2} . Die zugehörige Dichte ergibt sich durch Ableiten für $x < 0$ zu $f_{X^2}(x) = 0$ und

$$f_{X^2}(x) = \frac{d}{dx} F_{X^2}(x) = \frac{d}{dx} (2\Phi(\sqrt{x}) - 1) = 2\Phi'(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} = \frac{\Phi'(\sqrt{x})}{\sqrt{x}} = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}.$$

4.21 Transformationsformel für Dichtefunktionen bei Diffeomorphismen:

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$ eine Zufallsvariable mit Dichtefunktion $f_X : \mathbb{R}^d \rightarrow [0, \infty)$. Sei ferner $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ein \mathcal{C}^1 -Diffeomorphismus (bijektive stetig differenzierbare Abbildung, deren Umkehrabbildung wieder stetig differenzierbar ist). Dann ist auch $T \circ X : \Omega \rightarrow \mathbb{R}^d$ eine Zufallsvariable mit zugehöriger Dichtefunktion

$$f_{T \circ X}(x) = f_X(T^{-1}(x)) \cdot |\det \mathcal{J}T^{-1}(x)|, \quad x \in \Omega. \quad (21)$$

Dabei bezeichne $\mathcal{J}T^{-1}(x)$ die Jacobi-/Funktionalmatrix (die Matrix der ersten partiellen Ableitungen) von $T^{-1}(x)$.

Bemerkung: Formel (20) ist der eindimensionale Spezialfall von (21).¹²

¹² Ende der zwölften Vorlesung vom Mi, 21.05.14

5 Momente \mathbb{R} -wertiger Zufallsvariablen

Neben dem Wahrscheinlichkeitsmaß einer Zufallsvariablen, gibt es noch weitere wichtige Kenngrößen dieser, z.B. den Erwartungswert und die Varianz. Zur Definition dieser Begriffe benötigt man allerdings einen Integralbegriff bzgl. allgemeiner Maße.

5.1 Integration von Treppenfunktionen:

a) Sei (Ω, \mathcal{A}, P) ein Maßraum, d.h. $\Omega \neq \emptyset$, \mathcal{A} σ -Algebra und $P : \mathcal{A} \rightarrow [0, \infty]$ ein Maß. Für $A \in \mathcal{A}$ betrachte die Indikatorfunktion (oder auch charakteristische Funktion)

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

Offenbar ist $\mathbb{1}_A$ messbar $\Leftrightarrow A$ messbar. Setze nun

$$\int_{\Omega} \mathbb{1}_A dP = P(A).$$

b) Definition: Eine Funktion $f : \Omega \rightarrow [0, \infty)$ heißt nicht negative Treppenfunktion, falls es $n \in \mathbb{N}$, $a_1, \dots, a_n \geq 0$ und $A_1, \dots, A_n \in \mathcal{A}$ gibt, sodass

$$f = \sum_{k=1}^n a_k \mathbb{1}_{A_k} \quad (22)$$

gilt. Weiterhin ist

$$\mathcal{E}^+(\Omega, \mathcal{A}) := \{f : \Omega \rightarrow [0, \infty) \mid f \text{ ist eine nicht negative Treppenfunktion}\}$$

die Menge aller nichtnegativen Treppenfunktionen.

Definiere nun durch

$$\int_{\Omega} f dP = \sum_{k=1}^n a_k P(A_k)$$

das Integral für nichtnegative Treppenfunktionen.

Man verwendet ferner die Konventionen

$$0 \cdot x = 0 \quad \forall x \in [0, \infty] \quad \text{und} \quad \infty + x = \infty \quad \forall x \in [0, \infty]$$

5.2 Eigenschaften des Treppenintegrals:

Seien für $n \in \mathbb{N}$ $f, f_n, g, g_n \in \mathcal{E}^+(\Omega, \mathcal{A})$. Dann gilt:

a) Linearität: Für alle $a, b \geq 0$ gilt

$$\int_{\Omega} (af + bg) dP = a \int_{\Omega} f dP + b \int_{\Omega} g dP.$$

b) Monotonie: Für $f \leq g$ gilt

$$\int_{\Omega} f dP \leq \int_{\Omega} g dP.$$

c) Monotoner Limes: Sind f_n und g_n monoton wachsend. Falls $\lim_{n \rightarrow \infty} f_n = \sup_{n \in \mathbb{N}} f_n = \sup_{n \in \mathbb{N}} g_n = \lim_{n \rightarrow \infty} g_n$ gilt, so gilt auch

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n dP = \sup_{n \in \mathbb{N}} \int_{\Omega} f_n dP = \sup_{n \in \mathbb{N}} \int_{\Omega} g_n dP = \lim_{n \rightarrow \infty} \int_{\Omega} g_n dP.$$

d) Ist $f : \Omega \rightarrow [0, \infty)$ messbar, so existiert stets $(f_n) \subseteq \mathcal{E}^+(\Omega, \mathcal{A})$ mit $f_n \uparrow f$ für $n \rightarrow \infty$.

5.3 Integration messbarer Funktionen:

a) Sei $f : \Omega \rightarrow [0, \infty)$ eine positive messbare Funktion. Dann existiert nach 5.2 d) eine Folge $(f_n) \subseteq \mathcal{E}^+(\Omega, \mathcal{A})$ mit $f_n \uparrow f$ für $n \rightarrow \infty$. Definiere deshalb

$$\int_{\Omega} f dP := \lim_{n \rightarrow \infty} \int_{\Omega} f_n dP = \sup_{n \in \mathbb{N}} \int_{\Omega} f_n dP \in [0, \infty]. \quad (23)$$

Dieses Integral ist nach 5.2 c) wohldefiniert.

b) Definition: (Integrierbarkeit positiver und messbarer Funktionen) Eine positive Funktion $f : \Omega \rightarrow [0, \infty)$ heißt integrierbar, falls sie messbar ist und $\int_{\Omega} f dP < \infty$ ist.

c) Sei nun $f : \Omega \rightarrow \mathbb{R}$ messbar. Setze

$$f_+ := \max(f, 0) \geq 0 \text{ und } f_- := -\min(f, 0) \geq 0.$$

Insbesondere sind f_+, f_- messbar, da die Abbildungen \max, \min stetig sind. Ferner gilt

$$f = f_+ - f_- \text{ und } |f| = f_+ + f_-.$$

Nach Teil a) sind dann

$$\int_{\Omega} f_{\pm} dP, \int_{\Omega} |f| dP \in [0, \infty].$$

Also gilt

$$\int_{\Omega} |f| dP = \infty \Leftrightarrow \int_{\Omega} f_+ dP = \infty \text{ oder } \int_{\Omega} f_- dP = \infty.$$

d) Eine Funktion $f : \Omega \rightarrow \mathbb{R}$ heißt integrierbar (bzgl. P), falls f messbar und

$$\int_{\Omega} |f| dP < \infty \quad (\Leftrightarrow \int_{\Omega} f_+ dP < \infty \text{ und } \int_{\Omega} f_- dP < \infty).$$

In diesem Fall setze

$$\int_{\Omega} f dP := \int_{\Omega} f_+ dP - \int_{\Omega} f_- dP \in \mathbb{R}.$$

e) Eine Funktion $f : \Omega \rightarrow \mathbb{C}$ heißt integrierbar (bzgl. P), falls f messbar und $\operatorname{Re} f, \operatorname{Im} f$ integrierbar sind. In diesem Fall setze

$$\int_{\Omega} f dP := \int_{\Omega} \operatorname{Re} f dP + i \int_{\Omega} \operatorname{Im} f dP \in \mathbb{C}.$$

f) Für $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ definiere die Menge

$$\mathcal{L}_1(\Omega, \mathcal{A}, P) := \{f : \Omega \rightarrow \mathbb{K} \mid f \text{ integrierbar bzgl. } P\}.$$

$\mathcal{L}_1(\Omega, \mathcal{A}, P)$ ist ein Vektorraum, versehen mit der (Halb-)Norm (hängt vom verwendeten Maß P ab)

$$\|f\|_1 = \|f\|_{\mathcal{L}_1} := \int_{\Omega} |f| dP.$$

5.4 Fakt:

Das Integral

$$I : \mathcal{L}_1(\Omega, \mathcal{A}, P) \rightarrow \mathbb{K}, f \mapsto \int_{\Omega} f dP$$

ist ein lineares monotonen Funktional.

5.5 Beispiel: (Integration bzgl. diskreter Maße)

a) Sei $\Omega \neq \emptyset$ höchstens abzählbar und $\mathcal{A} = \mathcal{P}(\Omega)$. Wähle $(p_x)_{x \in \Omega} \subseteq [0, 1]$ mit $\sum_{x \in \Omega} p_x = 1$. Betrachte das assoziierte Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{A}) mit

$$P(A) := \sum_{x \in A} p_x \quad \forall A \in \mathcal{A}.$$

Für dieses Wahrscheinlichkeitsmaß kann man auch kurz

$$P(A) := \sum_{x \in \Omega} p_x \delta_x \quad \text{vgl. 4.7}$$

schreiben. In dieser Situation ist jede Funktion $f : \Omega \rightarrow \mathbb{K}$ messbar und es gilt:

$$\boxed{1} \quad f \in \mathcal{L}_1(\Omega, \mathcal{A}, P) \Leftrightarrow \sum_{x \in \Omega} p_x |f(x)| < \infty.$$

$$\boxed{2} \quad f \in \mathcal{L}_1(\Omega, \mathcal{A}, P) \Rightarrow \int_{\Omega} f dP = \sum_{x \in \Omega} p_x f(x) \in \mathbb{K}.$$

b) Sei $\Omega := \{0, 1, 2\}$, $\mathcal{A} := \mathcal{P}(\Omega)$ und $P := B_{2, \frac{1}{2}}$ Binomialverteilung mit $n = 2$ und $p = \frac{1}{2}$. Dann ist offenbar

$$P = \frac{1}{4}\delta_0 + \frac{1}{2}\delta_1 + \frac{1}{4}\delta_2.$$

Dann ist

$$\int_{\Omega} \sin(x) dP = \frac{1}{4} \sin(0) + \frac{1}{2} \sin(1) + \frac{1}{4} \sin(2).$$

5.6 Wiederholung: (Lebesgue-Maß auf \mathbb{R})

a) Auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ existiert genau ein Maß λ mit

$$\forall a < b \in \mathbb{R} : \lambda([a, b]) = b - a.$$

Dieses Maß heißt (eindimensionales) Lebesgue-Maß.

b) Für $a < b \in \mathbb{R}$ sei $f \in \mathcal{L}_1([a, b], \mathcal{B}([a, b]), \lambda)$ mit $\int_{\mathbb{R}} |f| d\lambda < \infty$. Dann ist f auch in $\mathcal{L}_1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ und es gilt

$$\int_{\mathbb{R}} f d\lambda = \int_{-\infty}^{\infty} f(x) dx.$$

5.7 Wiederholung: (Lebesgue-Maß auf \mathbb{R}^d)

a) Auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ existiert genau ein Maß λ^d mit

$$\forall a_1 < b_1, \dots, a_d < b_d \in \mathbb{R} : \lambda([a_1, b_1] \times \dots \times [a_d, b_d]) = \prod_{k=1}^d (b_k - a_k).$$

Dieses Maß heißt (d -dimensionales) Lebesgue-Maß.

5.8 Integration mit Dichtefunktionen bzgl. Wahrscheinlichkeitsmaßen auf \mathbb{R} bzw. \mathbb{R}^d :

a) Sei $f : \mathbb{R}^d \rightarrow [0, \infty)$ eine Dichtefunktion, d.h. f ist messbar und $\int_{\mathbb{R}^d} f d\lambda^d = 1$. Betrachte das zu f assoziierte Wahrscheinlichkeitsmaß $P \in \mathcal{M}^1(\mathbb{R}^d)$, d.h.

$$\forall A \in \mathcal{B}(\mathbb{R}^d) : P(A) = \int_A f(x) d\lambda^d(x) = \int_{\mathbb{R}^d} \mathbb{1}_A(x) f(x) d\lambda^d(x) \Leftrightarrow P = f\lambda^d.$$

Man beachte dabei, dass $P = f\lambda^d$ lediglich eine Schreibweise, aber kein „sinnvoller“ Ausdruck ist.

b) Satz: Sei $f : \mathbb{R}^d \rightarrow \mathbb{K}$ messbar. Dann gilt

$$\boxed{1} \quad g \in \mathcal{L}_1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P) \Leftrightarrow f \cdot g \in \mathcal{L}_1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P)$$

$$\boxed{2} \quad \forall g \in \mathcal{L}_1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), P) : \int_{\mathbb{R}^d} g dP = \int_{\mathbb{R}^d} g(x) \delta(x) d\lambda^d(x) = \int_{\mathbb{R}^d} g d(f\lambda^d) = \int_{\mathbb{R}^d} g f d\lambda^d.$$

c) Beispiel: Sei $f(x) := \begin{cases} e^{-x} & , x > 0 \\ 0 & , x \leq 0 \end{cases}$ die Dichte einer Exponentialverteilung P mit Parameter 1. Dann ist diese Verteilung gegeben durch $P = f \cdot \lambda$. Für $g(x) = x$ ergibt sich¹³

$$\int_{\mathbb{R}} x dP(x) = \int_{\mathbb{R}} x f(x) dx = \int_0^{\infty} x \cdot e^{-x} dx = 1$$

5.9 Satz: (Konvergenzsätze der Integrationstheorie)

Seien (Ω, \mathcal{A}, P) ein Maßraum und $(f_n) \subseteq \mathcal{L}_1(\Omega, \mathcal{A}, P)$ mit $f_n \rightarrow f$ punktweise für $n \rightarrow \infty$. Dann gelten die beiden wichtigen Sätze:

a) Satz über monotone Konvergenz: Falls f_n, f reellwertig sind und $f_n \uparrow f \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$, so gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n dP = \int_{\Omega} f dP.$$

b) Satz über majorisierte Konvergenz (Lebesgue): Sei $g \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ mit der Eigenschaft

$$\forall n \in \mathbb{N}, \forall \omega \in \Omega : |f_n(\omega)| \leq |g(\omega)|$$

eine so genannte (Lebesgue-) Majorante zu f_n . Falls es eine Funktion f gibt, sodass $f_n \rightarrow f$ punktweise für $n \rightarrow \infty$, dann ist $f \in \mathcal{L}(\Omega, \mathcal{A}, P)$ und es gilt

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n dP = \int_{\Omega} f dP.$$

5.10 Definition: (Erwartungswert)

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable.

a) Falls $X \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$, so heißt

$$E(X) := \int_{\Omega} X dP \in \mathbb{R}$$

Erwartungswert von X .

b) Sei $n \in \mathbb{N}$ und $X^n \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$, dann heißt

$$\mu_n := \mu_n(X) := E(X^n) = \int_{\Omega} X^n dP \in \mathbb{R}$$

das *n-te Moment* von X .

c) Man bezeichne mit

$$\mathcal{L}_2(\Omega, \mathcal{A}, P) := \{X : \Omega \rightarrow \mathbb{R} \mid X \text{ messbar, } \int_{\Omega} |X|^2 dP < \infty\}$$

den Vektorraum der quadratintegrierbaren Zufallsvariablen.

¹³ Ende der dreizehnten Vorlesung vom Fr, 23.05.14

5.11 Satz: (abstrakter Transformationssatz)

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\tilde{\Omega}, \tilde{\mathcal{A}})$ ein Messraum, $X : \Omega \rightarrow \tilde{\Omega}$ eine Zufallsvariable und $h : \tilde{\Omega} \rightarrow \mathbb{R}$ eine messbare Funktion. Dann gilt stets

$$h \circ X \in \mathcal{L}_1(\Omega, \mathcal{A}, P) \Leftrightarrow h \in \mathcal{L}_1(\tilde{\Omega}, \tilde{\mathcal{A}}, P_X) \quad (24)$$

und

$$\int_{\Omega} h \circ X \, dP = \int_{\tilde{\Omega}} h \, dP_X. \quad (25)$$

Beweis.

a) Sei zunächst $h := \mathbb{1}_A$ für $A \in \mathcal{A}$. Dann gilt natürlich

$$\int_{\tilde{\Omega}} h \, dP_X = P_X(A) = P(X^{-1}(A)) = \int_{\Omega} \mathbb{1}_{X^{-1}(A)} \, dP = \int_{\Omega} \mathbb{1}_A(X(\omega)) \, dP(\omega) = \int_{\Omega} h \circ X \, dP.$$

b) Aus Teil a) und der Linearität des Integrals folgt die Behauptung für alle $h \in \mathcal{E}^+(\tilde{\Omega}, \tilde{\mathcal{A}})$.

c) Sei $h \geq 0$ eine messbare Funktion. Nach 5.2 d) existiert dann eine Folge $(h_n) \subseteq \mathcal{E}^+(\tilde{\Omega}, \tilde{\mathcal{A}})$ mit $h_n \uparrow h$ punktweise für $n \rightarrow \infty$. Dann folgt auch $(h_n \circ X) \uparrow h \circ X$ für $n \rightarrow \infty$. Also ergibt sich mit dem Satz über monotone Konvergenz für alle $n \in \mathbb{N}$

$$\int_{\tilde{\Omega}} h \circ X \, dP \stackrel{n \rightarrow \infty}{\leftarrow} \int_{\tilde{\Omega}} h_n \circ X \, dP \stackrel{b)}{=} \int_{\tilde{\Omega}} h_n \, dP_X \stackrel{n \rightarrow \infty}{\rightarrow} \int_{\tilde{\Omega}} h \, dP_X.$$

d) Aus Teil c) und der Linearität des Integrals folgt die Behauptung für alle messbaren Funktionen h .

(Analog beweist man auch die Äquivalenz in Gleichung (24).) □

5.12 Berechnung von Erwartungswerten und Momenten:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Ferner sei $h : \mathbb{R} \rightarrow \mathbb{R}$ mit $h(x) := x^n$ für $n \in \mathbb{N}$. Falls $X^n \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ liegt, gilt

$$\mu_n(X) = \int_{\Omega} X^n \, dP = \int_{\Omega} h \circ X \, dP \stackrel{(25)}{=} \int_{\mathbb{R}} x^n \, dP_X(x). \quad (26)$$

Insbesondere gilt: Ist $P_X \in \mathcal{M}^1(\mathbb{R})$ bekannt, so kann man damit alle existierenden Momente $\mu_n(X)$ aus P_X berechnen. Umgekehrt kann man allerdings i.A. nicht die Verteilung P_X aus den Momenten berechnen!

5.13 Diskrete Beispiele:

a) Sei $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable, d.h. der Wertebereich

$$X(\Omega) := \{X(\omega) | \omega \in \Omega\} = \{x_1, x_2, \dots\} \subseteq \mathbb{R}$$

von X ist höchstens abzählbar. Dann ist P_X ein Wahrscheinlichkeitsmaß auf $X(\Omega)$ mit

$$P_X = \sum_{n \geq 1} p_{x_n} \delta_{x_n} \text{ mit } p_{x_n} = P(X = x_n) \geq 0.$$

Im Falle der Existenz folgt also nach 5.12

$$E(X) = \int_{\Omega} X \, dP = \int_{\mathbb{R}} x \, dP_X(x) = \sum_{n \geq 1} p_{x_n} x_n$$

und allgemeiner für alle $n \in \mathbb{N}$

$$\mu_n(X) = \int_{\Omega} X^n \, dP = \int_{\mathbb{R}} x^n \, dP_X(x) = \sum_{k \geq 1} p_{x_k} x_k^n.$$

b) Sei $X : \Omega \rightarrow \mathbb{R}$ eine geometrisch verteilte Zufallsvariable mit $p \in (0, 1)$, also gilt

$$P(X = k) = p(1-p)^k, \quad \forall k \in \mathbb{N}_0.$$

Damit ergibt sich für den Erwartungswert einer geometrisch verteilten Zufallsvariablen

$$E(X) = \sum_{k=0}^{\infty} p(1-p)^k \cdot k = p(1-p) \sum_{k=0}^{\infty} k \cdot (1-p)^{k-1} = \frac{p(1-p)}{(1-(1-p))^2} = \frac{1-p}{p}.$$

c) Sei $X : \Omega \rightarrow \mathbb{R}$ eine $B_{n,p}$ -verteilte Zufallsvariable mit $p \in (0, 1)$ und $n \in \mathbb{N}_0$, also gilt

$$B_{n,p}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \forall k \in \{0, \dots, n\}.$$

Damit ergibt sich für den Erwartungswert einer binomial verteilten Zufallsvariablen mit Hilfe des binomischen Satzes

$$\begin{aligned} E(X) &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \cdot k = \sum_{k=1}^n \frac{n(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{k-1} (1-p)^{(n-1)-(k-1)} \cdot p \\ &\stackrel{k=k-1}{=} n \cdot p \cdot \sum_{l=0}^{n-1} \binom{n-1}{l} p^l \cdot (1-p)^{(n-1)-l} = n \cdot p \cdot (p+1-p)^{n-1} = n \cdot p. \end{aligned}$$

5.14 Kontinuierliche Beispiele:

a) Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable. Die Verteilung $P_X \in \mathcal{M}^1(\mathbb{R})$ habe die Dichtefunktion $f \geq 0$. Im Falle der Existenz folgt also nach 5.12

$$E(X) = \int_{\Omega} X \, dP = \int_{\mathbb{R}} x \, dP_X(x) = \int_{\mathbb{R}} x \cdot f(x) \, dx.$$

und allgemeiner für alle $n \in \mathbb{N}$

$$\mu_n(X) = \int_{\Omega} X^n \, dP = \int_{\mathbb{R}} x^n \, dP_X(x) = \int_{\mathbb{R}} x^n \cdot f(x) \, dx.$$

b) Sei $X : \mathbb{R} \rightarrow \mathbb{R}$ eine exponentialverteilte Zufallsvariable mit Parameter $\lambda > 0$, also gilt

$$f_{\lambda}(x) = \begin{cases} 0 & , x \leq 0 \\ \lambda \cdot e^{-\lambda x} & , x > 0 \end{cases}, \quad \forall x \in \mathbb{R}$$

Damit ergibt sich für den Erwartungswert einer exponentialverteilten Zufallsvariablen

$$E(X) = \int_{\mathbb{R}} x \cdot f(x) \, dx = \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x} \, dx = \frac{1}{\lambda}.$$

c) Sei $X : \mathbb{R} \rightarrow \mathbb{R}$ eine N_{μ, σ^2} -verteilte Zufallsvariable mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, also gilt

$$f_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \forall x \in \mathbb{R}$$

Damit ergibt sich für den Erwartungswert einer normalverteilten Zufallsvariablen

$$\begin{aligned} E(X) &= \int_{\mathbb{R}} x \cdot f(x) \, dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx \stackrel{y:=x-\mu}{=} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (y + \mu) \cdot e^{-\frac{y^2}{2\sigma^2}} \, dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{\int_{-\infty}^{\infty} \mu \cdot e^{-\frac{y^2}{2\sigma^2}} \, dy}_{=2\pi\sigma^2\mu} + \frac{1}{\sqrt{2\pi\sigma^2}} \underbrace{\int_{-\infty}^{\infty} y \cdot e^{-\frac{y^2}{2\sigma^2}} \, dy}_{=0} = \mu. \end{aligned}$$

5.15 Rechenregeln für Erwartungswerte:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X, Y : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen.

a) Linearität: Für alle $a, b \in \mathbb{R}$ und $X, Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ folgt aus der Linearität des Integrals sofort

$$E(aX + bY) = a \cdot E(X) + b \cdot E(Y)$$

b) Für $X, Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ mit X, Y unabhängig gilt

$$E(X \cdot Y) = E(X) \cdot E(Y).$$

Für nicht unabhängige Zufallsvariablen ist dies i.A. nicht richtig.

c) Sei $n \in \mathbb{N}$, sodass das n -te Moment $\mu_n(X) = \int_{\Omega} X^n dP$ existiert. Dann existiert auch $\mu_m(X)$ für alle $m \leq n$. Speziell ist also

$$\mathcal{L}_1(\Omega, \mathcal{A}, P) \subseteq \mathcal{L}_2(\Omega, \mathcal{A}, P)$$

für Wahrscheinlichkeitsräume (Ω, \mathcal{A}, P) .

d) Es gilt die folgende Version der Cauchy-Schwarz'schen Ungleichung (CSU):

Seien $X, Y \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$, dann ist $X \cdot Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ und man hat

$$E(X \cdot Y)^2 \leq E(X^2) \cdot E(Y^2).$$

Beweis. a) Folgt sofort aus der Linearität des Integrals.

b) (für diskrete Zufallsvariablen X, Y)

Die Zufallsvariable X nehme Werte $\{x_1, x_2, \dots\} \subseteq \mathbb{R}$ an, Y nehme Werte $\{y_1, y_2, \dots\} \subseteq \mathbb{R}$ an. Dann nimmt die Zufallsvariable $X \cdot Y$ Werte in der abzählbaren Menge

$$\{x_i \cdot y_j \mid i, j \geq 1\} \subseteq \mathbb{R}$$

an. Dann ergibt sich für den Erwartungswert wegen der Unabhängigkeit von X und Y sofort

$$\begin{aligned} E(X \cdot Y) &= \sum_{i, j \geq 1} x_i y_j P(X = x_i, Y = y_j) = \sum_{i, j \geq 1} x_i P(X = x_i) y_j P(Y = y_j) \\ &= \sum_{i \geq 1} x_i P(X = x_i) \cdot \sum_{j \geq 1} y_j P(Y = y_j) = E(X) \cdot E(Y) \end{aligned}$$

(Analog zeigt man, dass alle vorkommenden Folgen tatsächlich absolut summierbar sind.)¹⁴

c) Für $m \leq n$ gilt

$$|\mu_m(X)| \leq \int_{\mathbb{R}} |x|^m dP_X(x) = \underbrace{\int_{-1}^1 |x|^m dP_X(x)}_{\leq 1} + \int_{\mathbb{R} \setminus [-1, 1]} \underbrace{|x|^m dP_X(x)}_{\leq |x|^n} \leq 1 + \mu_n(X) < \infty.$$

d) Unterscheide zwei Fälle:

Fall 1: Sei $0 = E(Y^2) = \int_{\Omega} Y^2 dP \Rightarrow P(Y = 0) = 1 \Rightarrow E(X \cdot Y) = 0$.

Fall 2: Sei $E(Y^2) > 0$, dann folgt

$$\begin{aligned} 0 &\leq E((E(Y^2) \cdot X - E(X \cdot Y) \cdot Y)^2) = E(Y^2)^2 \cdot E(X^2) + E(X \cdot Y)^2 E(Y^2) - 2E(Y^2)E(XY)^2 \\ &= E(Y^2) \cdot E(X^2) - E(XY)^2. \end{aligned}$$

□

¹⁴ Ende der vierzehnten Vorlesung vom Mi, 28.05.14

5.16 Beispiel:

Seien X_1, \dots, X_n unabhängige $B_{1,p}$ -verteilte Zufallsvariablen. Dann ist $X := X_1 + \dots + X_n$ eine $B_{n,p}$ -verteilte Zufallsvariable. Für den Erwartungswert folgt wegen der Linearität des Erwartungswertes sofort

$$E(X) = E(X_1) + \dots + E(X_n) = n \cdot E(X_1) = n(0 \cdot (1-p) + 1 \cdot p) = n \cdot p \quad (\text{vgl. 5.13})$$

5.17 Definition: (Varianz, Kovarianz, Streuung)

Seien $X, Y \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$. Dann ist wegen 5.15 d) auch $X \cdot Y \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ und man hat folgende Begriffe:

a) Die Zahl

$$Kov(X, Y) := E((X - E(X)) \cdot (Y - E(Y))) \in \mathbb{R}$$

heißt *Kovarianz* von X und Y .

b) Die Zahl

$$Var(X) := Kov(X, X) = E((X - E(X))^2) \in \mathbb{R}$$

heißt *Varianz* von X . Die Varianz beschreibt die mittlere quadratische Abweichung von X zum Erwartungswert $E(X)$.

c) Die Zahl

$$\sigma(X) := \sqrt{Var(X)} \geq 0$$

heißt *Standardabweichung* oder *Streuung* von X . $\sigma(X)$ ist auch ein Maß für die Abweichung von X zum Erwartungswert $E(X)$.

5.18 Rechenregeln für Varianz und Kovarianz:

Seien $X, Y, X_1, \dots, X_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$. Dann gilt:

a) Wegen der Linearität des Erwartungswertes erhält man

$$Kov(X, Y) = E(XY) - E(X)E(Y) \quad (27)$$

b) Aus Teil a) folgt dann sofort für die Varianz

$$Var(X) = E(X^2) - E(X)^2 \quad (28)$$

c) Wegen Teil b) erhält man für alle $a, b \in \mathbb{R}$

$$Var(aX + b) = a^2 Var(X) \quad \text{und} \quad \sigma(aX + b) = |a| \sigma(X),$$

wie man leicht nachrechnen kann.

d) Es gilt stets $Kov(X, Y) = Kov(Y, X)$.

e) Für die Summe mehrerer Zufallsvariablen gilt

$$Var\left(\sum_{k=1}^n X_k\right) = E\left(\left(\sum_{k=1}^n (X_k - E(X_k))\right)^2\right) = \sum_{k,l=1}^n E\left((X_k - E(X_k)) \cdot (X_l - E(X_l))\right) = \sum_{k,l=1}^n Kov(X_k, X_l).$$

f) Für X, Y unabhängig gilt

$$\text{Kov}(X, Y) = E(XY) - E(X)E(Y) \stackrel{5.15 \text{ b)}}{=} E(X)E(Y) - E(X)E(Y) = 0.$$

Aus den Teilen e) und f) folgt ferner die

5.19 Formel von Bienaymé

Seien $X_1, \dots, X_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$ unabhängig. Dann gilt:

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k) \quad (29)$$

5.20 Beispiele:

a) Sei $X : \Omega \rightarrow \mathbb{R}$ eine $B_{n,p}$ -verteilte Zufallsvariable. Was ist dann $\text{Var}(X)$?

Lösung 1: Mit der Definition der Varianz erhält man die schwierig zu berechnende Summe

$$\text{Var}(X) = E(X^2) - E(X)^2 = \sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} - (n \cdot p)^2.$$

Lösung 2: Seien X_1, \dots, X_n $B_{1,p}$ -verteilte Zufallsvariablen, dann ist

$$\tilde{X} := X_1 + \dots + X_n$$

eine $B_{n,p}$ -verteilte Zufallsvariable. Da X, \tilde{X} beides $B_{n,p}$ -verteilte Zufallsvariablen sind, gilt

$$E(X) = E(\tilde{X}) \text{ und } \text{Var}(X) = \text{Var}(\tilde{X}).$$

Mit der Formel von Bienaymé 5.19 folgt dann

$$\text{Var}(X) = \text{Var}(\tilde{X}) = \sum_{k=1}^n \text{Var}(X_k) = n \cdot \text{Var}(X_1) = n \cdot (E(X^2) - E(X)^2) = np(1-p).$$

Damit folgt dann insbesondere

$$\sum_{k=0}^n k^2 \binom{n}{k} p^k (1-p)^{n-k} = \text{Var}(X) + (np)^2 = np(1-p) - (np)^2 = np(1-p-np).$$

b) Sei nun X eine N_{μ, σ^2} -verteilte Zufallsvariable. Dann gilt $E(X) = \mu$ (vgl. 5.14) und es soll gezeigt werden, dass $\text{Var}(X) = \sigma^2$ ist. Betrachte dazu die transformierte Zufallsvariable $X - \mu$, diese ist dann N_{0, σ^2} -verteilt und wegen 5.18 b) gilt

$$\begin{aligned} \text{Var}(X) &= \text{Var}(X - \mu) = E((X - \mu)^2) - \underbrace{E(X - \mu)^2}_{=0} = \int_{\mathbb{R}} x^2 \cdot f_{0, \sigma^2}(x) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\sigma^2}} dx \stackrel{y:=\frac{x}{\sigma}}{=} \frac{\sigma^2}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy}_{=\sqrt{2\pi}} = \sigma^2 \end{aligned}$$

5.21 Definition: (Korrelation)

Seien $X, Y \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$ mit $\text{Var}(X) > 0$ und $\text{Var}(Y) > 0$.

a) Die Zahl

$$\rho_{X,Y} := \frac{\text{Kov}(X, Y)}{\sigma(X)\sigma(Y)}$$

heißt Korrelationskoeffizient von X und Y .

b) Falls $\rho_{X,Y} = 0$, so heißen X und Y nicht korreliert.

Falls $\rho_{X,Y} > 0$, so heißen X und Y positiv korreliert.

Falls $\rho_{X,Y} < 0$, so heißen X und Y negativ korreliert.

5.22 Fakten:

Seien $X, Y \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$ mit $\text{Var}(X) > 0$ und $\text{Var}(Y) > 0$. Dann gilt:

a) Aus der CSU (vgl. 5.15 d)) folgt $\rho_{X,Y} \in [-1, 1]$.

b) Es gilt stets $\rho_{X,X} = 1$ und $\rho_{X,-X} = -1$.

c) Sind X und Y unabhängig, so gilt stets $\rho_{X,Y} = 0$. Die Rückrichtung ist falsch.

5.23 Definition: (Erzeugendenfunktion)

Sei $X : \Omega \rightarrow \mathbb{N}_0$ eine Zufallsvariable mit $p_n := P(X = n)$ für $n \in \mathbb{N}_0$. Die sogenannte *Erzeugendenfunktion* $g_X : [-1, 1] \rightarrow \mathbb{R}$ ist definiert durch

$$g_X(t) := \sum_{n=0}^{\infty} p_n t^n = E(t^X).$$

Dies ist eine Potenzreihe in t mit Konvergenzradius $R \geq 1$.

Fakten: (Erzeugendenfunktion)

Sei $X : \Omega \rightarrow \mathbb{N}_0$ eine Zufallsvariable mit $p_n := P(X = n)$ für $n \in \mathbb{N}_0$ und $R > 1$. Dann gilt:

a) $g'_X(1) = E(X)$

b) $g''_X(1) = E((X-1)X) = E(X^2) - E(X)$

c) $\text{Var}(X) = g''_X(1) + g'_X(1)(1 - g'_X(1))$.

Beweis.

Gliedweises differenzieren für $t \in (-R, R)$ liefert

a) $g'_X(t) = \sum_{n=1}^{\infty} n p_n t^{n-1} \Rightarrow g'_X(1) = E(X)$.

b) $g''_X(t) = \sum_{n=2}^{\infty} n(n-1) p_n t^{n-2} \Rightarrow g'_X(1) = E((X-1)X) = E(X^2) - E(X)$.

c) $\text{Var}(X) = E(X^2) - E(X)^2 \stackrel{\text{b)}}{=} g''_X(1) + E(X) - E(X)^2 \stackrel{\text{a)}}{=} g''_X(1) + g'_X(1)(1 - g'_X(1))$. \square

5.24 Regeln für Erzeugendenfunktionen:

a) Seien $X, Y : \Omega \rightarrow \mathbb{N}_0$ unabhängige Zufallsvariablen. Dann gilt für $t \in [-1, 1]$

$$g_{X+Y}(t) = E(t^{X+Y}) = E(t^X \cdot t^Y) = E(t^X) \cdot E(t^Y) = g_X(t) \cdot g_Y(t).$$

b) Gilt für Zufallsvariablen $X, Y : \Omega \rightarrow \mathbb{N}_0$ stets $g_X(t) = g_Y(t)$ für alle $t \in [-1, 1]$, so folgt $P_X = P_Y$, denn sei $p_n^{(X)} := P(X = n)$ und $p_n^{(Y)} := P(Y = n)$ für $n \in \mathbb{N}_0$, dann folgt für alle $t \in [-1, 1]$

$$\sum_{n=0}^{\infty} p_n^{(X)} t^n = g_X(t) = g_Y(t) = \sum_{n=0}^{\infty} p_n^{(Y)} t^n.$$

Nach dem Identitätssatz für Potenzreihen folgt dann $p_n^{(X)} = p_n^{(Y)}$ für alle $n \in \mathbb{N}_0$ und somit $P_X = P_Y$.

5.25 Beispiel: (Poissonverteilung)

a) Sei X eine poissonverteilte Zufallsvariable mit Parameter $\lambda > 0$. Dann ist

$$g_X(t) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda t)^n}{n!} = e^{-\lambda} e^{\lambda t} = e^{\lambda(t-1)}.$$

Daraus folgt für Erwartungswert und Varianz einer poissonverteilten Zufallsvariablen X

$$\begin{aligned} E(X) &= g'_X(1) = \lambda e^{\lambda(t-1)}|_{t=1} = \lambda \\ \text{Var}(X) &= g''_X(1) + \lambda(1 - \lambda) = \lambda^2 e^{\lambda(t-1)}|_{t=1} + \lambda(1 - \lambda) = \lambda. \end{aligned}$$

b) Seien $X, Y : \Omega \rightarrow \mathbb{N}_0$ unabhängige Zufallsvariablen mit $p_X = \Pi_\lambda$ und $p_Y = \Pi_\mu$ für $\lambda, \mu > 0$. Dann ergibt sich

$$g_{X+Y}(t) = g_X(t)g_Y(t) = e^{\lambda(t-1)}e^{\mu(t-1)} = e^{-(\lambda+\mu)}e^{(\lambda+\mu)t}.$$

Also ist $g_{X+Y}(t)$ Erzeugendenfunktion einer $\Pi_{\lambda+\mu}$ -verteilten Zufallsvariablen. Mit 5.24

b) folgt dann, dass die Zufallsvariable $X + Y$ auch $\Pi_{\lambda+\mu}$ -verteilt ist.¹⁵

5.26 Satz: (Markov-Ungleichung)

Sei $I \subseteq \mathbb{R}$ ein Intervall, $f : I \rightarrow (0, \infty)$ eine monoton wachsende Funktion und $Y : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit $|Y(\omega)| \in I$ für alle $\omega \in \Omega$. Dann gilt für $\varepsilon > 0$ stets

$$P(|Y| \geq \varepsilon) \leq \frac{E(f(|Y|))}{f(\varepsilon)} \quad (30)$$

¹⁵ Ende der fünfzehnten Vorlesung vom Fr, 30.05.14

Beweis:

Da f monoton also insbesondere messbar ist, ist $f \circ Y$ wieder eine Zufallsvariable. Betrachte nun die Zufallsvariable

$$Z := \mathbf{1}_{\{|Y| \geq \varepsilon\}} f(\varepsilon)$$

mit der Eigenschaft $Z \leq f(|Y|)$, da f monoton wachsend ist. Dann folgt sofort

$$f(\varepsilon) \cdot E(\mathbf{1}_{\{|Y| \geq \varepsilon\}}) = f(\varepsilon) \cdot P(|Y| \geq \varepsilon) = E(Z) \leq E(f(|Y|))$$

□

5.27 Folgerungen:

Sei $Y : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, dann gilt

a) *Momentenungleichung:* Falls das k -te Moment von X existiert, dann gilt für alle $\varepsilon > 0$

$$P(|Y| \geq \varepsilon) \leq \frac{E(|X|^k)}{\varepsilon^k}. \quad (31)$$

Diese Gleichung ergibt sich sofort aus der Markov-Ungleichung (30) mit $f(x) = x^k$.

b) *Tschebyscheff-Ungleichung:* Für alle $X \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$ gilt stets

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}. \quad (32)$$

Diese Gleichung ergibt sich sofort aus Teil a) angewendet auf $X - E(X)$ mit $k = 2$.

c) *Exponentialungleichung:* Sei $\delta > 0$ mit $E(e^{\delta X}) < \infty$. Dann gilt für alle $\varepsilon > 0$

$$P(|X| \geq \varepsilon) \leq \frac{E(e^{\delta X})}{e^{\delta \varepsilon}}. \quad (33)$$

Diese Gleichung ergibt sich analog zur Markov-Ungleichung (30) mit $f(x) = e^{\delta x}$.

5.28 Beispiel: (Tailabschätzung für $N_{0,1}$)

In welcher Größenordnung liegt

$$1 - \Phi(\varepsilon) = \frac{1}{\sqrt{2\pi}} \int_{\varepsilon}^{\infty} e^{-\frac{x^2}{2}} dx$$

für $\varepsilon > 0$ groß?

Betrachte dazu eine $N_{0,1}$ -verteilte reellwertige Zufallsvariable X .

Ansatz 1: Die Tschebyscheff-Ungleichung liefert sofort

$$1 - \Phi(\varepsilon) = P(X \geq \varepsilon) = \frac{1}{2} P(|X| \geq \varepsilon) \leq \frac{1}{2\varepsilon^2}.$$

Ansatz 2: Die Exponentialungleichung liefert für $\delta > 0$

$$1 - \Phi(\varepsilon) = P(X \geq \varepsilon) \leq \frac{E(e^{\delta X})}{e^{\delta \varepsilon}}.$$

Für den Erwartungswert ergibt sich mit Hilfe des abstrakten Transformationsatzes 5.11 für $h(y) := e^{\delta y}$

$$E(e^{\delta X}) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\delta x} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\frac{\delta^2}{2}} e^{-\frac{(x-\delta)^2}{2}} dy = e^{\frac{\delta^2}{2}}.$$

Damit folgt weiterhin

$$1 - \Phi(\varepsilon) \leq e^{\frac{\delta^2}{2} - \varepsilon \delta}, \quad \forall \varepsilon > 0.$$

Der Ausdruck $e^{\frac{\delta^2}{2} - \varepsilon \delta}$ wird für $\delta = \varepsilon$ minimal, also folgt

$$1 - \Phi(\varepsilon) \leq e^{-\frac{\varepsilon^2}{2}}.$$

Diese Abschätzung ist viel besser als die, die durch die Tschebyscheff-Ungleichung zustande kommt.

5.29 Satz: (Schwaches Gesetz der großen Zahlen)

Seien $X_1, \dots, X_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$ unkorrelierte Zufallsvariablen, die alle den gleichen Erwartungswert $m := E(X_k)$, $k = 1, \dots, n$. Ferner sei $M > 0$ mit $Var(X_k) \leq M$ für alle $k = 1, \dots, n$. Das so genannte *empirische Mittel*

$$\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$$

ist eine „Schätzung“ für den Erwartungswert m und für alle $\varepsilon > 0$ gilt

$$P(|\bar{X}_n - m| \geq \varepsilon) \leq \frac{M}{\varepsilon^2 n} \rightarrow 0 \text{ für } n \rightarrow \infty. \quad (34)$$

Beweis:

Für den Erwartungswert des empirischen Mittels folgt sofort

$$E(\bar{X}_n) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \frac{m \cdot n}{n} = m.$$

Für die Varianz folgt dann mit der Formel von Bienayme 5.19

$$Var(\bar{X}_n) = \frac{1}{n^2} Var\left(\sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n Var(X_k) \leq \frac{n \cdot M}{n^2} = \frac{M}{n}.$$

Mit der Tschebyscheff-Ungleichung (32) folgt die Behauptung. \square

Bemerkungen: a) Für $P(|\bar{X}_n - m| \geq \varepsilon) \rightarrow 0$ für $n \rightarrow \infty$ sagt man auch, die Folge $(\bar{X}_n)_{n \in \mathbb{N}}$ der empirischen Mittel konvergiert *stochastisch*. Diese Konvergenz ist relativ schwach.

b) Andere Ungleichungen aus 5.27 liefern evtl. bessere Abschätzungen in (34).

5.30 Beispiel: (Wahlvorhersage)

Wie groß ist der Mindestumfang $n \in \mathbb{N}$ einer Stichprobe für eine Wahlvorhersage des Stimmanteils $p \in (0, 1)$ einer Partei vor einer Wahl, wenn man die Qualitätsforderung stellt, dass die Wahrscheinlichkeit dafür, dass ein absoluter Fehler größer als 0,01 auftritt, höchstens 0,05 beträgt?

Mathematisches Modell: Wähle n unabhängige Testpersonen, die die Bevölkerung repräsentieren. Seien X_1, \dots, X_n stochastisch unabhängige $\{0, 1\}$ -wertige Zufallsvariablen, wobei 1 der Wahl der Partei und 0 der Wahl einer anderen Partei entspricht. Man nimmt nun an, dass alle X_k stets $B_{1,p}$ -verteilt sind, wobei

$$B_{1,p} = p\delta_1 + (1-p)\delta_2.$$

Dann ist

$$S_n := \sum_{k=1}^n X_k$$

$B_{n,p}$ -verteilt und das empirische Mittel

$$\bar{X}_n = \frac{1}{n} S_n$$

ist eine gute Schätzung für p .

Das schwache Gesetz der großen Zahlen 5.29 liefert nun

$$P(|\bar{X}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{\varepsilon^2 n} \leq \frac{1}{4\varepsilon^2 n} = \frac{10^4}{4n} \stackrel{!}{\leq} 0,05 \Leftrightarrow n \geq 50\,000.$$

Andere Ungleichungen aus 5.27 liefern, dass die Qualitätsanforderung bereits für $n = 10\,000$ erfüllt ist.

6 Verteilungen von Summen unabhängiger Zufallsvariablen

6.1 Wiederholung: (Gemeinsame Verteilung)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen. Dann hat man:

a) Für alle $i = 1, \dots, n$ ist die Verteilung $P_{X_i} \in \mathcal{M}^1(\mathbb{R})$ von X_i gegeben durch

$$P_{X_i}(A) = P(X_i \in A) = P(\{\omega \in \Omega \mid X_i(\omega) \in A\}) = P(X_i^{-1}(A))$$

für alle $A \in \mathcal{B}(\mathbb{R})$.

b) Die gemeinsame Verteilung von X_1, \dots, X_n ist das eindeutige Wahrscheinlichkeitsmaß $P_{(X_1, \dots, X_n)} \in \mathcal{M}^1(\mathbb{R})$ mit der Eigenschaft

$$\forall A \in \mathcal{B}(\mathbb{R}) : P_{(X_1, \dots, X_n)}(A) = P((X_1, \dots, X_n) \in A) = P((X_1, \dots, X_n)^{-1}(A)).$$

c) Nach Satz 3.13 gilt:

$$X_1, \dots, X_n \text{ unabhängig} \Leftrightarrow P_{(X_1, \dots, X_n)} = P_{X_1} \times \dots \times P_{X_n}.$$

Beispiel: (Gemeinsame Verteilung)

Seien $n = 2$ und X_1, X_2 $\{1, 2\}$ -wertige Zufallsvariablen. Dann ist $P_{(X_1, X_2)}$ ein Wahrscheinlichkeitsmaß auf $\{1, 2\}^2$. Seien konkret

a)	<table style="border-collapse: collapse; text-align: center;"> <tr> <td style="border-right: 1px solid black; padding: 5px;"></td> <td style="border-right: 1px solid black; padding: 5px;">1</td> <td style="border-right: 1px solid black; padding: 5px;">2</td> <td style="border-right: 1px solid black; padding: 5px;">Σ</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">1</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{8}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{3}{8}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">2</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{3}{8}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{8}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">Σ</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> <td style="padding: 5px;"></td> </tr> </table>		1	2	Σ	1	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$	2	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{2}$	Σ	$\frac{1}{2}$	$\frac{1}{2}$	
	1	2	Σ														
1	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$														
2	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{2}$														
Σ	$\frac{1}{2}$	$\frac{1}{2}$															
b)	<table style="border-collapse: collapse; text-align: center;"> <tr> <td style="border-right: 1px solid black; padding: 5px;"></td> <td style="border-right: 1px solid black; padding: 5px;">1</td> <td style="border-right: 1px solid black; padding: 5px;">2</td> <td style="border-right: 1px solid black; padding: 5px;">Σ</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">1</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{4}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{4}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">2</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{4}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{4}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">Σ</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> <td style="border-right: 1px solid black; padding: 5px;">$\frac{1}{2}$</td> <td style="padding: 5px;"></td> </tr> </table>		1	2	Σ	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	Σ	$\frac{1}{2}$	$\frac{1}{2}$	
	1	2	Σ														
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$														
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$														
Σ	$\frac{1}{2}$	$\frac{1}{2}$															

In beiden Beispielen ist

$$P_{X_1} = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2 = P_{X_2}$$

und

$$P_{X_1} \times P_{X_2} = \frac{1}{4}\delta_{1,1} + \frac{1}{4}\delta_{1,2} + \frac{1}{4}\delta_{2,1} + \frac{1}{4}\delta_{2,2}.$$

Im Beispiel a) gilt offenbar $P_{X_1} \times P_{X_2} \neq P_{(X_1, X_2)}$ und deshalb sind X_1, X_2 abhängig.

Im Beispiel b) gilt offenbar $P_{X_1} \times P_{X_2} = P_{(X_1, X_2)}$ und deshalb sind X_1, X_2 unabhängig.

6.2 Sätze: (Fubini, Tonelli)

a) Seien $(\Omega_1, \mathcal{A}_1, P_1)$ und $(\Omega_2, \mathcal{A}_2, P_2)$ zwei beliebige Wahrscheinlichkeitsräume. Sei ferner $g \in \mathcal{L}_1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \times P_2)$, dann gilt

$$\int_{\Omega_1 \times \Omega_2} g(\omega_1, \omega_2) d(P_1 \times P_2)(\omega_1, \omega_2) = \int_{\Omega_1} \int_{\Omega_2} g(\omega_1, \omega_2) dP_2(\omega_2) dP_1(\omega_1) \quad (35)$$

b) Falls das Integral

$$\int_{\Omega_1} \int_{\Omega_2} |g(\omega_1, \omega_2)| dP_2(\omega_2) dP_1(\omega_1)$$

fast überall existiert, so ist $g \in \mathcal{L}_1(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \times P_2)$ und es gilt Gleichung (35).

Bemerkungen: a) Teil a) heißt Satz von *Fubini*, die Umkehrung dieses Satzes, also Teil

b) heißt Satz von *Tonelli*.

b) Die Reihenfolge der iterierten Integrale auf der rechten Seite in Gleichung (35) ist beliebig vertauschbar.

6.3 Beispiel: (Produktregel für Erwartungswerte)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X_1, X_2 \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ unabhängige Zufallsvariablen. Dann ist auch $X_1 \cdot X_2 \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ und es gilt

$$E(X_1 X_2) = E(X_1)E(X_2) \text{ [vgl. 5.15 b)].} \quad (36)$$

Beweis:

Offenbar gilt

$$\begin{aligned} \infty > E(X_1)E(X_2) &= \int_{\mathbb{R}} |x| dP_{X_1}(x) \int_{\mathbb{R}} |y| dP_{X_2}(y) = \int_{\mathbb{R}} \int_{\mathbb{R}} |xy| dP_{X_1}(x)dP_{X_2}(y) \\ &\stackrel{\text{Tonelli}}{=} \int_{\mathbb{R}^2} |xy| d(P_{X_1} \times P_{X_2})(x, y) \stackrel{\text{unabhängig}}{=} \int_{\mathbb{R}^2} |xy| dP_{(X_1, X_2)}(x, y) \\ &\stackrel{\text{Transformationssatz}}{\text{mit } h(x_1, x_2) = x_1 \cdot x_2}{=} \int_{\mathbb{R}} |z| dP_{X_1 \cdot X_2}(z) = E(|X_1 \cdot X_2|) \end{aligned}$$

Also ist $X_1 \cdot X_2 \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$. Die gleiche Rechnung ohne Beträge liefert dann Gleichung (36).¹⁶ □

¹⁶ Ende der sechzehnten Vorlesung vom Mi, 04.06.14

6.4 Faltungen von Wahrscheinlichkeitsmaßen auf \mathbb{R}^d :

Sei $A : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $(x, y) \mapsto x + y$ die Addition auf \mathbb{R}^d . (Diese Abbildung ist natürlich stetig und deshalb insbesondere messbar.)

Für $\mu, \nu \in \mathcal{M}^1(\mathbb{R}^d)$ heißt

$$\mu \star \nu := A(\mu \times \nu) \in \mathcal{M}^1(\mathbb{R}^d)$$

die *Faltung* von μ, ν .

Bemerkungen: a) Für alle $B \in \mathcal{B}(\mathbb{R}^d)$ gilt

$$(\mu \star \nu)(B) = (A(\mu \times \nu))(B) = (\mu \times \nu)(A^{-1}(B)) = (\mu \times \nu)(\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mid x + y \in B\}).$$

b) Seien $d = 1$, $b \in \mathbb{R}$, $B = (-\infty, b]$ und $F_{\mu+\nu}$ die Verteilungsfunktion von $\mu + \nu \in \mathcal{M}^1(\mathbb{R}^d)$. Dann gilt

$$F_{\mu+\nu}(b) = (\mu \star \nu)(B) = (\mu \times \nu)(\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \mid x + y \leq b\}).$$

6.5 Satz:

Seien $X_1, X_2 : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}^d$ unabhängige Zufallsvariablen. Dann ist

$$P_{X_1+X_2} = P_{X_1} \star P_{X_2}.$$

Beweis:

a) Für beliebige messbare Abbildungen A gilt

$$(\Omega, P) \xrightarrow{(X_1, X_2)} (\mathbb{R}^d \times \mathbb{R}^d, P_{X_1, X_2} = (X_1, X_2)(P)) \xrightarrow{A} (\mathbb{R}^d, A(P_{X_1, X_2}) = (A \circ (X_1, X_2))(P))$$

b) Also folgt insbesondere für die Addition A sofort

$$P_{X_1+X_2} = P_{A(X_1, X_2)} = A(P_{(X_1, X_2)})$$

□

6.6 Rechenregeln:

Seien $\mu, \nu \in \mathcal{M}^1(\mathbb{R}^d)$. Dann hat man:

a) Für alle $f \in \mathcal{L}_1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu \star \nu)$ gilt

$$\int_{\mathbb{R}^d} f d(\mu \star \nu) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x + y) d\mu(x) d\nu(y), \tag{37}$$

denn mit dem abstrakten Transformationssatz 5.11 und dem Satz von Fubini 6.2 folgt

$$\begin{aligned} \int_{\mathbb{R}^d} f d(\mu \star \nu) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} f dA(\mu \times \nu) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(A(x, y)) d(\mu \times \nu)(x, y) = \int_{\mathbb{R}^d \times \mathbb{R}^d} f(x + y) d(\mu \times \nu)(x, y) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x + y) d\mu(x) d\nu(y). \end{aligned}$$

b) Für alle $\lambda \in [0, 1]$ und $\mu, \nu, \rho \in \mathcal{M}^1(\mathbb{R}^d)$ besitzt die Faltung die folgenden algebraischen Eigenschaften

- i) $\mu \star \nu = \nu \star \mu$ (Kommutativität)
- ii) $(\mu \star \nu) \star \rho = \mu \star (\nu \star \rho)$ (Assoziativität)
- iii) $(\lambda\mu + (1 - \lambda)\nu) \star \rho = \lambda(\mu \star \rho) + (1 - \lambda)(\nu \star \rho)$ (Distributivität)

Diese Eigenschaften ergeben sich leicht durch Nachrechnen der Definition unter Verwendung von Teil a). Deshalb wird hier nur die Assoziativität bewiesen:

Für alle $B \in \mathcal{B}(\mathbb{R}^d)$ gilt:

$$\begin{aligned} ((\mu \star \nu) \star \rho)(B) &= \int_{\mathbb{R}^d} \mathbb{1}_B d((\mu \star \nu) \star \rho) \stackrel{a)}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B(x + y) d(\mu \star \nu)(x) d\rho(y) \\ &\stackrel{a)}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B((x + z) + y) d\mu(x) d\nu(z) d\rho(y) \stackrel{a)}{=} \dots \stackrel{a)}{=} \mu \star (\nu \star \rho). \end{aligned}$$

6.7 Diskrete Beispiele:

a) Für alle $x, y \in \mathbb{R}^d$ gilt

$$\delta_x \star \delta_y = \delta_{x+y}, \tag{38}$$

denn betrachte die beiden konstanten Zufallsvariablen

$$X = x \text{ und } Y = y.$$

Diese sind offenbar unabhängig und mit Satz 6.5 folgt sofort

$$\delta_{X+Y} = P_{X+Y} = P_X \star P_Y = \delta_X \star \delta_Y.$$

b) Seien $\mu, \nu \in \mathcal{M}^1(\mathbb{N}_0)$ mit zugehörigen Zähldichten $f_\mu, f_\nu : \mathbb{N}_0 \rightarrow [0, 1]$. Also lassen sich μ, ν darstellen als

$$\mu = \sum_{n \in \mathbb{N}_0} f_\mu(n) \delta_n \text{ und } \nu = \sum_{n \in \mathbb{N}_0} f_\nu(n) \delta_n.$$

7 Ferner ist $\mu \star \nu \in \mathcal{M}^1(\mathbb{N}_0)$ mit

$$\mu \star \nu = \left(\sum_{n \in \mathbb{N}_0} f_\mu(n) \delta_n \right) \star \left(\sum_{k \in \mathbb{N}_0} f_\nu(k) \delta_k \right) = \sum_{n, k \in \mathbb{N}_0} f_\mu(n) f_\nu(k) \delta_{n+k} = \sum_{m \in \mathbb{N}_0} \left(\sum_{k=0}^m f_\mu(k) f_\nu(m-k) \right) \delta_m.$$

Also ist $\mu \star \nu$ das Wahrscheinlichkeitsmaß auf \mathbb{N}_0 mit der Zähldichte

$$f_{\mu+\nu}(m) = \sum_{k=0}^m f_\mu(k) \cdot f_\nu(m-k) \text{ für } m \in \mathbb{N}_0.$$

c) Seien $\mu = \pi_{\lambda_1}$ und $\nu = \pi_{\lambda_2}$ Poissonverteilungen mit $\lambda_1, \lambda_2 > 0$. Dann gilt

$$\begin{aligned} f_{\pi_{\lambda_1} \star \pi_{\lambda_2}}(m) &= \sum_{k=0}^m e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{m-k}}{(m-k)!} = \frac{e^{-(\lambda_1+\lambda_2)} \lambda_2^m}{m!} \underbrace{\sum_{k=0}^m \binom{m}{k} \left(\frac{\lambda_1}{\lambda_2}\right)^k}_{\text{Binomischer Satz}} \\ &= \frac{e^{-(\lambda_1+\lambda_2)} \lambda_2^m}{m!} \cdot \left(1 + \frac{\lambda_1}{\lambda_2}\right)^m = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1 + \lambda_2)^m}{m!} = \pi_{\lambda_1+\lambda_2}(\{m\}). \end{aligned}$$

Also gilt $\pi_{\lambda_1} \star \pi_{\lambda_2} = \pi_{\lambda_1+\lambda_2}$.

6.8 Beispiel: (Negative Binomialverteilung)

Sei $r \in \mathbb{N}$ und $p \in (0, 1)$. Die Seiten einer Münze seien mit 0 und 1 bezeichnet, wobei Seite 1 mit Wahrscheinlichkeit p und 0 mit $1-p$ auftritt. Die Münze wird nun mehrfach unabhängig voneinander geworfen. Sei die \mathbb{N}_0 -wertige Zufallsvariable T_r die Zahl der Nullen bis zum r -ten Mal eine 1 auftritt. Was ist nun das Wahrscheinlichkeitsmaß P_{T_r} ? Für $r = 1$ hat man die geometrische Verteilung

$$P_{T_r} = \sum_{k=0}^{\infty} p(1-p)^k \delta_k.$$

Nun kann T_r als Summe der unabhängigen identisch verteilten Zufallsvariablen $T_1^{(1)}, \dots, T_1^{(r)}$ betrachtet werden. Es gilt also

$$P_{T_r} = P_{T_1^{(1)}} \star \dots \star P_{T_1^{(r)}} = P_{T_1} \star \dots \star P_{T_1} =: (P_{T_1})^{(r)}.$$

Diese Verteilung kann nun geschickt mit Hilfe der Erzeugendenfunktion berechnet werden:

$$\begin{aligned} g_{T_1}(t) &= \sum_{k=0}^{\infty} p(1-p)^k t^k = p \sum_{k=0}^{\infty} ((1-p)t)^k = \frac{p}{1-(1-p)t} \\ \stackrel{5.24 \text{ a)}}{\Rightarrow} g_{T_r}(t) &= \frac{p^r}{(1-(1-p)t)^r}. \end{aligned}$$

Für die k -te Ableitung folgt dann

$$g_{T_r}^{(k)}(t) = \frac{p^r(-r)(-r-1)\cdots(-r-k+1)(p-1)^k}{(1-(1-p)t)^{r+k}} \text{ für } k \in \mathbb{N}_0$$

und

$$g_{T_r}^{(k)}(0) = p^r(-r)(-r-1)\cdots(-r-k+1)(p-1)^k \text{ für } k \in \mathbb{N}_0.$$

Andererseits gilt aber auch

$$\sum_{k=0}^{\infty} P(T_r = k)t^k \stackrel{\text{Definition}}{=} g_{T_r}(t) \stackrel{\text{Taylor}}{=} \sum_{k=0}^{\infty} \frac{g_{T_r}^{(k)}(0)}{k!} t^k = \sum_{k=0}^{\infty} \binom{-r}{k} p^r (p-1)^k t^k.$$

Definition: (Negative Binomialverteilung)

Die *negative Binomialverteilung* $NB_{r,p}$ ist das Wahrscheinlichkeitsmaß auf \mathbb{N}_0 mit der Zähldichte

$$f_{r,p}(k) := \binom{-r}{k} p^r (p-1)^k.$$

6.9 Lemma:

Seien $\mu, \nu \in \mathcal{M}^1(\mathbb{R}^d)$, wobei ν eine Lebesgue-dichte $f_\nu \in \mathcal{L}_1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$ habe. Dann hat auch $\mu \star \nu \in \mathcal{M}^1(\mathbb{R}^d)$ eine Lebesgue-dichte $f_{\mu \star \nu} \in \mathcal{L}_1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$ mit

$$f_{\mu \star \nu}(x) = \mu \star f_\nu(x) := \int_{\mathbb{R}^d} f_\nu(x-y) d\mu(y) \text{ für fast alle } x \in \mathbb{R}^d.$$

Beweis:

Die Abbildung $x \mapsto \mu \star f_\nu(x)$ ist $[0, \infty]$ -wertig und wohldefiniert. Ferner gilt für alle $B \in \mathcal{B}(\mathbb{R}^d)$ stets

$$\begin{aligned} (\mu \star \nu)(B) &= \int_{\mathbb{R}^d} \mathbb{1}_B d(\mu \star \nu) \stackrel{6.6 \text{ a)}}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B(x+y) d\mu(x) \underbrace{d\nu(y)}_{=f_\nu(y)dy} \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B(x+y) d\mu(x) f_\nu(y) dy \stackrel{\text{Fubini}}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B(x+y) f_\nu(y) dy d\mu(x) \\ &\stackrel{z:=x+y}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B(z) f_\nu(z-x) dz d\mu(x) \stackrel{\text{Fubini}}{=} \int_{\mathbb{R}^d} \underbrace{\int_{\mathbb{R}^d} f_\nu(z-x) d\mu(x) \mathbb{1}_B(z) dz}_{= \mu \star f_\nu(z)} \\ &= ((\mu \star f_\nu)\lambda^d)(B) \end{aligned}$$

Also gilt $\mu \star \nu = (\mu \star f_\nu)\lambda^d$. □

6.10 Beispiel: (Faltung von Normalverteilungen auf \mathbb{R}^d)

Für alle $\mu_1, \mu_2 \in \mathbb{R}$ und alle $\sigma_1^2, \sigma_2^2 > 0$ gilt:

a) $N_{\mu_1, \sigma_1^2} \star N_{\mu_2, \sigma_2^2} = N_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2}$.

b) $\delta_{\mu_1} \star N_{0, \sigma_1^2} = N_{\mu_1, \sigma_1^2}$.

Beweis:

a) Wegen

$$N_{\mu_1, \sigma_1^2} \star N_{\mu_2, \sigma_2^2} \stackrel{b)}{=} \delta_{\mu_1} \star N_{0, \sigma_1^2} \star \delta_{\mu_2} \star N_{0, \sigma_2^2} = \delta_{\mu_1 + \mu_2} \star N_{0, \sigma_1^2} \star N_{0, \sigma_2^2}$$

und

$$\delta_{\mu_1 + \mu_2} \star N_{0, \sigma_1^2 + \sigma_2^2} \stackrel{b)}{=} N_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2}$$

genügt es

$$N_{0, \sigma_1^2} \star N_{0, \sigma_2^2} = N_{0, \sigma_1^2 + \sigma_2^2}$$

zu zeigen.

N_{0, σ_i^2} hat die Dichte $f_{0, \sigma_i^2}(x) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{x^2}{2\sigma_i^2}}$ für $i = 1, 2$. Damit ergibt sich nach einiger Rechnung am Ende des Tages

$$(f_{0, \sigma_1^2} \star f_{0, \sigma_2^2})(x) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{2\sigma_1^2}} e^{-\frac{y^2}{2\sigma_2^2}} dy = \dots = K \cdot e^{-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}}$$

für eine geeignete Konstante $K \in \mathbb{R}$. Man kann also festhalten, dass $f_{0, \sigma_1^2} \star f_{0, \sigma_2^2}$ bis auf eine multiplikative Konstante gleich der Dichte von $N_{0, \sigma_1^2 + \sigma_2^2}$. Da aber

$$\int_{\mathbb{R}} (f_{0, \sigma_1^2} \star f_{0, \sigma_2^2})(x) dx = 1 = \int_{\mathbb{R}} f_{0, \sigma_1^2 + \sigma_2^2}(x) dx$$

gilt, folgt $f_{0, \sigma_1^2} \star f_{0, \sigma_2^2} = f_{0, \sigma_1^2 + \sigma_2^2}$ und somit die Behauptung.

b) Die Verteilung N_{0, σ_1^2} hat die zugehörige Dichtefunktion $f_{0, \sigma_1^2}(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{x^2}{2\sigma_1^2}}$. Mit Lemma 6.9 folgt dann sofort, dass $\delta_{\mu_1} \star N_{0, \sigma_1^2}$ die Dichte

$$(\delta_{\mu_1} \star f_{0, \sigma_1^2})(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \int_{\mathbb{R}} e^{-\frac{(x-y)^2}{2\sigma_1^2}} d\delta_{\mu_1}(y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} = f_{\mu_1, \sigma_1^2}(x)$$

besitzt. Dies ist die Dichte von N_{μ_1, σ_1^2} .¹⁷ □

¹⁷ Ende der siebzehnten Vorlesung vom Fr, 06.06.14

6.11 Definition: (Gamma-Verteilung und χ^2 -Verteilung)

a) Für $\alpha, \nu > 0$ ist die Gamma-Verteilung $\Gamma_{\alpha, \nu}$ definiert als das Wahrscheinlichkeitsmaß auf \mathbb{R} mit der Dichtefunktion

$$f_{\alpha, \nu}(x) := \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x} & , x > 0 \\ 0 & , x \leq 0 \end{cases}.$$

Dabei ist $\Gamma(\nu) := \int_0^\infty x^{\nu-1} e^{-x} dx \in (0, \infty)$ die *Gamma-Funktion*.

b) Die χ^2 -Verteilung χ_k^2 mit so genanntem Freiheitsgrad $k \in \mathbb{N}$ ist definiert als

$$\chi_k^2 := \Gamma_{\frac{1}{2}, \frac{k}{2}}.$$

6.12 Lemma:

Für $\alpha, \nu, \mu > 0$ gilt:

a) $\Gamma_{\alpha, \nu} \star \Gamma_{\alpha, \mu} = \Gamma_{\alpha, \nu+\mu}$ und $f_{\alpha, \nu} \star f_{\alpha, \mu} = f_{\alpha, \nu+\mu}$.

b) Für unabhängige $N_{0,1}$ -verteilte Zufallsvariablen $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ gilt

$$P_{X_1^2 + \dots + X_n^2} = \chi_n^2.$$

c) $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

d) Für die *Beta-Funktion* gilt

$$B(s, t) := \int_0^1 (1-x)^{s-1} x^{t-1} dx = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}$$

mit $s, t > 0$.

Beweis:

a) Für $x \leq 0$ gilt offenbar $(f_{\alpha, \nu} \star f_{\alpha, \mu})(x) = 0$. Für $x > 0$ ergibt sich nach einiger Rechnung am Ende des Tages

$$(f_{\alpha, \nu} \star f_{\alpha, \mu})(x) = \dots = f_{\alpha, \mu+\nu}(x) \cdot \frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)} \int_0^1 (1-z)^{\mu-1} z^{\nu-1} dz.$$

Dieselbe Argumentation wie im Beweis von Beispiel 6.10 a) liefert dann, dass

$$\frac{\Gamma(\mu)\Gamma(\nu)}{\Gamma(\mu+\nu)} \int_0^1 (1-z)^{\mu-1} z^{\nu-1} dz = 1$$

sein muss und somit Teil a) und d) bewiesen sind.

b) Beispiel 4.20 d) besagt: Ist X_1 eine \mathbb{R} -wertige $N_{0,1}$ -verteilte Zufallsvariable, so hat X_1^2 eine Verteilung auf \mathbb{R} mit der Dichte

$$f_{\frac{1}{2}, \frac{1}{2}}(x) := \begin{cases} \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} & , x > 0 \\ 0 & , x \leq 0 \end{cases}.$$

Die Zufallsvariable X_1^2 ist also χ_1^2 -verteilt und man erhält

$$\frac{\sqrt{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} = \frac{1}{\sqrt{2\pi}} \Leftrightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Also gilt Teil c).

Seien nun X_1, \dots, X_n unabhängige χ_1^2 -verteilte Zufallsvariablen. Dann sind auch X_1^2, \dots, X_n^2 unabhängige χ_1^2 -verteilte Zufallsvariablen und man erhält

$$P_{X_1^2 + \dots + X_n^2} = P_{X_1^2} \star \dots \star P_{X_n^2} = (\chi_1^2)^{(n)} \stackrel{a)}{=} \chi_n^2.$$

□

6.13 Definition und Lemma: (Totalvariationsabstand)

Für $\mu, \nu \in \mathcal{M}^1(\mathbb{N}_0)$ mit zugehörigen Zähldichten $f_\mu, f_\nu : \mathbb{N}_0 \rightarrow [0, 1]$ definiert man den sogenannten *Totalvariationsabstand*

$$d(\mu, \nu) := \sum_{n=0}^{\infty} |f_\mu(n) - f_\nu(n)| = \|f_\mu - f_\nu\|_1.$$

Für diese Metrik auf $\mathcal{M}^1(\mathbb{N}_0)$ und für alle $\mu, \nu, \rho \in \mathcal{M}^1(\mathbb{N}_0)$ gilt stets

$$d(\mu \star \rho, \nu \star \rho) \leq d(\mu, \nu).$$

(Man bezeichnet diese Eigenschaft auch als „Glättung durch Faltung“.)

Beweis:

Mit der Dreiecksungleichung für den Betrag folgt

$$\begin{aligned} d(\mu \star \rho, \nu \star \rho) &= \sum_{n=0}^{\infty} |(f_\mu \star f_\rho)(n) - (f_\nu \star f_\rho)(n)| = \sum_{n=0}^{\infty} \left| \sum_{k=0}^n (f_\mu(k) - f_\nu(k)) \cdot f_\rho(n-k) \right| \\ &\leq \sum_{n=0}^{\infty} \sum_{k=0}^n |f_\mu(k) - f_\nu(k)| \cdot f_\rho(n-k) \stackrel{l:=n-k}{=} \sum_{k=0}^{\infty} |f_\mu(k) - f_\nu(k)| \cdot \underbrace{\sum_{l=0}^{\infty} f_\rho(l)}_{=1} \\ &= d(\mu, \nu). \end{aligned}$$

□

6.14 Poissonapproximation von $B_{n,p}$:

Für alle $n \in \mathbb{N}$ und $p \in (0, 1)$ gilt

$$d(B_{n,p}, \pi_{n,p}) \leq 2np^2 = 2 \frac{(np)^2}{n}.$$

Beweis:

Bekanntermaßen gilt

$$B_{n,p} = \underbrace{B_{1,p} \star \cdots \star B_{1,p}}_{n\text{-mal}} = (B_{1,p})^{(n)}$$

und

$$\pi_{n,p} = \underbrace{\pi_p \star \cdots \star \pi_p}_{n\text{-mal}} = (\pi_p)^{(n)}.$$

Daraus folgt sofort mit der Dreiecksungleichung für Metriken

$$\begin{aligned} d(B_{n,p}, \pi_{n,p}) &= d((B_{1,p})^{(n)}, (\pi_p)^{(n)}) \stackrel{\Delta\text{-Ungl.}}{\leq} \sum_{k=0}^{n-1} d((\pi_p)^{(k)} \star (B_{1,p})^{(n-k)}, (\pi_p)^{(k+1)} \star (B_{1,p})^{(n-(k+1))}) \\ &\stackrel{6.13}{\leq} \sum_{k=0}^{n-1} d(B_{1,p}, \pi_p) = n \cdot d(B_{1,p}, \pi_p). \end{aligned}$$

Ferner ist

$$d(B_{1,p}, \pi_p) = \underbrace{|(1-p) - e^{-p}|}_{\leq 0} + \underbrace{|p - pe^{-p}|}_{\geq 0} + \sum_{n \geq 2} \underbrace{\left| 0 - e^{-p} \cdot \frac{p^n}{n!} \right|}_{\leq 0 \ \forall n \geq 2} = 2 \cdot p(1 - e^{-p}) \leq 2p^2.$$

Denn es gilt $\sum_{n=0}^{\infty} (B_{1,p}(\{n\}) - \pi_p(\{n\})) = 0$ und $1 - e^{-p} \leq p$. □

7 Konvergenz von Zufallsvariablen

7.1 Definition: (Nullmenge)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

a) Eine Menge $A \in \mathcal{A}$ heißt $(P-)$ Nullmenge, falls $P(A) = 0$ ist. A heißt *fast sicher* (f.s.), falls $P(A) = 1$ bzw. $P(\Omega \setminus A) = 0$ ist.

b) Konvention: Seien $X_1, X_2 : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen. Dann bedeutet „ $X_1 = X_2$ f.s.“, dass

$$P(\{\omega \in \Omega \mid X_1(\omega) = X_2(\omega)\}) = 1$$

ist.

Bemerkungen:

(i) Sei $(A_n) \subseteq \mathcal{A}$ eine Folge von Nullmengen, dann ist auch

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

wieder eine Nullmenge.

(ii) Sei $(A_n) \subseteq \mathcal{A}$ eine Folge fast sicherer Mengen, dann ist auch

$$\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

wieder fast sicher.

Im Folgenden seien immer (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X_n, X : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen. Ferner wird die Schreibweise

$$|x| := \|x\|_2 = \sqrt{x_1^2 + \dots + x_d^2}, \quad \forall x \in \mathbb{R}^d$$

verwendet.

7.2 Definition: (Konvergenzbegriffe)

a) Man sagt $X_n \rightarrow X$ für $n \rightarrow \infty$ *stochastisch*, falls

$$\forall \varepsilon > 0 : P(|X_n - X| > \varepsilon) = P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}) \rightarrow 0 \text{ für } n \rightarrow \infty.$$

b) Für $p = 1, 2$ konvergiert (X_n) im p -ten Mittel gegen X , falls $X_n, X \in \mathcal{L}_p(\Omega, \mathcal{A}, P)$ für alle $n \in \mathbb{N}$ und

$$\|X_n - X\|_p^p = \int_{\Omega} |X_n(\omega) - X(\omega)|^p dP(\omega) \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Man schreibt dann auch kurz $X_n \xrightarrow{\|\cdot\|_p} X$ für $n \rightarrow \infty$.

c) Man sagt $X_n \rightarrow X$ für $n \rightarrow \infty$ *fast sicher*, falls

$$P(X_n \xrightarrow[n \rightarrow \infty]{} X) = P(\{\omega \in \Omega | X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)\}) = 1.$$

Bemerkungen:

(i) Die fast sichere Konvergenz ist einfach die punktweise Konvergenz außerhalb einer Nullmenge.

(ii) Die stochastische Konvergenz tritt im schwachen Gesetz der großen Zahlen (vgl. 5.29) auf.

(iii) Die fast sichere Konvergenz ist eine „schwache“ Konvergenz, die stochastische Konvergenz ist sogar noch „schwächer“.

(iv) Auf dem Raum $\mathcal{L}_p(\Omega, \mathcal{A}, P)$ wird durch

$$\|X_n - X\|_p := \left(\int_{\Omega} |X_n(\omega) - X(\omega)|^p dP(\omega) \right)^{\frac{1}{p}}$$

die sogenannte \mathcal{L}_p -(Halb-)Norm definiert.¹⁸

7.3 Satz:

Es gilt

a) $X_n \xrightarrow{\|\cdot\|_2} X \Rightarrow X_n \xrightarrow{\|\cdot\|_1} X$ für $n \rightarrow \infty$.

b) $X_n \xrightarrow{\|\cdot\|_1} X \Rightarrow X_n \xrightarrow{\text{stoch.}} X$ für $n \rightarrow \infty$.

Beweis:

a) Mit der Cauchy-Schwarz Ungleichung folgt

$$\|X_n - X\|_1 = \int_{\Omega} 1 \cdot |X_n - X| dP \leq \|X_n - X\|_2 \cdot \|1\|_2 = \|X_n - X\|_2 \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Man hat also die Inklusion $\mathcal{L}_2(\Omega, \mathcal{A}, P) \subseteq \mathcal{L}_1(\Omega, \mathcal{A}, P)$.

b) Sei $\varepsilon > 0$ fest, dann folgt mit der Markov-Ungleichung 5.26 sofort

$$P(|X_n - X| > \varepsilon) \leq \frac{E(|X_n - X|)}{\varepsilon} = \frac{1}{\varepsilon} \int_{\Omega} |X_n - X| dP = \frac{1}{\varepsilon} \|X_n - X\|_1 \rightarrow 0 \text{ für } n \rightarrow \infty.$$

□

¹⁸ Ende der achtzehnten Vorlesung vom Mi, 11.06.14

Bemerkung: Die Rückrichtungen in 7.3 sind jeweils falsch (vgl. Beispiele in den Übungen).

7.4 Definition: (Limes Inferior und Limes Superior von Mengen)

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(A_n) \subseteq \mathcal{A}$ eine Folge von Mengen. Dann definiert man

$$\liminf_{n \rightarrow \infty} A_n := \bigcup_{i \in \mathbb{N}} \left(\bigcap_{j \geq i} A_j \right) \in \mathcal{A}$$

und

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{i \in \mathbb{N}} \left(\bigcup_{j \geq i} A_j \right) \in \mathcal{A}.$$

7.5 Lemma:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(A_n) \subseteq \mathcal{A}$ eine Folge von Mengen. Dann gilt:

- a) $\limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N}\}.$
- b) $\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ bis auf endlich viele } n \in \mathbb{N}\}.$
- c) $\liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n.$
- d) $P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n)$

Beweis.

a) Sei $\omega \in \Omega$. Dann gilt

$$\omega \in \limsup_{n \rightarrow \infty} A_n \Leftrightarrow \omega \in \bigcap_{i \in \mathbb{N}} \left(\bigcup_{j \geq i} A_j \right) \Leftrightarrow \forall i \in \mathbb{N} : \omega \in \bigcup_{j \geq i} A_j \Leftrightarrow \forall i \in \mathbb{N} \exists j \geq i : \omega \in A_j.$$

Also liegt $\omega \in A_j$ für unendlich viele j .

b) Analog zu Teil a) [vgl. Übungen].

c) Folgt sofort aus Teil a) und b).

d) Betrachte die Menge $B_n := \bigcup_{k \geq n} A_k$. Für diese gilt natürlich $B_{n+1} \subseteq B_n$ und $\limsup_{n \rightarrow \infty} A_n =$

$\bigcap_{n \in \mathbb{N}} B_n$ für alle $n \in \mathbb{N}$. Weiter folgt mit 4.4 a) sofort

$$P(\limsup_{n \rightarrow \infty} A_n) = P\left(\bigcap_{n \in \mathbb{N}} B_n\right) = \lim_{n \rightarrow \infty} P(B_n) = \limsup_{n \rightarrow \infty} P(B_n) \stackrel{B_n \supseteq A_n}{\geq} \limsup_{n \rightarrow \infty} P(A_n).$$

Die Ungleichung $P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n)$ folgt analog [vgl. Übungen]. \square

7.6 Lemma: (Borel-Cantelli)

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A_n \in \mathcal{A}$ für alle $n \in \mathbb{N}$.

a) Falls $\sum_{n=1}^{\infty} P(A_n) < \infty$, so gilt $P(\limsup_{n \rightarrow \infty} A_n) = 0$.

b) Falls $\sum_{n=1}^{\infty} P(A_n) = \infty$ und alle A_n unabhängig sind, so gilt $P(\limsup_{n \rightarrow \infty} A_n) = 1$.
(So etwas wird auch ein „0-1-Gesetz“ genannt.)

Beweis:

a) Man erhält sofort

$$0 \leq P(\limsup_{n \rightarrow \infty} A_n) = P\left(\bigcap_{i \in \mathbb{N}} \left(\bigcup_{j \geq i} A_j\right)\right) \stackrel{\forall i \in \mathbb{N}}{\leq} P\left(\bigcup_{j \geq i} A_j\right) \leq \sum_{j=i}^{\infty} P(A_j) \rightarrow 0 \text{ für } i \rightarrow \infty,$$

denn nach Voraussetzung gilt $\sum_{n=1}^{\infty} P(A_n) < \infty$.

b) Bekanntermaßen gilt für alle $x \in \mathbb{R}$ stets

$$1 - x \leq e^{-x}.$$

Deshalb gilt auch für alle $1 \leq m \leq k \in \mathbb{N}$

$$0 \leq \prod_{i=m}^k (1 - P(A_i)) \leq \prod_{i=m}^k e^{-P(A_i)} = e^{-\sum_{i=m}^k P(A_i)} \rightarrow 0 \text{ für } k \rightarrow \infty.$$

Also gilt für alle $m \in \mathbb{N}$

$$\lim_{k \rightarrow \infty} \prod_{i=m}^k (1 - P(A_i)) = 0. \quad (39)$$

Weiterhin folgt

$$\begin{aligned} 1 - P(\limsup_{n \rightarrow \infty} A_n) &= 1 - \lim_{n \rightarrow \infty} P\left(\bigcup_{j \geq n} A_j\right) = \lim_{n \rightarrow \infty} P\left(\Omega \setminus \left(\bigcup_{j \geq n} A_j\right)\right) = \lim_{n \rightarrow \infty} P\left(\bigcap_{j \geq n} A_j^C\right) \\ &= \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} P\left(\bigcap_{j=n}^k A_j^C\right) = \lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \prod_{j=n}^k (1 - P(A_j)) \stackrel{(39)}{=} \lim_{n \rightarrow \infty} 0 = 0. \end{aligned}$$

□

7.7 Beispiel:

Eine Box enthalte anfangs eine weiße und eine schwarze Kugel. Man zieht nun zufällig eine Kugel, notiert die Farbe und legt die Kugel samt einer weißen Kugel in die Box. Dieser Vorgang wird beliebig oft wiederholt.

Bei der ersten Ziehung hat man:

Einmal weiß und einmal schwarz in der Box.

Bei der n -ten Ziehung hat man:

n -mal weiß und einmal schwarz in der Box.

Für das Ereignis

$$A_n := \{, \text{Bei der } n\text{-ten Ziehung wird schwarz gezogen} \}$$

hat man

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

und deshalb

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

Offenbar sind alle A_n unabhängig und mit 7.6 b) folgt

$$P(\limsup_{n \rightarrow \infty} A_n) = 1.$$

Das Ereignis unendlich oft schwarz zu ziehen bei unendlich vielen Ziehungen ist also fast sicher.

7.8 Satz:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X_n, X : \Omega \rightarrow \mathbb{R}$ Zufallsvariablen. Dann gilt

$$X_n \rightarrow X \text{ f.s.} \Rightarrow X_n \xrightarrow[\text{stoch.}]{} X \text{ für } n \rightarrow \infty.$$

Beweis:

Sei $\varepsilon > 0$ und $n \in \mathbb{N}$. Betrachte die Menge

$$B_n := \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \varepsilon\}.$$

Dann gilt

$$\begin{aligned} 0 \leq \liminf_{n \rightarrow \infty} P(B_n) &\leq \limsup_{n \rightarrow \infty} P(B_n) \stackrel{7.5 \text{ d)}}{\leq} P(\limsup_{n \rightarrow \infty} B_n) \\ &\stackrel{7.5 \text{ a)}}{\leq} P(|X_n - X| > \varepsilon \text{ für unendlich viele } n \in \mathbb{N}) \\ &\leq P(\{\omega \in \Omega \mid X_n(\omega) \not\rightarrow X(\omega) \text{ für } n \rightarrow \infty\}) = 0, \end{aligned}$$

da $X_n \rightarrow X$ f.s. .

□

7.9 Satz: (Ein starkes Gesetz der großen Zahlen)

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(X_n) \subseteq \mathcal{L}_2(\Omega, \mathcal{A}, P)$ eine Folge von reellwertigen unkorrelierten Zufallsvariablen mit $M := \sup_{n \in \mathbb{N}} \text{Var}(X_n) < \infty$. Dann gilt

$$Z_n := \frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \rightarrow 0 \text{ f.s. für } n \rightarrow \infty.$$

Beweis:

Sei O.E. $E(X_k) = 0$ für alle $k \in \mathbb{N}$, sonst betrachte die Zufallsvariable $X_k - E(X_k)$ für alle $k \in \mathbb{N}$.

a) Es wird zunächst gezeigt, dass

$$Z_{n^2} \rightarrow 0 \text{ f.s. für } n \rightarrow \infty$$

gilt. Betrachte dazu

$$\text{Var}(Z_{n^2}) = \frac{1}{n^4} \text{Var}\left(\sum_{k=1}^{n^2} X_k\right) = \frac{1}{n^4} \sum_{k=1}^{n^2} \text{Var}(X_k) \leq \frac{1}{n^4} \cdot n^2 \cdot M = \frac{M}{n^2}.$$

Nun liefert die Tschebyscheff-Ungleichung 5.27 b) für alle $\varepsilon > 0$

$$P(|Z_{n^2}| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var}(Z_{n^2}) \leq \frac{M}{n^2 \varepsilon^2}.$$

Betrachte ferner die Menge

$$A_n := \{|Z_{n^2}| \geq \varepsilon\} = \{\omega \in \Omega \mid |Z_{n^2}(\omega)| \geq \varepsilon\} \in \mathcal{A}.$$

Dann ist

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

Das Lemma von Borel-Cantelli 7.6 a) liefert nun

$$P(\limsup_{n \rightarrow \infty} A_n) = 0.$$

Also gilt für $\varepsilon > 0$ stets

$$1 = P(\Omega \setminus \limsup_{n \rightarrow \infty} A_n) = P(\{\omega \in \Omega \mid |Z_{n^2}(\omega)| \geq \varepsilon \text{ für nur endlich viele } n \in \mathbb{N}\}).$$

Daraus folgt nun

$$\begin{aligned} 1 &= P\left(\bigcap_{k \in \mathbb{N}} \{\omega \in \Omega \mid |Z_{n^2}(\omega)| \geq \frac{1}{k} \text{ für nur endlich viele } n \in \mathbb{N}\}\right) \\ &= P\left(\{\omega \in \Omega \mid \forall k \in \mathbb{N} \exists m \in \mathbb{N} \forall n \geq m : |Z_{n^2}(\omega)| < \frac{1}{k}\}\right) \\ &= P(\{\omega \in \Omega \mid Z_{n^2}(\omega) \rightarrow 0 \text{ für } n \rightarrow \infty\}). \end{aligned}$$

b) Zu jedem $m \in \mathbb{N}$ wähle ein $n(m) \in \mathbb{N}$ mit der Eigenschaft

$$n(m)^2 \leq m \leq (n(m) + 1)^2.$$

Vergleiche nun Z_m und $Z_{n(m)^2}$. Dazu definiert man

$$S_k := \sum_{j=1}^k X_j.$$

Dann gilt natürlich

$$\text{Var}(S_m - S_{n(m)^2}) \leq M(m - n(m)^2).$$

Nun liefert die Tschebyscheff-Ungleichung 5.27 b) für alle $\varepsilon > 0$

$$P(|S_m - S_{n(m)^2}| \geq \varepsilon \cdot n(m)^2) \leq \frac{1}{\varepsilon^2 \cdot n(m)^4} M(m - n(m)^2).$$

Daraus folgt weiter

$$\sum_{m=1}^{\infty} P\left(\frac{1}{n(m)^2} |S_m - S_{n(m)^2}| \geq \varepsilon\right) \leq \frac{M}{\varepsilon^2} \underbrace{\sum_{n=1}^{\infty} \left(\sum_{m=n^2}^{(n+1)^2-1} \frac{m - n^2}{n^4} \right)}_{= \frac{1}{n^4} \sum_{k=0}^{2n} k = \frac{1}{n^4} \frac{1n(2n+1)}{2}} < \infty.$$

Wie in Teil a) liefert nun das Lemma von Borel-Cantelli 7.6 a) für alle $k \in \mathbb{N}$

$$1 = P(\{\omega \in \Omega \mid \frac{1}{n(m)^2} |S_m - S_{n(m)^2}| \geq \frac{1}{k} \text{ für endlich viele } m \in \mathbb{N}\}).$$

Daraus folgt analog zu a) wieder

$$\frac{1}{n(m)^2} (S_m - S_{n(m)^2}) \rightarrow 0 \text{ f.s. für } m \rightarrow \infty.$$

Nach a) hat man allerdings

$$\frac{1}{n(m)^2} S_{n(m)^2} = Z_{n(m)^2} \rightarrow 0 \text{ f.s. für } m \rightarrow \infty.$$

Daraus folgt sofort

$$\frac{1}{n(m)^2} S_m \rightarrow 0 \text{ f.s. für } m \rightarrow \infty.$$

Ferner gilt $m \geq n(m)^2$, also auch

$$\frac{1}{m} \leq \frac{1}{n(m)^2}$$

und deshalb folgt letztlich

$$\frac{1}{m} S_m \rightarrow 0 \text{ f.s. für } m \rightarrow \infty.$$

□

Bemerkungen:

a) Seien $X_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$ unkorreliert und identisch verteilt, d.h. $P(X_n) = P(X_1) \forall n \in \mathbb{N}$. Dann gilt natürlich auch

$$E(X_n) = E(X_1) \text{ und } Var(X_n) = Var(X_1) = M \quad \forall n \in \mathbb{N}.$$

In dieser Situation liefert dann 7.9

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow E(X_1) \text{ f.s. für } n \rightarrow \infty.$$

b) Falls alle X_n unabhängig sind, sind sie nach 5.22 c) auch unkorreliert.

c) Satz 7.9 und Satz 7.8 implizieren die folgende Version des schwachen Gesetzes der großen Zahlen¹⁹

$$\frac{1}{n} \sum_{k=1}^n (X_k - E(X_k)) \xrightarrow{\text{stoch}} 0 \text{ für } n \rightarrow \infty. \quad [\text{vgl. 5.29}]$$

Es gibt mehrere Gesetze der großen Zahlen. Die folgende Variante von Kolmogorov wird ohne Beweis angegeben.

7.10 Satz: (Ein starkes Gesetz der großen Zahlen von Kolmogorov)

Seien $X_n \in \mathcal{L}_1(\Omega, \mathcal{A}, P)$ unabhängig und identisch verteilt. Dann gilt

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow E(X_1) \text{ f.s. für } n \rightarrow \infty.$$

7.11 Definition: (Konvergenz in Verteilung und schwache Konvergenz von Maßen)

a) Der Träger $\text{supp } f$ einer Funktion $f: \mathbb{R}^d \rightarrow \mathbb{C}$ ist der Abschluss der *Nichtnullstellenmenge* von f

$$\text{supp } f := \overline{\{x \in \mathbb{R}^d \mid f(x) \neq 0\}} \subseteq \mathbb{R}.$$

b) Man schreibt

$$\mathcal{C}_B(\mathbb{R}^d) := \{f: \mathbb{R}^d \rightarrow \mathbb{C} \mid f \text{ stetig und beschränkt}\}$$

für den Vektorraum der stetigen beschränkten Funktionen auf \mathbb{R}^d und

$$\mathcal{C}_C(\mathbb{R}^d) := \{f \in \mathcal{C}_B(\mathbb{R}^d) \mid \text{supp } f \text{ kompakt}\}$$

¹⁹ Ende der neunzehnten Vorlesung vom Fr, 13.06.14

für den Vektorraum der stetigen Funktionen mit kompaktem Träger auf \mathbb{R}^d .

c) Seien $\mu, \mu_n \in \mathcal{M}^1(\mathbb{R}^d)$. Man sagt μ_n *konvergiert schwach* gegen μ , falls für alle $f \in \mathcal{C}_B(\mathbb{R}^d)$ stets

$$\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu \text{ für } n \rightarrow \infty$$

gilt. Man schreibt dann

$$\mu_n \xrightarrow{\text{schwach}} \mu \text{ für } n \rightarrow \infty.$$

d) Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X_n, X : \Omega \rightarrow \mathbb{R}^d$ Zufallsvariablen. Man sagt X_n *konvergiert in Verteilung* gegen X , falls

$$P_{X_n} \xrightarrow{\text{schwach}} P_X \text{ für } n \rightarrow \infty.$$

Man schreibt dann

$$X_n \xrightarrow{\text{Vert.}} X \text{ für } n \rightarrow \infty.$$

7.12 Lemma:

Seien $\mu, \mu_n \in \mathcal{M}^1(\mathbb{R}^d)$. Falls für alle $f \in \mathcal{C}_C(\mathbb{R}^d)$ stets

$$\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu \text{ für } n \rightarrow \infty$$

gilt, so folgt

$$\mu_n \xrightarrow{\text{schwach}} \mu \text{ für } n \rightarrow \infty.$$

Beweis:

Sei $f \in \mathcal{C}_B(\mathbb{R}^d)$ und $\varepsilon > 0$. Wähle nun $f_k \in \mathcal{C}_C(\mathbb{R}^d)$ mit

$$f_k(x) := \begin{cases} 1 & , \|x\| \leq k \\ 0 & , \|x\| \geq k+1 \\ \in [0, 1] & , \text{sonst} \end{cases}$$

Dann gilt offenbar $f_k \uparrow 1$ punktweise für $k \rightarrow \infty$ und der Satz über monotone Konvergenz 5.9 a) liefert

$$\int_{\mathbb{R}^d} f_k d\mu \uparrow \int_{\mathbb{R}^d} 1 d\mu = 1,$$

da μ ein Wahrscheinlichkeitsmaß ist. Daraus folgt nun aber, dass ein $k_0 \in \mathbb{N}$ existiert mit der Eigenschaft

$$\int_{\mathbb{R}^d} f_{k_0} d\mu > 1 - \frac{\varepsilon}{2}. \quad (40)$$

Nach Konstruktion ist auch f_{k_0} in $\mathcal{C}_C(\mathbb{R}^d)$ und man erhält nach Voraussetzung

$$\int_{\mathbb{R}^d} f_{k_0} d\mu_n \rightarrow \int_{\mathbb{R}^d} f_{k_0} d\mu \text{ für } n \rightarrow \infty.$$

Für $n \geq n_0$ folgt

$$\left| \int_{\mathbb{R}^d} \underbrace{f(1-f_{k_0})}_{\geq 0} d\mu_n \right| \leq \|f\|_{\text{sup}} \int_{\mathbb{R}^d} 1 - f_{k_0} d\mu_n = \|f\|_{\text{sup}} \left(1 + \int_{\mathbb{R}^d} f_{k_0} d\mu_n \right) \stackrel{(40)}{<} \|f\|_{\text{sup}}(1 - (1 - \varepsilon)) = \|f\|_{\text{sup}}\varepsilon.$$

Analog folgt auch

$$\left| \int_{\mathbb{R}^d} f(1-f_{k_0}) d\mu \right| < \|f\|_{\text{sup}}\varepsilon.$$

Daraus folgt nun abschließend

$$\begin{aligned} \left| \int_{\mathbb{R}^d} f d\mu_n - \int_{\mathbb{R}^d} f d\mu \right| &\stackrel{\Delta\text{-Ungl.}}{\leq} \underbrace{\left| \int_{\mathbb{R}^d} f(1-f_{k_0}) d\mu_n \right|}_{\leq \|f\|_{\text{sup}}\varepsilon} + \underbrace{\left| \int_{\mathbb{R}^d} f(1-f_{k_0}) d\mu \right|}_{\leq \|f\|_{\text{sup}}\varepsilon} + \underbrace{\left| \int_{\mathbb{R}^d} f f_{k_0} d\mu_n - \int_{\mathbb{R}^d} f f_{k_0} d\mu \right|}_{\leq \varepsilon \text{ für } n \geq n_0 \text{ groß genug, da } f \cdot f_{k_0} \in \mathcal{C}_C(\mathbb{R}^d)} \\ &\leq \varepsilon(\|f\|_{\text{sup}} + 1). \end{aligned}$$

Also folgt für alle $f \in \mathcal{C}_B(\mathbb{R}^d)$ stets

$$\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu \text{ für } n \rightarrow \infty.$$

Also gilt per Definition

$$\mu_n \xrightarrow{\text{schwach}} \mu \text{ für } n \rightarrow \infty.$$

□

7.13 Satz:

Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und X_n, X \mathbb{R}^d -wertige Zufallsvariablen. Falls $X_n \xrightarrow{\text{stoch}} X$ für $n \rightarrow \infty$, so folgt

$$X_n \xrightarrow{\text{Vert.}} X \text{ für } n \rightarrow \infty.$$

Beweis:

Nach Lemma 7.12 genügt es zu zeigen, dass für alle $f \in \mathcal{C}_C(\mathbb{R}^d)$ stets

$$\int_{\mathbb{R}^d} f dP_{X_n} \rightarrow \int_{\mathbb{R}^d} f dP_X \text{ für } n \rightarrow \infty$$

gilt. Sei also $f \in \mathcal{C}_C(\mathbb{R}^d)$, dann ist f insbesondere glm. stetig und deshalb gilt

$$\forall \varepsilon > 0 \exists \delta > 0 \forall x, y \in \mathbb{R}^d : \|x - y\| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon. \quad (41)$$

Sei nun $\varepsilon > 0$ und δ gemäß (41) gewählt. Dann folgt für

$$A_n := \{\omega \in \Omega \mid \|X_n(\omega) - X(\omega)\| > \delta\} \subseteq \mathcal{A}$$

sofort

$$\begin{aligned} \left| \int_{\mathbb{R}^d} f dP_{X_n} - \int_{\mathbb{R}^d} f dP_X \right| &\stackrel{5.11}{=} \left| \int_{\Omega} f(X_n(\omega)) dP(\omega) - \int_{\Omega} f(X(\omega)) dP(\omega) \right| \\ &\stackrel{\Delta\text{-Ungl.}}{\leq} \int_{A_n} |f(X_n(\omega)) - f(X(\omega))| dP(\omega) + \int_{\Omega \setminus A_n} |f(X_n(\omega)) - f(X(\omega))| dP(\omega) \\ &\leq 2 \|f\|_{\sup} \underbrace{P(A_n)}_{\rightarrow 0} + \varepsilon. \end{aligned}$$

□

Auf \mathbb{R} kann man die schwache Konvergenz von Wahrscheinlichkeitsmaßen über die dazu assoziierten Verteilungsfunktionen charakterisieren.

7.14 Satz:

Seien $\mu, \mu_n \in \mathcal{M}^1(\mathbb{R})$ mit Verteilungsfunktionen F, F_n . Dann konvergiert μ_n genau dann schwach gegen μ , wenn F_n in allen Stetigkeitsstellen punktweise gegen F konvergiert.

Beweis:

„ \Rightarrow “: μ_n konvergiere schwach gegen μ . Sei $x \in \mathbb{R}$, sodass F in x stetig ist, dann gilt

$$\forall \varepsilon > 0 \exists \delta > 0 \forall y \in \mathbb{R} : |x - y| \leq \delta \Rightarrow |f(x) - f(y)| \leq \varepsilon. \quad (42)$$

Wähle nun $f_1, f_2 \in \mathcal{C}_C(\mathbb{R})$ mit

$$f_1(z) := \begin{cases} 1 & , z \in (-\infty, x - \delta] \\ -\frac{1}{\delta} \cdot z + \frac{x}{\delta} & , z \in [x - \delta, x] \\ 0 & , z \in [x, \infty) \end{cases} \quad \text{und} \quad f_2(z) := \begin{cases} 1 & , z \in (-\infty, x] \\ -\frac{1}{\delta} \cdot z + \frac{x}{\delta} + 1 & , z \in [x, x + \delta] \\ 0 & , z \in [x + \delta, \infty) \end{cases} .$$

Dann gilt

$$f_1(x) \leq \mathbb{1}_{(-\infty, x]} \leq f_2(x) \quad \forall x \in \mathbb{R}$$

und es folgt sofort

$$F_n(x) = \mu_n((-\infty, x]) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty, x]} d\mu_n \leq \int_{\mathbb{R}} f_2 d\mu_n$$

und

$$F_n(x) = \mu_n((-\infty, x]) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty, x]} d\mu_n \geq \int_{\mathbb{R}} f_1 d\mu_n.$$

Man erhält nun wegen der Stetigkeit von F in x

$$\limsup_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} \int_{\mathbb{R}} f_2 d\mu_n = \int_{\mathbb{R}} f_2 d\mu \leq \int_{\mathbb{R}} \mathbb{1}_{(-\infty, x+\delta]} d\mu = F(x+\delta) \leq F(x) + \varepsilon.$$

Da dies für alle $\varepsilon > 0$ gilt, folgt

$$\limsup_{n \rightarrow \infty} F_n(x) \leq F(x).$$

Andererseits gilt wegen der Stetigkeit von F in x

$$\liminf_{n \rightarrow \infty} F_n(x) \geq \liminf_{n \rightarrow \infty} \int_{\mathbb{R}} f_1 d\mu_n = \int_{\mathbb{R}} f_1 d\mu \geq \int_{\mathbb{R}} \mathbb{1}_{(-\infty, x-\delta]} d\mu = F(x-\delta) \geq F(x) - \varepsilon.$$

Da dies für alle $\varepsilon > 0$ gilt, folgt

$$\liminf_{n \rightarrow \infty} F_n(x) \geq F(x).$$

Damit folgt dann insgesamt für alle Stetigkeitsstellen x von F

$$\limsup_{n \rightarrow \infty} F_n(x) = \liminf_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} F_n(x) = F(x).$$

„ \Leftarrow “: F_n konvergiere in allen Stetigkeitsstellen punktweise gegen F . Nach Lemma 7.12 genügt es zu zeigen, dass für alle $f \in \mathcal{C}_C(\mathbb{R})$ stets

$$\int_{\mathbb{R}} f d\mu_n \rightarrow \int_{\mathbb{R}} f d\mu \quad \text{für } n \rightarrow \infty$$

gilt. Sei also $f \in \mathcal{C}_C(\mathbb{R})$ und $\varepsilon > 0$. Dann existiert ein $k \in \mathbb{N}$ mit

$$f(x) = 0 \quad \text{für } |x| \geq k.$$

Ferner ist f auf \mathbb{R} glm. stetig. Deshalb existiert für $r \in \mathbb{N}$ eine Zerlegung

$$Z_k := \{t_0 := -k < t_1 < \dots < t_r := k\}$$

mit

$$|f(x) - f(t_{i-1})| < \frac{\varepsilon}{r} \quad \forall x \in (t_{i-1}, t_i]$$

für $i = 1, \dots, r$. Die Stützstellen t_0, \dots, t_r können so gewählt werden, dass F in diesen Punkten stetig ist, da F höchstens abzählbar viele Unstetigkeitsstellen besitzt (vgl. Bemerkung 7.15). Dann gilt für $i = 1, \dots, r$

$$\begin{aligned} |\mu_n((t_{i-1}, t_i]) - \mu((t_{i-1}, t_i])| &= |F_n(t_i) - F_n(t_{i-1}) - F(t_i) + F(t_{i-1})| \\ &\stackrel{\Delta\text{-Ungl.}}{\leq} |F_n(t_i) - F(t_i)| + |F_n(t_{i-1}) - F(t_{i-1})| < \frac{\varepsilon}{r} \end{aligned}$$

für n groß genug nach Voraussetzung. Daraus folgt nun

$$\begin{aligned} \left| \int_{\mathbb{R}^d} f d\mu_n - \int_{\mathbb{R}^d} f d\mu \right| &= \left| \sum_{i=1}^r \left[\int_{t_{i-1}}^{t_i} f(t) d\mu_n(t) - \int_{t_{i-1}}^{t_i} f(t) d\mu(t) \right] \right| \\ &= \left| \sum_{i=1}^r \left[\int_{t_{i-1}}^{t_i} (f(t) - f(t_{i-1})) d\mu_n(t) - \int_{t_{i-1}}^{t_i} (f(t) - f(t_{i-1})) d\mu(t) \right] \right. \\ &\quad \left. + \sum_{i=1}^r f(t_{i-1}) \cdot [\mu_n((t_{i-1}, t_i]) - \mu((t_{i-1}, t_i])] \right| \\ &\stackrel{\Delta\text{-Ungl.}}{\leq} \sum_{i=1}^r \left[\int_{t_{i-1}}^{t_i} \underbrace{|f(t) - f(t_{i-1})|}_{< \frac{\varepsilon}{r}} d\mu_n(t) + \int_{t_{i-1}}^{t_i} \underbrace{|f(t) - f(t_{i-1})|}_{< \frac{\varepsilon}{r}} d\mu(t) \right] \\ &\quad + \sum_{i=1}^r \underbrace{|f(t_{i-1})|}_{\leq \|f\|_{\text{sup}}} \cdot \underbrace{|\mu_n((t_{i-1}, t_i]) - \mu((t_{i-1}, t_i])|}_{< \frac{\varepsilon}{r}} \\ &\leq \varepsilon(2 + \|f\|_{\text{sup}}) \text{ für } n \text{ hinreichend groß.} \end{aligned}$$

Also gilt

$$\mu_n \xrightarrow{\text{schwach}} \mu \text{ für } n \rightarrow \infty.$$

□

7.15 Bemerkung:

Sei $F : \mathbb{R} \rightarrow [0, 1]$ eine beliebige Verteilungsfunktion. Dann hat F höchstens abzählbar viele Unstetigkeitsstellen.

Beweis:

F ist unstetig in x bedeutet, da F als Verteilungsfunktion monoton wachsend ist

$$F(x_-) := \lim_{y \uparrow x} F(y) < F(x).$$

Daher gilt

$$\{x \in \mathbb{R} \mid F \text{ unstetig in } x\} = \bigcup_{k \in \mathbb{N}} \{x \in \mathbb{R} \mid F(x) - F(x_-) > \frac{1}{k}\}.$$

Offenbar gilt $|\{x \in \mathbb{R} \mid F(x) - F(x_-) > \frac{1}{k}\}| \leq k$ und es folgt die Behauptung.²⁰ \square

7.16 Beispiel:

a) Seien $x, x_n \in \mathbb{R}^d$ Punkte mit $x_n \rightarrow x$ für $n \rightarrow \infty$. Dann gilt

$$\delta_{x_n} \xrightarrow{\text{schwach}} \delta_x \text{ für } n \rightarrow \infty,$$

denn für alle $f \in \mathcal{C}_B(\mathbb{R}^d)$ folgt wegen der Stetigkeit

$$\int_{\mathbb{R}^d} f d\delta_{X_n} = f(x_n) \rightarrow f(x) = \int_{\mathbb{R}^d} f d\delta_X.$$

b) Der Grenzwertsatz 4.11 besagt

$$P_{p_n X_n} \xrightarrow{\text{schwach}} \exp_1 \text{ für } n \rightarrow \infty.$$

Es folgt ein Kriterium für die schwache Konvergenz von Verteilungen auf \mathbb{N}_0 .

7.17 Lemma:

Seien $\mu, \mu_n \in \mathcal{M}^1(\mathbb{N}_0) \subseteq \mathcal{M}^1(\mathbb{R})$ mit zugehörigen Zähldichten $f, f_n : \mathbb{N}_0 \rightarrow [0, 1]$. Dann gilt:

μ_n konvergiert genau dann schwach gegen μ , wenn $f_n(k)$ für alle $k \in \mathbb{N}_0$ punktweise gegen $f(k)$ konvergiert.

Beweis: „ \Rightarrow “: Setze für $k \in \mathbb{N}$

$$g(x) := \begin{cases} 0 & , x < k-1 \\ x+1-k & , k-1 \leq x \leq k \\ -x+k+1 & , k \leq x \leq k+1 \\ 0 & , x > k+1 \end{cases} \in \mathcal{C}_C(\mathbb{R}).$$

Dann gilt offenbar

$$f_n(k) = \sum_{l \in \mathbb{N}_0} g(l) f_n(l) = \int_{\mathbb{R}} g d\mu_n \rightarrow \int_{\mathbb{R}} g d\mu = \sum_{l \in \mathbb{N}_0} g(l) f(l) = f(k).$$

²⁰ Ende der zwanzigsten Vorlesung vom Mi, 18.06.14

Also konvergiert $f_n(k)$ für alle $k \in \mathbb{N}_0$ punktweise gegen $f(k)$.

„ \Leftarrow “: Nach Lemma 7.12 kann o.E. $f \in \mathcal{C}_C(\mathbb{R})$ gewählt werden. Dann ist $g(k) \neq 0$ nur für endlich viele $k \in \mathbb{N}_0$ und man erhält

$$\int_{\mathbb{R}} g d\mu_n = \sum_{k \in \mathbb{N}_0} g(k) f_n(k) \rightarrow \sum_{k \in \mathbb{N}_0} g(k) f(k) = \int_{\mathbb{R}} g d\mu.$$

Also konvergiert μ_n schwach gegen μ . □

Beispiel: Nach Lemma 7.12 und der Poissonapproximation der Binomialverteilung 6.14 gilt

$$B_{n,p_n} \xrightarrow{\text{schwach}} \pi_\lambda \text{ für } n \rightarrow \infty,$$

falls $n \cdot p \rightarrow \lambda > 0$ für $n \rightarrow \infty$.

Nun wird ein wesentlicher Satz über schwache Konvergenz ohne Beweis angegeben.

7.18 Theorem: (Zentraler Grenzwertsatz)

Sei $(X_n) \subseteq \mathcal{L}_2(\Omega, \mathcal{A}, P)$ eine Folge reellwertiger unabhängiger identisch verteilter Zufallsvariablen mit $m := E(X_n)$ und $\sigma^2 := \text{Var}(X_n) > 0$ für alle $n \in \mathbb{N}$. Dann definiert man die Zufallsvariablen

$$S_n^* := \frac{X_1 + \dots + X_n - nm}{\sqrt{n\sigma^2}}$$

mit $E(S_n^*) = 0$ und $\text{Var}(S_n^*) = 1$. Für die Verteilung von S_n^* gilt dann

$$P_{S_n^*} \xrightarrow{\text{schwach}} N_{0,1} \text{ für } n \rightarrow \infty. \quad (43)$$

Beweis:

Der Beweis erfolgt erst in der Stochastik II Vorlesung. □

Bemerkungen:

a) Nach Satz 7.14 gilt für $x \in \mathbb{R}$ stets

$$F_n(x) = P(S_n^* \leq x) \rightarrow \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \text{ für } n \rightarrow \infty.$$

b) Fehlerabschätzung von *Berry-Esseen*:

$$\|F_n - \Phi\|_{\text{sup}} \leq \frac{1}{\sqrt{n}} \cdot \underbrace{\frac{2 \cdot E(|X_1 - m|^3)}{\sigma^3}}_{\text{konstant}} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \text{ für } n \rightarrow \infty.$$

Da die X_n identisch verteilt sind, ist es in der Konstante egal, welche Zufallsvariable in dem Erwartungswert steht. Hier wurde o.E. die Zufallsvariable X_1 verwendet.

7.19 Beispiel: (Anwendung auf Binomialverteilung)

Seien $p \in (0, 1)$ und X_n unabhängige $B_{1,p}$ -verteilte Zufallsvariablen. Dann gilt nach 5.13 c) $m = E(X_n) = p$ und nach 5.20 a) $\sigma^2 = \text{Var}(X_n) = p(1-p)$ für alle $n \in \mathbb{N}$. Sei ferner $a < b \in \mathbb{R}$, dann liefert Theorem 7.18 für

$$S_n^* := \frac{X_1 + \dots + X_n - np}{\sqrt{np(1-p)}}$$

sofort

$$\begin{aligned} N_{0,1}([a, b]) &= \Phi(b) - \Phi(a) \stackrel{7.18}{=} \lim_{n \rightarrow \infty} [P(S_n^* \leq b) - P(S_n^* \leq a)] = \lim_{n \rightarrow \infty} P(S_n^* \in (a, b]) \\ &= \lim_{n \rightarrow \infty} P\left(\frac{X_1 + \dots + X_n - np}{\sqrt{np(1-p)}} \in (a, b]\right) \\ &= \lim_{n \rightarrow \infty} P\left(X_1 + \dots + X_n \in (np + a\sqrt{np(1-p)}, np + b\sqrt{np(1-p)})\right). \end{aligned}$$

Daraus folgt der

7.20 Grenzwertsatz: (v. Moivre-Laplace)

Für alle $a < b \in \mathbb{R}$ gilt

$$\Phi(b) - \Phi(a) = \lim_{n \rightarrow \infty} P\left(X_1 + \dots + X_n \in (np + a\sqrt{np(1-p)}, np + b\sqrt{np(1-p)})\right).$$

7.21 Beispiele:

a) Man würfle mit einem fairen Würfeln $n = 6000$ mal. Wie groß ist dann die Wahrscheinlichkeit

$$A := P(\text{„Die Zahl 6 wird zwischen 950 und 1050 mal geworfen“})?$$

1. Lösung: (exakt) Nach den Regeln der Binomialverteilung folgt für $p = \frac{1}{6}$ und $n = 6000$ sofort der auch für Computerprogramme sehr schwer berechenbare Ausdruck

$$A = \sum_{k=950}^{1050} \binom{6000}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{6000-k} = ???$$

2. Lösung: (approximativ) Mit dem Satz von Moivre-Laplace 7.20 erhält man

$$A \approx B_{6000, \frac{1}{6}}((950, 1050]) \stackrel{!}{\approx} B_{6000, \frac{1}{6}}\left(\left(1000 + a\sqrt{\frac{5000}{6}}, 1000 + b\sqrt{\frac{5000}{6}}\right)\right) \stackrel{7.20}{\approx} \Phi(b) - \Phi(a).$$

Die Werte für $\Phi(b)$ und $\Phi(a)$ sind tabelliert! Man kann nun a und b berechnen und erhält dafür

$$a = -\frac{50}{\sqrt{\frac{5000}{6}}} = -\sqrt{3} \approx -1,73 \text{ und } b = \frac{50}{\sqrt{\frac{5000}{6}}} = \sqrt{3} \approx 1,73.$$

Damit ergibt sich nun

$$A \approx N_{0,1}([a, b]) = N_{0,1}([-b, b]) = \Phi(b) - \Phi(-b) = 2\Phi(b) - 1 \stackrel{\text{Tabelle}}{\approx} 0,91.$$

b) Wahlvorhersage: Gesucht ist der Stichprobenumfang $n \in \mathbb{N}$ einer Umfrage, um den Stimmanteil $p \in (0, 1)$ einer bestimmten Partei bei einer Wahl zu schätzen, wobei das Gütekriterium

$$P(|\text{„Fehler“}| \geq 0,01) \leq 0,05 \quad [\text{vgl. 5.30}]$$

erfüllt sein soll. Man definiert zur Modellierung die Zufallsvariablen $X_k \in \{0, 1\}$ als Abstimmergebnis der k -ten befragten Person. Dabei bedeutet 0, dass die Person die Partei nicht wählt und 1, dass sie diese Partei wählt. Die X_k können als unabhängig und $B_{1,p}$ verteilt angenommen werden. Dann ist die Zufallsvariable

$$S_n := X_1 + \dots + X_n$$

$B_{n,p}$ verteilt. Durch $\frac{1}{n}S_n$ ist dann eine sinnvolle Schätzung für p gegeben und man erhält

$$\begin{aligned} P\left(\frac{S_n}{n} - p \in [-0,01, 0,01]\right) &= P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \in \left[-\frac{0,01\sqrt{n}}{\sqrt{np(1-p)}}, \frac{0,01\sqrt{n}}{\sqrt{np(1-p)}}\right]\right) \\ &\stackrel{7.18}{\approx} N_{0,1}\left(\left[-\frac{0,01\sqrt{n}}{\sqrt{np(1-p)}}, \frac{0,01\sqrt{n}}{\sqrt{np(1-p)}}\right]\right) \\ &\geq N_{0,1}\left(\left[-\frac{\sqrt{n}}{50}, \frac{\sqrt{n}}{50}\right]\right) = 2\Phi\left(\frac{\sqrt{n}}{50}\right) - 1 \stackrel{\text{Qualitäts-}}{\geq} \text{bedingung} 0,95. \end{aligned}$$

Umformen und Nachschlagen in einer Tabelle für Φ liefert $n \approx 9600$.

Abschließend kann man also festhalten, dass das Gütekriterium für die Umfrage ab ca. 10000 Befragten erfüllt ist.

8 Markov-Ketten und stochastische Matrizen

Motivation: Die Modellierung von zufälligen Phänomenen mit zeitlicher Entwicklung (z.B. Wetter, Aktienkurse, ...) erfolgt oft mit Hilfe so genannter stochastischer Prozesse.

8.1 Definition: (stochastische Prozesse)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, (E, \mathcal{B}) ein Messraum und $\phi \neq I \subseteq \mathbb{R}$ ein Intervall. Eine Familie $(X_t)_{t \in I}$ von Zufallsvariablen $X_t : \Omega \rightarrow E$ heißt ein *stochastischer Prozess* mit *Zeitbereich* I , *Zustandsraum* (E, \mathcal{B}) und *Grundwahrscheinlichkeitsraum* (Ω, \mathcal{A}, P) .

Bemerkungen: (Wichtige Spezialfälle)

a) Häufig hat man einen Zustandsraum

$$(E, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \text{ oder } (E, \mathcal{B}) \text{ diskret, d.h. } E \text{ höchstens abzählbar und } \mathcal{B} = \mathcal{P}(E).$$

b) Häufig hat man einen Zeitbereich

$$I = \mathbb{N}_0, \quad I = \{0, 1, \dots, n \in \mathbb{N}\}, \quad I = [0, \infty) \text{ oder } I = [0, T \in \mathbb{R}].$$

In der Stochastik I beschränkt man sich auf $I = \mathbb{N}_0$ und einen diskreten Zustandsraum $E = \{a_1, a_2, \dots\}$.

Rekursiv definierte Verteilung der X_t :

a) Man hat die Startverteilung $P_{X_0} \in \mathcal{M}^1(E)$ mit konkreter Zähldichte f_0 .

b) Man definiert Übergangswahrscheinlichkeiten von Zeiten $\{0, 1, \dots, t\}$ zum Zeitpunkt $\{t+1\}$ für $t \in \mathbb{N}_0$ mit Hilfe einer so genannten *bedingten Zähldichte*

$$f_{t+1}(b|b_0, \dots, b_t) = P_{(b_0, \dots, b_t), b} := P(X_{t+1} = b | X_0 = b_0, \dots, X_t = b_t) \in [0, 1].$$

Man stellt die üblichen Forderungen an die bedingte Zähldichte

$$\forall b \in E : f_{t+1}(b|b_0, \dots, b_t) \geq 0 \text{ und } \sum_{b \in E} f_{t+1}(b|b_0, \dots, b_t) = 1.$$

Idee: Durch Startverteilung und Folge $(f_t)_{t \in \mathbb{N}_0}$ der bedingten Zähldichten wird der gesuchte stochastische Prozess beschrieben. ²¹

²¹ Ende der einundzwanzigsten Vorlesung vom Fr, 20.06.14

Beispiele:

a) Es seien $I := \mathbb{N}_0$, $E := \mathbb{Z}$ und (X_n) eine Folge unabhängiger identisch verteilter Zufallsvariablen mit Verteilung

$$\mu := \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1.$$

Dann gilt für alle $t \in \mathbb{N}_0$ stets $P_{X_t} = \mu$ und insbesondere hat man die Startverteilung $P_{X_0} = \mu$. Die Übergangswahrscheinlichkeiten sind

$$P_{(b_0, \dots, b_t), b} = \begin{cases} \frac{1}{2} & , b \in \{-1, 1\} \\ 0 & , \text{sonst} \end{cases}.$$

b) Es seien $I := \mathbb{N}_0$, $E := \mathbb{Z}$ und (X_n) eine Folge unabhängiger identisch verteilter Zufallsvariablen mit Verteilung

$$\mu := \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1.$$

Definiere nun den neuen Prozess $(S_t)_{t \in \mathbb{N}_0}$ auf $E = \mathbb{Z}$ durch

$$S_0 := 0 \text{ und } S_n := \sum_{k=1}^n X_k \text{ für } n \in \mathbb{N}.$$

Dann hat man die Startverteilung $P_{S_0} = \delta_0$ und für alle $t \in \mathbb{N}_0$ gilt für die Übergangswahrscheinlichkeiten stets

$$P_{(b_0, \dots, b_t), b} = \begin{cases} \frac{1}{2} & , b - b_t \in \{-1, 1\} \\ 0 & , \text{sonst} \end{cases}.$$

Diese Prozesse bezeichnet man auch als „Irrfahrten“ auf \mathbb{Z} .

c) Man betrachte das Polya'sche Urnenmodell mit weißen und schwarzen Kugeln. Definiere die Zufallsvariable

$$X_t := \text{Farbe der Kugel bei } (t+1)\text{-ter Ziehung}$$

und den Zustandsraum

$$E := \{\text{schwarz, weiß}\}.$$

Hier hängt $P_{(b_0, \dots, b_t), b}$ von allen b_0, \dots, b_t ab [vgl. 2.15].

d) *Ruinsspiele:* Zwei Spieler starten ein Spiel mit Startkapital $N_1, N_2 \in \mathbb{N}$. In jeder Runde gewinnt/verliert Spieler 1 von /an Spieler 2 eine feste Menge seines Kapitals mit der festen Wahrscheinlichkeit $p \in (0, 1)$. Das Spiel endet ohne weitere Kapitalabänderungen, wenn einer der beiden Spieler pleite ist. Modelliere nun den Kapitalverlauf von Spieler 1. Dazu definiere die Zufallsvariable

$$X_t := \text{Kapital von Spieler 1 nach } t \text{ Runden.}$$

Der Zustandsraum ist dann

$$E := \{0, 1, \dots, N_1 + N_2\}$$

und man hat offenbar $X_0 = N_1$ und deshalb die Startverteilung $P_{X_0} = \delta_{N_1}$. Die Übergangswahrscheinlichkeiten für $b_0, \dots, b_t \in E$ sind

$$P_{(b_0, \dots, b_t), b} = \begin{cases} p & , \text{ falls } b_t \neq 0, b_t \neq N_1 + N_2 \text{ und } b = b_t + 1 \\ 1 - p & , \text{ falls } b_t \neq 0, b_t \neq N_1 + N_2 \text{ und } b = b_t - 1 \\ 1 & , \text{ falls } b = b_t = 0 \text{ oder } b = b_t = N_1 + N_2 \\ 0 & , \text{ sonst} \end{cases}.$$

(Es ist oft sinnvoll solche Prozesse durch Übergangsgraphen zu veranschaulichen, besonders, wenn viele Übergangswahrscheinlichkeiten 0 sind.)

8.2 Definition: (Markov-Ketten)

Ein stochastischer Prozess $(X_t)_{t \in \mathbb{N}_0}$ mit diskreten Zuständen in E heißt *Markov-Kette*, falls die so genannte *Markov-Bedingung*

$$(M) \quad \forall n \in \mathbb{N} \forall b_0, \dots, b_{n+1} \in E \text{ mit } P(X_0 = b_0, \dots, X_n = b_n) > 0 : \\ P(X_{n+1} = b_{n+1} | X_0 = b_0, \dots, X_n = b_n) = P(X_{n+1} = b_{n+1} | X_n = b_n) =: P_{b_n, b_{n+1}}^{(n+1)}$$

erfüllt ist.

Deutung: Bei Kenntnis der Gegenwart zur Zeit n ist die Zukunft $n + 1$ von der Vergangenheit $0, 1, \dots, n - 1$ unabhängig.

8.3 Rechenregeln:

Für eine Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ gilt:

a) Für alle $n \in \mathbb{N}$ und alle $b_0, \dots, b_n \in E$ mit $P(X_k = b_k) > 0$ für $k = 1, \dots, n - 1$ gilt

$$P(X_0 = b_0, \dots, X_n = b_n) = P(X_0 = b_0) \cdot \prod_{k=1}^n P(X_k = b_k | X_{k-1} = b_{k-1}).$$

Dies folgt aus 2.3 a) und der Markov-Bedingung.

b) Für alle $n \in \mathbb{N}_0, m \in \mathbb{N}$ und alle $b_n, \dots, b_{n+m} \in E$ mit $P(X_k = b_k) > 0$ für $k = n, \dots, n+m-1$ gilt

$$P(X_{n+1} = b_{n+1}, \dots, X_{n+m} = b_{n+m} | X_n = b_n) = \prod_{k=n+1}^{n+m} P(X_k = b_k | X_{k-1} = b_{k-1}).$$

Dies folgt ebenfalls aus 2.3 a) und der Markov-Bedingung.

c) *Chapman-Kolmogorov Gleichungen:*

Für alle $0 \leq n < m < k \in \mathbb{N}_0$ und $a, b \in E$ mit $P(X_n = a) > 0$ gilt

$$P(X_k = b | X_n = a) = \sum_{c \in E} P(X_m = c | X_n = a) \cdot P(X_k = b | X_m = c), \text{ denn}$$

$$\begin{aligned}
 P(X_k = b, X_n = a) &= \sum_{c \in E} P(X_k = b, X_m = c, X_n = a) = \sum_{c \in E} P(X_k = b | X_m = c, X_n = a) \cdot P(X_m = c, X_n = a) \\
 &\stackrel{(M)}{=} \sum_{c \in E} P(X_k = b | X_m = c) \cdot P(X_m = c, X_n = a)
 \end{aligned}$$

und Division durch $P(X_n = a)$ liefert die Behauptung.

8.4 Übergangsmatrizen:

Zur Beschreibung einer Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ auf $E := \{a_1, a_2, \dots\}$ geht man wie folgt vor:

a) Die zur Startverteilung $P_{X_0} \in \mathcal{M}^1(E)$ gehörende Zähldichte wird als Zeilenvektor

$$(p_1, p_2, \dots) \text{ mit } p_j := P(X_0 = a_j) \quad \forall j \in \mathbb{N}$$

dargestellt. Dabei gilt

$$p_j \in [0, 1] \forall j \in \mathbb{N} \text{ und } \sum_{j \geq 1} p_j = 1.$$

b) Für $i = 1, 2, \dots$ und $j = 1, 2, \dots$ definiert man

$$p_{i,j}^{n,m} := P(X_m = a_j | X_n = a_i)$$

für alle $0 \leq n < m \in \mathbb{N}$ und $a_i, a_j \in E$. Daraus bildet man nun die so genannte *Übergangsmatrix*

$$S^{n,m} := (p_{i,j}^{n,m})_{i,j=1,2,\dots}$$

Eigenschaften:

Für eine stochastische Matrix $S^{n,m} := (p_{i,j}^{n,m})_{i,j=1,2,\dots}$ hat man:

a) Offenbar gilt für alle $0 \leq n < m$ und $i, j = 1, 2, \dots$ stets

$$p_{i,j}^{n,m} \in [0, 1]$$

und die *Zeilensumme* der stochastischen Matrix ist stets 1, also gilt für alle $i = 1, 2, \dots$ stets

$$\sum_{j \geq 1} p_{i,j}^{n,m} = 1.$$

b) Man hat die *Chapman-Kolmogorov Gleichungen*:

Für alle $0 \leq n < m < k \in \mathbb{N}_0$ und $i, j = 1, 2, \dots$ gilt wegen 8.3 c)

$$\sum_{l \geq 1} p_{i,l}^{n,m} p_{l,j}^{m,k} = p_{i,j}^{n,k} \Leftrightarrow S^{n,m} \cdot S^{m,k} = S^{n,k}.$$

c) Für alle $n \in \mathbb{N}$, $j = 1, 2, \dots$ und $a_j \in E$ gilt

$$P(X_n = a_j) = \sum_{i \geq 1} P(X_n = a_j, X_0 = a_i) = \sum_{i \geq 1} \underbrace{P(X_0 = a_i)}_{=p_i} \cdot \underbrace{P(X_n = a_j, X_0 = a_i)}_{=p_{i,j}^{0,n}} = \underbrace{\left((p_1, p_2, \dots) \cdot S^{0,n} \right)}_{\substack{j\text{-te Komponente} \\ \text{dieses Zeilenvektors}}}_j.$$

Konvention: Verteilungen auf E entsprechen Zeilenvektoren. Dann gilt

$$\text{Startvektor} \times \text{Übergangsmatrix} = \text{Verteilung des Prozesses.}$$

Es folgt eine Untersuchung von Matrizen für endliche Zustandsräume E .

8.5 Definition: (stochastische Matrix)

Eine Matrix $S := (p_{ij})_{i,j=1,\dots,n} \in \mathcal{M}_n(\mathbb{R})$ heißt *stochastisch*, falls

$$\forall i, j = 1, \dots, n : p_{ij} \in [0, 1] \text{ und } \forall i = 1, \dots, n : \sum_{j=1}^n p_{ij} = 1.$$

8.6 Fakten:

a) Seien $S^{(1)}, S^{(2)} \in \mathcal{M}_n(\mathbb{R})$ stochastisch. Dann ist offensichtlich auch $S^{(1)} \cdot S^{(2)} \in \mathcal{M}_n(\mathbb{R})$ stochastisch.

b) Sei $(S^{(k)})_{k \in \mathbb{N}} \subseteq \mathcal{M}_n(\mathbb{R})$ mit $S^{(k)} = (p_{ij}^{(k)})_{i,j=1,\dots,n}$ eine Folge stochastischer Matrizen, die gegen eine Matrix $S = (p_{ij})_{i,j=1,\dots,n} \in \mathcal{M}_n(\mathbb{R})$ konvergiert (d.h. komponentenweise konvergiert). Dann ist auch S wieder stochastisch, denn offenbar sind alle $p_{ij} \in [0, 1]$ und es gilt

$$\sum_{j=1}^n p_{ij} = \sum_{j=1}^n \lim_{k \rightarrow \infty} p_{ij}^{(k)} = \lim_{k \rightarrow \infty} \sum_{j=1}^n p_{ij}^{(k)} = \lim_{k \rightarrow \infty} 1 = 1.$$

c) Sei $S \in \mathcal{M}_n(\mathbb{R})$ stochastisch. Dann ist der Vektor

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

ein Eigenvektor von S zum Eigenwert 1, denn

$$S \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^n p_{1j} \\ \vdots \\ \sum_{j=1}^n p_{nj} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

8.7 Lemma:

Sei $\lambda \in \mathbb{C}$ Eigenwert einer stochastischen Matrix $S \in \mathcal{M}_n(\mathbb{R})$, dann gilt:

a) $|\lambda| \leq 1$

b) Falls $|\lambda| = 1$, so stimmen geometrische und algebraische Vielfachheit von λ überein.

Beweis:

Sei $S \in \mathcal{M}_n(\mathbb{R})$ eine stochastische Matrix. Also ist nach 8.6 a) auch $S^k \in \mathcal{M}_n(\mathbb{R})$ für alle $k \in \mathbb{N}$ stochastisch und es gilt

$$S^k \in [0, 1]^{n \times n} \subseteq \mathbb{R}^{n \times n}$$

für alle $k \in \mathbb{N}$. Also ist $(S^k) \subseteq [0, 1]^{n \times n}$ eine beschränkte Folge. Für $1 \leq r \leq n$ sei

$$J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_r \end{pmatrix} = T^{-1}ST, \quad T \in \text{GL}_n(\mathbb{R})$$

die *Jordansche Normalform* von S mit den Jordanblöcken

$$J_l := \begin{pmatrix} \lambda_l & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 0 & \cdots & \cdots & \lambda_l \end{pmatrix}$$

für $l = 1, \dots, r$. Für S^k hat man dann

$$S^k = (TJT^{-1})^k = \underbrace{TJT^{-1}TJT^{-1}\dots TJT^{-1}}_{k \text{ Faktoren}} = TJ^kT^{-1} = T \begin{pmatrix} J_1^k & & 0 \\ & \ddots & \\ 0 & & J_r^k \end{pmatrix} T^{-1}.$$

Dabei ist

$$J_l^k := \begin{pmatrix} \lambda_l^k & k\lambda_l^{k-1} & & \cdots & \\ \vdots & \ddots & \ddots & & \\ \vdots & & \ddots & k\lambda_l^{k-1} & \\ 0 & \cdots & \cdots & \lambda_l^k & \end{pmatrix}$$

für $l = 1, \dots, r$.

Da (S^k) beschränkt ist, muss wegen $S^k = TJ^kT^{-1}$ auch J^k beschränkt. Aus der expliziten Form der Jordanblöcke von J^k kann man also ablesen, dass für alle $l = 1, \dots, r$ stets $|\lambda_l| \leq 1$ gelten muss, denn sonst wäre J^k unbeschränkt. Insbesondere folgt deshalb auch für $|\lambda_l| = 1$, dass die Ausdrücke in der Nebendiagonalen unbeschränkt sind. Um keinen Widerspruch zur Beschränktheit von S^k zu erhalten, darf also der zugehörige Jordanblock nur den Eigenwert λ_l enthalten, sodass es gar keine Nebendiagonale gibt. Dies ist genau dann der Fall, wenn die algebraische mit der geometrischen Vielfachheit von λ_l übereinstimmt. ²² \square

²² Ende der zweiundzwanzigsten Vorlesung vom Mi, 25.06.14

8.8 Satz: (Ergodensatz)

Es sei $S \in \mathcal{M}_N(\mathbb{R})$ stochastisch, dann ist auch

$$Q := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n S^k \in \mathcal{M}_N(\mathbb{R})$$

stochastisch mit

$$Q^2 = Q = QS = SQ.$$

Beweis:

Sei $S \in \mathcal{M}_n(\mathbb{R})$ eine stochastische Matrix. Also ist nach 8.6 a) auch $S^k \in \mathcal{M}_n(\mathbb{R})$ für alle $k \in \mathbb{N}$ stochastisch. Für $1 \leq r \leq N$ sei

$$J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_r \end{pmatrix} = T^{-1}ST, \quad T \in \text{GL}_N(\mathbb{R})$$

die *Jordansche Normalform* von S , wobei nach Lemma 8.7 für $l = 1, \dots, r$ nur Jordanblöcke der Form

- (a) $\lambda_l = 1 \Rightarrow J_l = (1)$
- (b) $|\lambda_l| = 1, \lambda_l \neq 1 \Rightarrow J_l = (\lambda_l)$
- (c) $|\lambda_l| < 1 \Rightarrow J_l := \begin{pmatrix} \lambda_l & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 0 & \cdots & \cdots & \lambda_l \end{pmatrix}$

auftreten können. Ferner gilt für alle $k \in \mathbb{N}$ stets

$$\frac{1}{n} \sum_{k=1}^n S^k = \frac{1}{n} \sum_{k=1}^n T J^k T^{-1} = T \begin{pmatrix} \tilde{J}_1 & & 0 \\ & \ddots & \\ 0 & & \tilde{J}_r \end{pmatrix} T^{-1}$$

mit

$$\tilde{J}_l = \frac{1}{n} \sum_{k=1}^n J_l^k.$$

Mit der Fallunterscheidung von oben folgt für $l = 1, \dots, r$ sofort

(a) $\lambda_l = 1 \Rightarrow \tilde{J}_l = \frac{1}{n} \sum_{k=1}^n (1)^k = (1) \rightarrow (1)$ für $n \rightarrow \infty$.

(b) $|\lambda_l| = 1, \lambda_l \neq 1 \Rightarrow \tilde{J}_l = \frac{1}{n} \sum_{k=1}^n (\lambda_l)^k = \frac{1}{n} \cdot \lambda_l \sum_{k=0}^{n-1} (\lambda_l)^k \stackrel{\text{geom. Summe}}{=} \left(\frac{1}{n} \cdot \lambda_l \cdot \underbrace{\frac{1 - \lambda_l^n}{1 - \lambda_l}}_{\text{beschränkt}} \right) \rightarrow (0)$ für $n \rightarrow \infty$.

(c) $|\lambda_l| < 1 \Rightarrow J_l = \begin{pmatrix} \lambda_l & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 0 & \dots & \dots & \lambda_l \end{pmatrix} \Rightarrow J_l^k = \begin{pmatrix} \lambda_l^k & \binom{k}{1} \lambda_l^{k-1} & \dots & \binom{k}{d-1} \lambda_l^{k-d+1} \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \binom{k}{1} \lambda_l^{k-1} \\ 0 & \dots & 0 & \lambda_l^k \end{pmatrix} \rightarrow \begin{pmatrix} 0 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & 0 \end{pmatrix}$

für $k \rightarrow \infty$ mit $d := \dim J_l \Rightarrow \tilde{J}_l = \frac{1}{n} \sum_{k=1}^n J_l^k \rightarrow 0$ für $k \rightarrow \infty \forall l = 1, \dots, r$.

Insgesamt hat man also

$$\frac{1}{n} \sum_{k=1}^n S^k \rightarrow T \cdot \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \cdot T^{-1} = Q.$$

Die Q Eigenschaften von Q sind:

- (a) Q ist stochastisch, denn $\frac{1}{n} \sum_{k=1}^n S^k$ ist für jedes $n \in \mathbb{N}$ stochastisch und nach 8.6 b) ist dann auch der Limes Q stochastisch.
- (b) Durch Nachrechnen sieht man sofort

$$Q^2 = T \cdot \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \cdot T^{-1} T \cdot \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \cdot T^{-1} = T \cdot \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix} \cdot T^{-1} = Q.$$

- (c) Wegen $\frac{1}{n} S \rightarrow 0$ und $\frac{1}{n} S^{n+1} \rightarrow 0$ für $n \rightarrow \infty$ folgt

$$QS = \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n S^k \right) \cdot S = \lim_{n \rightarrow \infty} (S^2 + \dots + S^{n+1}) = \lim_{n \rightarrow \infty} (S + \dots + S^n) = Q.$$

- (d) Analog zu (c) zeigt man auch $SQ = Q$. □

Bemerkungen:

- a) Insbesondere sind alle Zeilen von Q Zeileneigenvektoren von S zum Eigenwert 1.
 b) Eine Folge $(S^k)_{k \in \mathbb{N}}$ ist im Allgemeinen nicht konvergent, denn betrachte

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Rightarrow S^{2k} = I, \quad S^{2k+1} = S.$$

- c) Die Matrix Q ist nie die Nullmatrix und $\text{rang} Q$ entspricht nach 8.6 c) immer der algebraischen (geometrischen) Vielfachheit des Eigenwerts 1 von S .

8.9 Definition: (Stationäre Verteilungen)

Es sei $S \in \mathcal{M}_N(\mathbb{R})$ eine stochastische Matrix. Ein Zeilenvektor $v := (p_1, \dots, p_N) \in [0, 1]^N$ mit $\sum_{k=1}^N p_k = 1$ heißt *stationäre Verteilung* zu S , falls

$$v \cdot S = v.$$

Deutung: Bei einer Markov-Kette mit Einschrittübergangswahrscheinlichkeit S und Startverteilung v hat nun nach einem Schritt und damit auch zu jedem späteren Zeitpunkt stets die Verteilung v als Verteilung der Markov-Kette zu diesen Zeiten.

Der Ergodensatz 8.8 liefert sofort das

8.10 Korollar:

Jede stochastische Matrix $S \in \mathcal{M}_N(\mathbb{R})$ besitzt mindestens eine stationäre Verteilung.

Beweis:

Nehme eine Zeile v von Q aus 8.8. □

8.11 Zeithomogene Markov-Ketten:

- a) Betrachte eine Markov-Kette $(X_n)_{n \in \mathbb{N}}$ auf $E := \{a_1, a_2, \dots\}$ höchstens abzählbar unendlich. Die Startverteilung P_{X_0} entspricht dann einem Zeilenvektor. Man hat also die Übergangsmatrizen $S^{n,m}$ mit $0 \leq n < m$ wie in 8.4.
 b) Definition: Falls $S := S^{n,m}$ unabhängig von $n \in \mathbb{N}_0$ ist, so heißt $(X_n)_{n \in \mathbb{N}}$ eine *zeithomogene Markov-Kette* auf E mit Startverteilung P_{X_0} und Einschrittübergangsmatrix.
 c) In dieser Situation gilt dann für alle $n < m$ stets

$$S^{n,m} = S^{n,n+1} \cdot S^{n+1,n+2} \cdot \dots \cdot S^{m-1,m} = S^{m-n}.$$

d) Beispiel: (Ruinspiel) Zwei Spieler starten ein Spiel mit Startkapital $N_1, N_2 \in \mathbb{N}$. In jeder Runde gewinnt/verliert Spieler 1 von /an Spieler 2 eine feste Menge seines Kapitals mit der festen Wahrscheinlichkeit $p \in (0, 1)$. Hier ist

$$S = \begin{pmatrix} 1 & 0 & 0 & \dots & \dots & \dots & 0 \\ 1-p & 0 & p & & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & 1-p & 0 & p \\ 0 & \dots & \dots & \dots & 0 & 0 & 1 \end{pmatrix} \in \mathcal{M}_{N_1+N_2}(\mathbb{R}).$$

8.12 Auftreffwahrscheinlichkeiten:

Es sei E ein endlicher Zustandsraum, $\phi \neq A \subseteq E$ und $(X_n)_{n \in \mathbb{N}}$ eine zeithomogene Markov-Kette auf E mit Übergangsmatrix S . Betrachte

$$T_A(\omega) := \inf\{n \in \mathbb{N}_0 \mid X_n(\omega) \in A\} := \begin{cases} \min\{n \in \mathbb{N}_0 \mid X_n(\omega) \in A\} & , \text{ falls } \{n \in \mathbb{N}_0 \mid X_n(\omega) \in A\} \neq \emptyset \\ \infty & , \text{ falls } \{n \in \mathbb{N}_0 \mid X_n(\omega) \in A\} = \emptyset \end{cases}$$

T_A ist eine $\mathbb{N}_0 \cup \{\infty\}$ -wertige Zufallsvariable.

Deutung: T_A ist die Zeit des ersten Auftreffens in der Menge A .

Eine häufig auftretende Frage ist, wie groß die Wahrscheinlichkeit

$$\alpha_r^A := P(T_A < \infty \mid X_0 = r)$$

bei Startpunkt r irgendwann die Menge A zu treffen, ist.

8.13 Satz:

Für alle $\phi \neq A \subseteq E$ und $r \in E$ gilt

$$\alpha_r^A = \begin{cases} 1 & , \text{ falls } r \in A \\ \sum_{b \in E} p_{rb} \alpha_b^A & , \text{ falls } r \in E \setminus A \end{cases}$$

Dies führt im Allgemeinen auf ein inhomogenes LGS. Dabei ist $S = (p_{rb})_{r,b \in E}$.

Beweis: Der erste Fall für $r \in A$ ist klar.

Sei also $r \in E \setminus A$ und $n \in \mathbb{N}$, dann gilt

$$\begin{aligned}
P(T_A \leq n+1 | X_0 = r) &= P(\exists k \in \{1, \dots, n+1\} \text{ mit } x_k \in A | X_0 = r) \\
&= \sum_{\substack{(b_1, \dots, b_{n+1}) \in E^{n+1} \text{ mit} \\ \exists i \in \{1, \dots, n+1\}: b_i \in A}} P(X_1 = b_1, \dots, X_{n+1} = b_{n+1} | X_0 = r) \\
&= \sum_{\substack{(b_1, \dots, b_{n+1}) \in E^{n+1} \text{ mit} \\ \exists i \in \{1, \dots, n+1\}: b_i \in A}} p_{rb_1} p_{b_1 b_2} \cdots p_{b_n b_{n+1}} \\
&\stackrel{c:=b_1}{=} \sum_{c \in E \setminus A} \sum_{\substack{(b_2, \dots, b_{n+1}) \in E^n \text{ mit} \\ \exists i \in \{2, \dots, n+1\}: b_i \in A}} p_{rc} p_{cb_2} \cdots p_{b_n b_{n+1}} + \sum_{c \in A} \sum_{(b_2, \dots, b_{n+1}) \in E^n} p_{rc} p_{cb_2} \cdots p_{b_n b_{n+1}} \\
&= \sum_{c \in E \setminus A} p_{rc} \sum_{\substack{(b_1, \dots, b_n) \in E^n \text{ mit} \\ \exists i \in \{1, \dots, n\}: b_i \in A}} p_{cb_2} \cdots p_{b_n b_{n+1}} + \sum_{c \in E} p_{rc} \\
&= \sum_{c \in E \setminus A} p_{rc} P(T_A \leq n | X_0 = c) + \sum_{c \in A} p_{rc} \underbrace{P(T_A \leq n | X_0 = c)}_{=1, \text{ da } c \in A} \\
&= \sum_{c \in E} p_{rc} P(T_A \leq n | X_0 = c).
\end{aligned}$$

Also folgt

$$\begin{aligned}
P(T_A < \infty | X_0 = r) &= P\left(\bigcup_{n \in \mathbb{N}} \underbrace{\{\omega \in \Omega | T_A(\omega) \leq n\}}_{:= B_n \subseteq B_{n+1}} \mid X_0 = r\right) = \frac{P\left(\bigcup_{n \in \mathbb{N}} \{\omega \in \Omega | T_A(\omega) \leq n\} \cap \{X_0 = r\}\right)}{P(X_0 = r)} \\
&= \lim_{n \rightarrow \infty} \frac{P(T_A \leq n+1, X_0 = r)}{P(X_0 = r)} = \lim_{n \rightarrow \infty} P(T_A \leq n+1, X_0 = r) \\
&= \lim_{n \rightarrow \infty} \left[\sum_{b \in E} p_{rb} P(T_A \leq n, X_0 = b) \right] = \sum_{b \in E} p_{rb} \lim_{n \rightarrow \infty} P(T_A \leq n, X_0 = b) \\
&= \sum_{b \in E} p_{rb} P(T_A < \infty, X_0 = b).
\end{aligned}$$

Daraus folgt die Behauptung. \square

8.14 Beispiel: (Ruinspiel)

Zwei Spieler starten ein Spiel mit Startkapital $N_1, N_2 \in \mathbb{N}$. In jeder Runde gewinnt/verliert Spieler 1 von /an Spieler 2 eine feste Menge seines Kapitals mit der festen Wahrscheinlichkeit $p = \frac{1}{2}$.

In dieser Situation ist natürlich die Frage, ob oder wann Spieler 1 ruiniert wird, besonders interessant.

Um diese Frage zu beantworten setzt man $E := \{0, 1, \dots, N_1 + N_2\}$ und $A := \{0\}$. Gesucht ist nun die Wahrscheinlichkeit

$$\alpha_i := \alpha_i^A = P(T_A < \infty | X_0 = i)$$

für $i = 0, 1, \dots, N_1 + N_2$.

Offenbar gilt

$$\alpha_0 = 1, \alpha_{N_1+N_2} = 0 \text{ und für alle } i = 1, 2, \dots, N_1 + N_2 - 1 : \alpha_i = \frac{1}{2}(\alpha_{i+1} + \alpha_{i-1}) \quad (\star).$$

Betrachte nun den Vektorraum

$$V := \{(\alpha_0, \dots, \alpha_{N_1+N_2}) \in \mathbb{R}^{N_1+N_2+1} \mid (\star) \text{ gilt}\}.$$

Dies ist ein zweidimensionaler Unterraum vom $\mathbb{R}^{N_1+N_2+1}$ mit Basis

$$(1, \dots, 1) \text{ und } (0, 1, \dots, N_1 + N_2).$$

Als Ansatz wählt man nun

$$(\alpha_0, \dots, \alpha_{N_1+N_2}) = c(1, \dots, 1) + d(0, 1, \dots, N_1 + N_2).$$

Dann liefern die Nebenbedingungen $\alpha_0 = 1$ und $\alpha_{N_1+N_2} = 0$ sofort

$$c = 1 \text{ und } d = -\frac{1}{N_1 + N_2}.$$

Dann folgt für alle $i \in \{0, 1, \dots, N_1 + N_2\}$ sofort

$$\alpha_i = 1 - \frac{i}{N_1 + N_2}.$$

Besonders interessant ist nun, wann der Spieler 1 mit Startkapital $i = N_1$ pleite ist. Mit der obigen Formel ergibt sich dafür sofort ²³

$$\alpha_{N_1} = 1 - \frac{N_1}{N_1 + N_2} = \frac{N_2}{N_1 + N_2}.$$

8.15 Langzeitverhalten zeithomogener Markovketten:

Es sei $(X_n)_{n \in \mathbb{N}_0}$ eine zeithomogene Markov-Kette auf $E = \{a_1, \dots, a_N\}$ mit Startverteilung

$$\mathcal{M}^1(E) \ni P_{X_0} \hat{=} P = (p_1, \dots, p_N) \in [0, 1]^N, \quad \sum_{k=1}^N p_k = 1$$

und Übergangsmatrix

$$S = (p_{ij})_{i,j=1,\dots,N} \in \mathcal{M}_N(\mathbb{R}),$$

wobei die p_{ij} die Wahrscheinlichkeiten sind vom Zustand a_i zum Zustand a_j zu kommen. Dann gilt natürlich

$$p \cdot S \hat{=} P_{X_1} \text{ und } p \cdot S^n \hat{=} P_{X_n} \quad \forall n \in \mathbb{N}_0 \text{ (mit } S^0 := I_N).$$

²³ Ende der dreiundzwanzigsten Vorlesung vom Fr, 27.06.14

Man interessiert sich nun in vielen Fällen dafür, ob P_{X_n} konvergiert und was $\lim_{n \rightarrow \infty} S^n$ ist. Einerseits kann man die Eigenwerte und Eigenvektoren von S und somit die Jordansche Normalform berechnen und damit alles direkt ausrechnen.

Ein anderer Zugang verwendet den Banachschen Fixpunktsatz.

8.16 Fakten:

Versehe den Zeilenvektorraum \mathbb{R}^N mit der $\|\cdot\|_1$ -Norm. Man definiert

$$\mathcal{M}_N^1 := \{(p_1, \dots, p_N) \in [0, 1]^N \mid \sum_{k=1}^N p_k = 1\} \subseteq \mathbb{R}^N.$$

Dies ist eine kompakte Teilmenge des \mathbb{R}^N , aber kein Untervektorraum. \mathcal{M}_N^1 entspricht der Menge aller Wahrscheinlichkeitsmaße auf E . Durch

$$d(p, q) := \|p - q\|_1 \quad p, q \in \mathcal{M}_N^1$$

wird eine Metrik auf \mathcal{M}_N^1 definiert. Also ist (\mathcal{M}_N^1, d) ein kompakter metrischer Raum und insbesondere vollständig. Sei nun $S \in \mathcal{M}_N(\mathbb{R})$ eine stochastische Matrix. Betrachte die lineare Abbildung

$$\varphi_S : \mathcal{M}_N^1 \rightarrow \mathcal{M}_N^1, \quad \varphi_S(p) := p \cdot S.$$

Dann gilt:

a) $\|\varphi_S(p) - \varphi_S(q)\|_1 \leq \|p - q\|_1 \quad \forall p, q \in \mathcal{M}_N^1$, denn

$$\|\varphi_S(p) - \varphi_S(q)\|_1 = \|(p - q)S\|_1 \leq \|S\| \|p - q\|_1 \leq \|p - q\|_1.$$

Also ist φ_S Lipschitz-stetig.

b) Nun sei $S \in \mathcal{M}_N(\mathbb{R})$ eine stochastische Matrix, sodass alle Einträge mindestens eine Spalte nur aus $(0, 1]$ sind. Sei $j_0 \in \{1, \dots, N\}$ der Index dieser Spalte. Setze

$$\varepsilon := \min_{i=1, \dots, N} p_{ij_0} \in (0, 1].$$

Ferner definiere

$$E := \begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & \underbrace{1}_{j_0} & 0 & \dots & 0 \end{pmatrix} \in \mathcal{M}_N(\mathbb{R}).$$

Dann gilt offenbar

$$(p - q)E = 0 \quad \forall p, q \in \mathcal{M}_N^1.$$

d) Ferner sei $\tilde{S} := \frac{1}{1-\varepsilon}(S - \varepsilon E) \in \mathcal{M}_N(\mathbb{R})$ für $\varepsilon < 1$. Diese Matrix ist stochastisch, da alle Einträge zwischen 0 und 1 sind und alle Zeilensummen 1 sind.

Weiterhin gilt für alle $p, q \in \mathcal{M}_N^1$ stets

$$p\tilde{S} - q\tilde{S} = \frac{1}{1-\varepsilon}(pS - qS - \underbrace{\varepsilon(p - q)E}_{=0}) = \frac{1}{1-\varepsilon}(p - q)S.$$

Daraus folgt sofort mit Teil a)

$$\|pS - qS\|_1 = (1 - \varepsilon) \underbrace{\| (p - q)\tilde{S} \|_1}_{=\varphi_{\tilde{S}}(p) - \varphi_{\tilde{S}}(q)} \leq \underbrace{(1 - \varepsilon)}_{<1} \|p - q\|_1 < \|p - q\|_1.$$

Die Lipschitz-Konstante ist also echt kleiner 1 und deshalb ist φ_S eine Kontraktion, sofern die Matrix S die Bedingung aus Teil b) erfüllt.

8.17 Satz:

Sei $S \in \mathcal{M}_N(\mathbb{R})$ eine stochastische Matrix, sodass für eine Potenz $k \in \mathbb{N}$ die Matrix S^k mindestens eine Spalte nur mit Einträgen aus $(0, 1]$ hat. Dann existiert genau eine stationäre Verteilung $p \in \mathcal{M}_N^1$ und man hat

$$\lim_{n \rightarrow \infty} S^n = \begin{pmatrix} p \\ \vdots \\ p \end{pmatrix} \in \mathcal{M}_N(\mathbb{R}),$$

wobei p ein Zeilenvektor ist.

Beweis:

a) Wir beweisen den Satz zunächst für den Fall $k = 1$. Sei also $S \in \mathcal{M}_N(\mathbb{R})$ eine stochastische Matrix mit mindestens einer Spalte nur mit Einträgen aus $(0, 1]$. Dann sind die Resultate aus 8.16 b) anwendbar und $\varphi_S : \mathcal{M}_N^1 \rightarrow \mathcal{M}_N^1$ ist eine Kontraktion auf dem vollständigen metrischen Raum $\mathcal{M}_N(\mathbb{R})$. Deshalb folgt mit dem *Banachschen Fixpunktsatz* sofort, dass es genau eine Verteilung $p \in \mathcal{M}_N^1$ mit $\varphi_S(p) = p$, also eine stationäre Verteilung, gibt. Der Banachsche Fixpunktsatz liefert auch

$$S^{n+1} = S \cdot S^n = \begin{pmatrix} s_1 \\ \vdots \\ s_N \end{pmatrix} \cdot S^n = \begin{pmatrix} s_1 \cdot S^n \\ \vdots \\ s_N \cdot S^n \end{pmatrix} = \begin{pmatrix} \varphi_S^n(s_1) \\ \vdots \\ \varphi_S^n(s_N) \end{pmatrix} \rightarrow \begin{pmatrix} p \\ \vdots \\ p \end{pmatrix} \text{ für } n \rightarrow \infty,$$

wobei $s_1, \dots, s_N \in \mathcal{M}_N^1$ die Zeilen von S sind.

b) Sei nun $k \in \mathbb{N}$ beliebig und $S \in \mathcal{M}_N(\mathbb{R})$ eine stochastische Matrix, sodass S^k mindestens eine Spalte nur mit Einträgen aus $(0, 1]$ besitzt. Nach Teil a) besitzt S^k dann genau eine stationäre Verteilung $p \in \mathcal{M}_N^1$ mit

$$\lim_{n \rightarrow \infty} (S^k)^n = \begin{pmatrix} p \\ \vdots \\ p \end{pmatrix}.$$

Sei nun $q \in \mathcal{M}_N^1$ eine stationäre Verteilung von S , dann gilt

$$q = qS = qS^2 = \dots = qS^{k-1} = qS^k.$$

Also ist q auch stationäre Verteilung von S^k . Nach Teil a) besitzt aber S^k nur die stationäre Verteilung p . Also besitzt auch S nur die stationäre Verteilung p .

Sei nun $r \in \{0, \dots, k-1\}$, dann hat man

$$S^{r+n \cdot k} = S^r \cdot (S^k)^n = \begin{pmatrix} s_1 \\ \vdots \\ s_N \end{pmatrix} \cdot (S^k)^n = \begin{pmatrix} s_1 \cdot (S^k)^n \\ \vdots \\ s_N \cdot (S^k)^n \end{pmatrix} = \begin{pmatrix} \varphi_S^{kn}(s_1) \\ \vdots \\ \varphi_S^{kn}(s_N) \end{pmatrix} \rightarrow \begin{pmatrix} p \\ \vdots \\ p \end{pmatrix} \text{ für } n \rightarrow \infty,$$

wobei $s_1, \dots, s_N \in \mathcal{M}_N^1$ die Zeilen von S^r sind.

Also konvergieren alle Teilfolgen von S^n gegen $\begin{pmatrix} p \\ \vdots \\ p \end{pmatrix}$ und deshalb auch S^n selber. \square

8.18 Beispiele:

a) Es sei

$$S := \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \in \mathcal{M}_5(\mathbb{R}).$$

Dann ist

$$S^2 := \begin{pmatrix} \frac{5}{9} & \frac{3}{9} & \frac{1}{9} & 0 & 0 \\ \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{9} & 0 \\ \frac{1}{9} & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{9} \\ 0 & \frac{1}{9} & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \\ 0 & 0 & \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \end{pmatrix} \in \mathcal{M}_5(\mathbb{R}).$$

Die dritte Spalte von S^2 besitzt nur Einträge aus $(0, 1]$. Nach Satz 8.17 besitzt S also genau eine stationäre Verteilung $p \in \mathcal{M}_N^1$. Um diese zu berechnen, löst man das LGS

$$pS = p \Leftrightarrow p(S - I_5) = 0.$$

Also Lösung in diesem Fall erhält man für die stationäre Verteilung

$$p = \frac{1}{5} \cdot (1, 1, 1, 1, 1).$$

Also gilt

$$\lim_{n \rightarrow \infty} S^n = \frac{1}{5} \cdot \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}.$$

b) Ruinspiel: Zwei Spieler starten ein Spiel mit Startkapital $N_1, N_2 \in \mathbb{N}$. In jeder Runde gewinnt/verliert Spieler 1 von /an Spieler 2 eine feste Menge seines Kapitals mit der

festen Wahrscheinlichkeit $p \in (0, 1)$.

Man hat dann die Übergangsmatrix

$$S = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 1-p & 0 & p & & & & \vdots \\ 0 & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & 1-p & 0 & p \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix} \in \mathcal{M}_{N_1+N_2}(\mathbb{R}).$$

Offenbar gilt

$$S^k = \begin{pmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ * & * & * & * & * & * & * \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix} \quad \forall k \in \mathbb{N}.$$

Satz 8.17 ist also nicht anwendbar, da in jeder Spalte mindestens eine 0 steht. In diesem Fall gibt es nicht eine eindeutige stationäre Verteilung, sondern unendlich viele stationäre Verteilung von der Form

$$(q, 0, \dots, 0, 1-q) \in [0, 1]^{N_1+N_2}.$$

Andererseits gilt aber trotzdem stets

$$\lim_{n \rightarrow \infty} S^n = \begin{pmatrix} q_0 & 0 & \cdots & 0 & 1-q_0 \\ \vdots & \vdots & & \vdots & \vdots \\ q_{N_1+N_2} & 0 & \cdots & 0 & 1-q_{N_1+N_2} \end{pmatrix} \in \mathcal{M}_{N_1+N_2}(\mathbb{R})$$

für $q_0, \dots, q_{N_1+N_2} \in [0, 1]$ passend.

c) Es sei

$$S := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}).$$

Dann ist offenbar

$$S^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2 \quad \text{und} \quad S^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = S \quad \forall k \in \mathbb{N}.$$

Offenbar ist auch hier Satz 8.17 nicht anwendbar. Tatsächlich ist die Folge S^n auch nicht konvergent. Andererseits existiert genau eine stationäre Verteilung

$$p = \frac{1}{2}(1, 1) \in \mathcal{M}_2^1.$$

9 Statistik: Grundlagen des Schätzens

Motivation: Eine Münze mit Kopf (1) und Zahl (0) werde 100-mal unabhängig geworfen. Die Wahrscheinlichkeit eine 1 zu erhalten sei $p \in (0, 1)$ und die Wahrscheinlichkeit eine 0 zu erhalten sei $1 - p \in (0, 1)$. Ziel ist es nun aus dem erzielten Ergebnis Rückschlüsse auf die Wahrscheinlichkeit p zu erhalten und einen möglichst genauen Wert für p zu *schätzen*.

Mathematisch formuliert hat man $\Omega := \{0, 1\}^{100}$ mit Wahrscheinlichkeitsmaßen

$$P_p := (p\delta_1 + (1-p)\delta_0) \times \cdots \times (p\delta_1 + (1-p)\delta_0) \in \mathcal{M}^1(\Omega).$$

Betrachte die Zufallsvariablen X_1, \dots, X_{100} mit

$$X_k : \Omega \rightarrow \{0, 1\}, (\omega_1, \dots, \omega_{100}) \mapsto \omega_k \text{ für } k = 1, \dots, 100.$$

X_k entspricht dann dem Ergebnis des k -ten Wurfs. Eine sinnvolle Schätzung für p aus dem Ergebnis der 100 Würfe ist sicherlich das *empirische Mittel*

$$\overline{X_{100}}(\omega) := \bar{X}(\omega) := \frac{1}{100} \sum_{k=1}^{100} X_k(\omega).$$

Es ergeben sich nun einige Fragen: Wann ist ein Schätzer gut? Was bedeutet in diesem Zusammenhang überhaupt „gut“? Was ist überhaupt ein Schätzer? Wie findet man gute Schätzer?

9.1 Definition:

Es seien (Ω, \mathcal{A}) ein Messraum mit Ω als Menge der möglichen Ergebnisse und (Θ, \mathcal{B}) ein Messraum, der die Indexmenge aller möglichen Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) beschreibt (meist ist $\Theta \subseteq \mathbb{R}^d$). Ferner sei

$$\mathcal{M}^\Theta := \{P_\vartheta | \vartheta \in \Theta\} \subseteq \mathcal{M}^1(\Omega, \mathcal{A})$$

die Menge aller möglichen Wahrscheinlichkeitsmaße, die dem Experiment zu Grunde liegen.

Jede messbare Abbildung

$$\hat{\vartheta} : (\Omega, \mathcal{A}) \rightarrow (\Theta, \mathcal{B})$$

heißt *Schätzer* für $\vartheta \in \Theta$.

[Achtung: Diese Definition sagt nichts über die Qualität eines Schätzers aus.]

Deutung: Die echte Verteilung $p \in \mathcal{M}^1(\Omega, \mathcal{A})$ ist unbekannt. Man weiß nur, dass auf jeden Fall $p \in \mathcal{M}^\Theta$ gelten muss. Dann möchte man von der Beobachtung $\omega \in \Omega$ des Experiments möglichst gut auf das $\vartheta \in \Theta$ mit $P = P_\vartheta$ schließen.

Im Beispiel aus der Motivation zu Anfang dieses Kapitels hat man

$$\Theta := [0, 1] \text{ und } p \stackrel{\wedge}{=} \vartheta$$

und

$$P_\vartheta := (\vartheta\delta_1 + (1 - \vartheta)\delta_0) \times \cdots \times (\vartheta\delta_1 + (1 - \vartheta)\delta_0) \in \mathcal{M}^1(\Omega).$$

Weiterhin sind

$$\Omega = \{0, 1\}^{100}, \mathcal{A} = \mathcal{P}(\Omega), \mathcal{B} = \mathcal{B}([0, 1])$$

und für die Zufallsvariablen hat man

$$\omega \mapsto \hat{\vartheta}(\omega) := \overline{X_{100}}(\omega) = \frac{1}{100} \sum_{k=1}^{100} \omega_k.$$

Diese Abbildung ist per Definition messbar und deshalb ein Schätzer.

Es existieren allgemeine Methoden, um in verschiedenen Fällen gute Schätzer explizit zu konstruieren.²⁴

9.2 Maximum-Likelihood-Verfahren:

Idee: Ein Experiment liefert eine Realisierung $\omega \in \Omega$. Wähle nun einen Schätzer $\hat{\vartheta}(\omega) \in \Theta$ so, dass die „Wahrscheinlichkeit $P_\vartheta(\{\omega\})$ für des gegebene $\omega \in \Omega$ maximal ist unter allen $\vartheta \in \Theta$ “.

(Man möchte also aus der Wirkung $\omega \in \Omega$ auf die Ursache $\hat{\vartheta}(\omega) \in \Theta$ schließen.)

1. Diskreter Fall: Sei Ω höchstens abzählbar und $\mathcal{A} := \mathcal{P}(\Omega)$. Dann betrachtet man die *Likelihood-Funktion*

$$L : \Omega \times \Theta \rightarrow [0, 1] \text{ mit } L(\omega, \vartheta) = P_\vartheta(\{\omega\}).$$

2. Kontinuierlicher Fall: Sei $\Omega \subseteq \mathbb{R}^d$ und $\mathcal{A} := \mathcal{B}(\Omega)$. Man nimmt nun an, dass P_ϑ für $\vartheta \in \Theta$ eine Dichte f_ϑ bzgl. des Lebesgue-Maßes λ^d auf \mathbb{R}^d besitzt. Dann betrachtet man die *Likelihood-Funktion*

$$L : \Omega \times \Theta \rightarrow [0, \infty) \text{ mit } L(\omega, \vartheta) = f_\vartheta(\omega).$$

In beiden Fällen ist $\hat{\vartheta} : \Omega \rightarrow \Theta$ genau dann der *Maximum-Likelihood-Schätzer* für $\vartheta \in \Theta$, wenn für alle $\omega \in \Omega$ stets

$$L(\omega, \hat{\vartheta}(\omega)) = \max_{\vartheta \in \Theta} L(\omega, \vartheta)$$

gilt.

²⁴ Ende der vierundzwanzigsten Vorlesung vom Mi, 02.07.14

9.3 Beispiele:

a) Eine Münze mit den Seiten 1 und 0 mit Wahrscheinlichkeit $p \in (0, 1)$ für 1 und $1 - p$ für 0 werde n -mal unabhängig voneinander geworfen. Dann ist $\Omega := \{0, 1\}^n$, $\Theta := [0, 1]$ und man hat

$$P_\vartheta = \underbrace{(p\delta_1 + (1-p)\delta_0) \times \cdots \times (p\delta_1 + (1-p)\delta_0)}_{n\text{-mal}} \in \mathcal{M}^1(\Omega).$$

Für $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ setze

$$k := k(\omega) := \sum_{i=1}^n \omega_i.$$

Hier hängt für alle $\vartheta \in \Theta$ das P_ϑ nur von ϑ und $k(\omega)$ ab. Also gilt

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

Dabei ist $p = \vartheta \in [0, 1]$. Man sucht nun das Maximum:

$$d_p L(\omega, \vartheta) = p^{k-1}(1-p)^{n-k-1}[(1-p)k - p(n-k)] \stackrel{!}{=} 0 \Rightarrow p_0 = 0, p_1 = 1, p_2 = \frac{k}{n}.$$

Da L eine stetige Funktion auf einem kompakten Intervall ist, existiert auch ein Maximum und es liegt bei $p_2 = \frac{k}{n}$. Man hat also den Maximum-Likelihood-Schätzer

$$\overline{X}_n := \frac{1}{n} \sum_{i=1}^n \omega_i.$$

b) Binomialverteilung: Sei $\Omega := \{0, \dots, n\}$ und $\Theta := [0, 1]$. Für $\vartheta = p$ betrachte $P_\vartheta = B_{n,\vartheta}$ und benutze die Konvention $B_{n,0} := \delta_0$, $B_{n,1} := \delta_n$.

Man hat dann die *Likelihood-Funktion*

$$L(\omega, \vartheta) = B_{n,\vartheta}(\{\omega\}) = \binom{n}{k} p^\omega (1-p)^{n-\omega}.$$

Analog zu Teil a) ergibt sich der Maximum-Likelihood-Schätzer für $\hat{\vartheta} = p \in [0, 1]$

$$\hat{\vartheta}(\omega) = \frac{\omega}{n}.$$

c) Poissonverteilung: Sei $\Omega := \mathbb{N}_0$ und $\vartheta \in \Theta := [0, \infty)$. Für $P_\vartheta := \Pi_\vartheta$ mit der Konvention $\Pi_0 := \delta_0$ hat man die Likelihood-Funktion

$$L(\omega, \vartheta) = e^{-\vartheta} \cdot \frac{\vartheta^\omega}{\omega!}$$

und als Maximum erhält man

$$d_\vartheta L(\omega, \vartheta) = -e^{-\vartheta} \cdot \frac{\vartheta^\omega}{\omega!} + e^{-\vartheta} \cdot \omega \cdot \frac{\vartheta^{\omega-1}}{\omega!} = e^{-\vartheta} \cdot \frac{\vartheta^{\omega-1}}{\omega!} (\omega - \vartheta) \stackrel{!}{=} 0 \Rightarrow \vartheta_0 = 0, \vartheta_1 = \omega.$$

Das Maximum liegt offenbar bei $\vartheta_2 = \omega$. Analog zu Teil a) ergibt sich der Maximum-Likelihood-Schätzer für $\hat{\vartheta} \in [0, \infty)$

$$\hat{\vartheta}(\omega) = \omega.$$

d) Exponentialverteilung: Sei $\Omega := \mathbb{R}$ und $\vartheta \in \Theta := (0, \infty)$. Für $P_\vartheta := \exp_\vartheta$ mit der Dichtefunktion

$$f_\vartheta(x) = \vartheta e^{-\vartheta x} \mathbf{1}_{(0, \infty)}$$

hat man die Likelihood-Funktion

$$L(x, \vartheta) = \vartheta e^{-\vartheta x}$$

und als Maximum erhält man

$$d_\vartheta L(x, \vartheta) = e^{-\vartheta x} - x\vartheta e^{-\vartheta x} = e^{-\vartheta x} (1 - x\vartheta) \stackrel{!}{=} 0 \Rightarrow \vartheta = \frac{1}{x}.$$

Analog zu Teil a) ergibt sich der Maximum-Likelihood-Schätzer für $\hat{\vartheta} \in (0, \infty)$

$$\hat{\vartheta}(x) = \frac{1}{x}.$$

e) μ -Schätzung für N_{μ, σ^2} -verteilte Messfehler: Häufig sind Messfehler N_{μ, σ^2} -verteilt mit unbekanntem $\mu \in \mathbb{R}$ und bekanntem $\sigma^2 > 0$. Man möchte nun μ möglichst gut schätzen. Setze dazu $\Omega := \mathbb{R}$ und $\mu = \vartheta \in \Theta := \mathbb{R}$. Für $P_\vartheta := N_{\vartheta, \sigma^2}$ mit der Dichtefunktion

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

hat man die Likelihood-Funktion

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

und als Maximum erhält man analog zu a)

$$\theta = x.$$

Ebenfalls analog zu Teil a) ergibt sich der Maximum-Likelihood-Schätzer für $\hat{\vartheta} \in \mathbb{R}$

$$\hat{\vartheta}(x) = x.$$

In vielen Fällen ist der Erwartungswert der unbekannte Parameter.

9.4 Standard-Mittelwertschätzer:

Gegeben: Indexmenge $\Theta, \mathcal{M}^\Theta \subseteq \mathcal{M}^1(\mathbb{R}), \Omega \subset \mathbb{R}$, Zufallsexperiment mit Ergebnissen in Ω .
 Oft: $\Theta, \mathcal{M}^\Theta$ unübersichtlich und groß.

Bekannt: Für alle $\vartheta \in \Theta$ ist der Erwartungswert

$$\mu(\vartheta) := \int_{\mathbb{R}} x dP_\vartheta(x) \in \mathbb{R}.$$

Idee: Schätze nur $\mu(\vartheta)$ statt ϑ , da es oft zu schwierig ist das „echte“ θ zu schätzen.

Praxis: Wiederhole das Experiment n -mal unabhängig. Dann erhält man n unabhängige identisch verteilte P_ϑ -verteilte Zufallsvariablen X_1, \dots, X_n mit einem unbekanntem $\vartheta \in \theta$.
 Schätzer für $\mu(\vartheta)$ statt für ϑ nennt man auch Standard-Mittelwertschätzer:

$$\overline{X}_n := \frac{1}{n} \sum_{k=1}^n X_k.$$

Dieser hat die Eigenschaften:

a) Sei $\vartheta \in \Theta$. Man nimmt nun an, dass das ϑ das Richtige ist. Dann ist auch P_ϑ das richtige Wahrscheinlichkeitsmaß. Zu jedem $\vartheta \in \Theta$ gibt es einen Erwartungswert E_ϑ mit

$$E_\vartheta := E_\vartheta(\overline{X}_n) = \frac{1}{n} \cdot n E(X_1) = \int_{\mathbb{R}} x dP_\vartheta(x) = \mu(\vartheta).$$

Man sagt : „ \overline{X}_n ist *erwartungstreuer Schätzer* für $\mu(\vartheta)$ “.

b) Nach den Gesetzen der großen Zahlen gilt

$$\overline{X}_n \rightarrow \mu(\vartheta) \text{ für } n \rightarrow \infty \text{ f.s. (also insbesondere stochastisch).}$$

Man sagt : „Die Folge (\overline{X}_n) ist *konsistent*“.

Die Eigenschaften a) und b) sind natürlich wünschenswert für einen Schätzer.

9.5 Schätzen von Parametern:

Gegeben: Große Indexmenge Θ , sowie $(\Omega, \mathcal{A}), \mathcal{M}^\Theta \subseteq \mathcal{M}^1(\Omega)$.

Idee: $\Theta, \mathcal{M}^\Theta$ sehr groß und unübersichtlich. Schätze statt des echten ϑ eine passende Kenngröße $g(\vartheta) \subseteq \mathbb{R}^d$.

Beispiel: Sei $\Omega = \mathbb{R}, \mathcal{A} = \mathcal{B}(\mathbb{R}), \mathcal{M}^\Theta = \{P \in \mathcal{M}^1(\mathbb{R}, \mathcal{B}(\mathbb{R})) | P \text{ hat 2.tes Moment}\}$.

Setze $g := (g_1, g_2)^T$ und

$$g_1(\vartheta) := \mu(\vartheta) = \int_{\mathbb{R}} x dP_\vartheta(x)$$

sowie

$$g_2(\vartheta) := \sigma^2(\vartheta) = \int_{\mathbb{R}} x^2 dP_\vartheta(x) - g_1(\vartheta)^2.$$

Ziel: SchlieÙe von einem oder mehreren unabhängigen Ergebnissen auf $g(\vartheta)$ möglichst gut!

Definition: In obiger Situation heißt eine messbare Abbildung

$$\hat{g}: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$$

ein Schätzer für g .

9.6 Beispiel: (Ein Mittelwertschätzer)

n -fache Wiederholung eines Experiments. Setze

$$\Theta := \left\{ P \in \mathcal{M}^1(\mathbb{R}) : \int_{\mathbb{R}} x dP(x) \text{ existiert} \right\}$$

und wähle n unabhängige \mathbb{R} -wertige identisch $P_{\vartheta} \in \mathcal{M}^{\Theta}$ -verteilte Ergebnisse X_1, \dots, X_n , wobei $\Omega = \mathbb{R}$ und

$$\mathcal{M}^{\Theta} := \left\{ \underbrace{P \times \dots \times P}_{d\text{-mal}} \in \mathcal{M}^1(\mathbb{R}^d) \mid P \in \Theta \right\}.$$

Ziel: Schätze Mittelwert

$$g: \Theta \rightarrow \mathbb{R} \text{ mit } g(P) = \int_{\mathbb{R}} x dP(x).$$

Nehme den *Standard-Mittelwertschätzer*

$$\hat{g}(x) = \frac{1}{n} \sum_{k=1}^n x_k$$

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

9.7 Erwartungstreue Schätzer:

Seien $(\Omega, \mathcal{A}), (\Theta, \mathcal{B}), \mathcal{M}^{\Theta}$ und $g: \Theta \rightarrow \mathbb{R}$ Parameterfunktion mit $d = 1$ wie in 9.5.

Definition: Ein Schätzer $\hat{g}: \Omega \rightarrow \mathbb{R}$ heißt *erwartungstreuer Schätzer für g* , falls für alle $\vartheta \in \Theta$ stets

$$E_{\vartheta}(\hat{g}) = \int_{\Omega} \hat{g}(\omega) dP_{\vartheta}(\omega) = g(\vartheta)$$

gilt.

Deutung: In jeder vorliegenden Situation ergibt sich also die Schätzung \hat{g} im Mittel bezüglich des echten ϑ das echte $g(\vartheta)$.²⁵

²⁵ Ende der fünfundzwanzigsten Vorlesung vom Fr, 04.07.14

9.8 Parameterschätzungen bei Normalverteilungen:

Seien X_1, \dots, X_n reellwertige, unabhängige, identisch N_{μ, σ^2} -verteilte Zufallsvariablen mit $\mu \in \mathbb{R}, \sigma^2 > 0$ unbekannt. Schätze μ, σ^2 aus Beobachtungen $X_1(\omega), \dots, X_n(\omega)$. Ein Maximum-Likelihood Ansatz (ML-Ansatz) liefert:

$$\vartheta = (\mu, \sigma^2) \in \Theta := \mathbb{R} \times (0, \infty)$$

$\Rightarrow (X_1, \dots, X_n)$ hat Dichte in \mathbb{R}^n

$$f_{\mu, \sigma^2}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \underbrace{\frac{1}{\sigma^n} \cdot e^{-\frac{1}{2\sigma^2} \cdot \sum_{k=1}^n (x_k - \mu)^2}}_{=: \tilde{f}_{\mu, \sigma^2}(x_1, \dots, x_n)}.$$

Um den Rechenaufwand zu verringern, maximiert man diese Funktion, indem man $\log(\tilde{f}_{\mu, \sigma^2}(x_1, \dots, x_n))$ maximiert. Man setzt dazu wie gewöhnlich den Gradienten gleich dem Nullvektor und verifiziert mit der Definitheit der Hesse-Matrix, dass es sich tatsächlich um ein Maximum der ML-Funktion handelt. Als Maximum erhält man durch diese Rechnung

$$\mu = \frac{1}{n} \sum_{k=1}^n x_k \quad \text{und} \quad \sigma^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \mu)^2.$$

Fazit: Die ML-Schätzer für (μ, σ^2) sind

$$\overline{X_n} = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{und} \quad \hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \overline{X_n})^2.$$

Auch allgemein sind dies sinnvolle Schätzer für μ und σ^2 . Ist μ bekannt, verwendet man natürlich in der Formel des Schätzers für σ^2 den echten Wert μ statt den zugehörigen Schätzer $\overline{X_n}$. Solche Situationen treten auch tatsächlich auf, beispielsweise bei Spannungsschwankungen.

9.9 Allgemeine Varianzschätzer:

Seien $X_1, \dots, X_n \in \mathcal{L}_2(\Omega, \mathcal{A}, P)$ reellwertige, unabhängige, identisch verteilte Zufallsvariablen mit Erwartungswert $\mu \in \mathbb{R}$, Varianz $\sigma^2 > 0$ und unbekannter Verteilung $Q \in \mathcal{M}^1(\mathbb{R})$. Dann gilt:

a) Der erwartungstreue Standardmittelwertschätzer ist

$$\overline{X_n} = \frac{1}{n} \sum_{k=1}^n X_k.$$

b) Der Standardvarianzschätzer bei bekanntem Mittelwert $\mu \in \mathbb{R}$ motiviert durch 9.8 ist

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2.$$

$\hat{\sigma}_n^2$ ist erwartungstreu, denn

$$E(\hat{\sigma}_n^2) = \frac{1}{n} \sum_{k=1}^n \underbrace{E((X_k - \mu)^2)}_{=\sigma^2} = \frac{1}{n} \cdot n \cdot \sigma^2 = \sigma^2.$$

c) Der Standardvarianzschätzer bei unbekanntem Mittelwert ist

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

$\tilde{\sigma}_n^2$ ist nicht erwartungstreu, denn

$$E(\tilde{\sigma}_n^2) = \frac{1}{n} \sum_{k=1}^n E((X_k - \bar{X}_n)^2) = \frac{1}{n} \cdot n \cdot E((X_1 - \bar{X}_n)^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2.$$

Da $\frac{n-1}{n} < 1 \forall n \in \mathbb{N}$ wird durch $\tilde{\sigma}_n^2$ die echte Varianz unterschätzt. Man kann allerdings $\tilde{\sigma}_n^2$ mit $\frac{n}{n-1}$ durchmultiplizieren und erhält dann offenbar einen erwartungstreuen Schätzer für σ^2 . Also wählt man

$$\hat{s}_n^2 := \frac{n}{n-1} \cdot \tilde{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

als Standardvarianzschätzer bei unbekanntem Mittelwert.

Es folgen nun weitere Qualitätskriterien für Schätzer.

9.10 Konsistenz:

Zum Schätzen eines Parameters $g(\vartheta) \in \mathbb{R}$ werde ein Experiment n -mal unabhängig wiederholt. Dabei sei n variabel. Was passiert nun für $n \rightarrow \infty$?

Definition: Eine Folge (\hat{g}_n) von Schätzern für den unbekannt Parameter g heißt *konsistent*, falls für jedes $\vartheta \in \Theta$ die Zufallsvariablen $\hat{g}_n(\vartheta)$ gegen $g(\vartheta)$ im folgenden Sinne konvergieren:

a) *schwache Konsistenz* : \Leftrightarrow stochastisch konvergent, d.h.

$$\forall \vartheta \in \Theta \forall \varepsilon > 0 : P_\vartheta(|\hat{g}_n(\vartheta) - g(\vartheta)| > \varepsilon) \rightarrow 0 \text{ für } n \rightarrow \infty.$$

b) *starke Konsistenz* : \Leftrightarrow fast sicher konvergent bzgl P_ϑ , d.h.

$$\forall \vartheta \in \Theta : \hat{g}_n(\vartheta) \rightarrow g(\vartheta) \quad P_\vartheta \text{ f.s. für } n \rightarrow \infty.$$

Beispiele:

a) Sei $(X_k) \subseteq \mathcal{L}_2(\Omega, \mathcal{A}, P_\vartheta)$ eine Folge reellwertiger unabhängiger identisch verteilter Zufallsvariablen mit Erwartungswert $\mu \in \mathbb{R}$. Dann gilt

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mu \quad P_\vartheta \text{ f.s. für } n \rightarrow \infty$$

nach dem starken Gesetz der großen Zahlen 7.9.

b) Der Standardvarianzschätzer ist für $(X_k) \subseteq \mathcal{L}_4(\Omega, \mathcal{A}, P_\vartheta)$ auch stark konsistent nach dem starken Gesetz der großen Zahlen 7.9.

Ein quantitatives Qualitätskriterium für Schätzer ist

9.11 Mittlerer quadratischer Fehler:

Sei \hat{g} ein Schätzer für den Parameter $g(\vartheta) \in \mathbb{R}$. Der assoziierte mittlere quadratische Fehler ist

$$R(\vartheta, \hat{g}) := E_\vartheta((\hat{g} - g(\vartheta))^2) \geq 0.$$

Dies ist ein Maß dafür, wie stark \hat{g} von dem echten $g(\vartheta)$ abweicht, sofern ϑ das Richtige ist.

Ziel: Man möchte natürlich $R(\vartheta, \hat{g})$ möglichst klein für möglichst viele (oder sogar alle) $\vartheta \in \Theta$ haben.

Definition: Ein Schätzer \hat{g} heißt *gleichmäßig bester Schätzer* für $g(\vartheta)$, falls für alle möglichen Schätzer \tilde{g} von g und $\vartheta \in \theta$ stets

$$R(\vartheta, \hat{g}) \leq R(\vartheta, \tilde{g})$$

gilt.

[Es gibt Situationen, in denen solche glm. besten Schätzer existieren, dies ist allerdings nur die Ausnahme.]

10 Konfidenzintervalle

Motivation: Gegeben seien ein Messraum (Ω, \mathcal{A}) und eine Menge

$$P^\Theta := \{P_\vartheta | \vartheta \in \Theta\} \subseteq \mathcal{M}^1(\Omega, \mathcal{A}).$$

Dabei ist das echte $\vartheta_0 \in \Theta$ unbekannt. Für eine Parameterfunktion

$$g : \Theta \rightarrow M \subseteq \mathbb{R}$$

ist das korrekte $g(\vartheta_0)$ zu schätzen.

$$\hat{g} : \Omega \rightarrow \mathbb{R}$$

„gut“, falls

$$|\hat{g} - g(\vartheta_0)|$$

„mit hoher Wahrscheinlichkeit klein ist.“

10.1 Definition: (Konfidenzintervall)

Fixiere $\alpha > 0$ als eine kleine Fehlerwahrscheinlichkeit. Eine Familie $(I(\omega))_{\omega \in \Omega}$ von Intervallen heißt *Konfidenzintervall/Konfidenzbereich* zu der zu schätzenden Größe g mit Niveau $1 - \alpha$, falls

$$\forall \vartheta \in \Theta : P_\vartheta(\{\omega \in \Omega | g(\vartheta) \in I(\omega)\}) \geq 1 - \alpha. \quad (44)$$

Also: In jeder möglichen Situation $\vartheta \in \Theta$ wird also aus der Realisierung $\omega \in \Omega$ des Experiments ein Intervall $I(\omega) \subseteq \mathbb{R}$ bestimmt, sodass der echte Parameter $g(\vartheta)$ in diesem Intervall liegt mit einer P_ϑ -Wahrscheinlichkeit von mehr als $1 - \alpha$.

Praxis: Die Problemstellung gibt das $\alpha > 0$ vor. Die Aufgabe ist nun zu allen $\omega \in \Omega$ passende Konfidenzintervalle $I(\omega)$ zu finden.²⁶

10.2 Mittelwertschätzer bei Normalverteilungen:

Seien X_1, \dots, X_n reellwertige, unabhängige, identisch N_{μ, σ^2} -verteilte Zufallsvariablen mit unbekanntem Erwartungswert $\mu \in \mathbb{R}$ und bekannter Varianz $\sigma^2 > 0$. Man hat dann die *Punktschätzung* für μ

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \text{ mit Verteilung } P_{\bar{X}_n} = N_{\mu, \frac{1}{n}\sigma^2}.$$

Sei ferner $0 < \alpha \ll 1$ gegeben.

a) *Zweiseitige Konfidenzintervalle:* Zu α suche das so genannte *zweiseitige Quantil* $\lambda > 0$ mit

$$N_{0,1}([- \lambda, \lambda]) = 1 - \alpha.$$

²⁶ Ende der sechszwanzigsten Vorlesung vom Mi, 09.07.14

Damit ergibt sich das zweiseitige Konfidenzintervall

$$\begin{aligned}
 1 - \alpha &= N_{0,1}([- \lambda, \lambda]) = N_{\mu, \frac{1}{n}\sigma^2} \left(\left[\mu - \lambda \sqrt{\frac{\sigma^2}{n}}, \mu + \lambda \sqrt{\frac{\sigma^2}{n}} \right] \right) \\
 &= P_{\mu} \left(\mu - \lambda \sqrt{\frac{\sigma^2}{n}} \leq \bar{X}_n \leq \mu + \lambda \sqrt{\frac{\sigma^2}{n}} \right) \\
 &= P_{\mu} \left(\bar{X}_n - \lambda \sqrt{\frac{\sigma^2}{n}} \leq \mu \leq \bar{X}_n + \lambda \sqrt{\frac{\sigma^2}{n}} \right).
 \end{aligned}$$

Also ist

$$\left(\left[\bar{X}_n(\omega) - \lambda \sqrt{\frac{\sigma^2}{n}}, \bar{X}_n(\omega) + \lambda \sqrt{\frac{\sigma^2}{n}} \right] \right)_{\omega \in \Omega}$$

ein zweiseitiges Konfidenzintervall zum Niveau $1 - \alpha$.

b) *Einseitige Vertrauensintervalle*: Man habe aus der Praxis die Vorgabe: zu kleine Schätzungen sind unerheblich und große große sind sehr unerwünscht. Man soll also sicherstellen, dass nur mit einer Wahrscheinlichkeit kleiner als α zu große Schätzungen auftreten. Vernachlässige auf der anderen Seite aber die zu kleinen Schätzungen.

Man sucht nun zu α aus einer Tabelle das so genannte *einseitige Quantil* $u > 0$ zu α mit

$$N_{0,1}([-u, \infty)) = N_{0,1}((-\infty, u]) = 1 - \alpha.$$

Damit ergibt sich das einseitige Konfidenzintervall

$$\begin{aligned}
 1 - \alpha &= N_{0,1}([-u, \infty)) = N_{\mu, \frac{1}{n}\sigma^2} \left(\left[\mu - u \sqrt{\frac{\sigma^2}{n}}, \infty \right) \right) \\
 &= P_{\mu} \left(\bar{X}_n \geq \mu - u \sqrt{\frac{\sigma^2}{n}} \right) \\
 &= P_{\mu} \left(\mu \leq \bar{X}_n + u \sqrt{\frac{\sigma^2}{n}} \right)
 \end{aligned}$$

Also ist

$$\left(\left(-\infty, \bar{X}_n(\omega) + u \sqrt{\frac{\sigma^2}{n}} \right] \right)_{\omega \in \Omega}$$

ein einseitiges Konfidenzintervall zum Niveau $1 - \alpha$.

10.3 Binomialverteilter Fall:

Seien X_1, \dots, X_n reellwertige, unabhängige, identisch $(p\delta_1 + (1-p)\delta_0)$ -verteilte Zufallsvariablen mit $p \in (0, 1)$ unbekannt.

Man hat dann den Standardmittelwertschätzer

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

Sei ferner $0 < \alpha \ll 1$ gegeben.

Eine Tabelle liefert dann für das zweiseitige Quantil $\lambda > 0$ zu α

$$\begin{aligned} 1 - \alpha &= N_{0,1}([- \lambda, \lambda]) \stackrel{7.18}{\approx} P_p \left(\frac{X_1 + \dots + X_n - np}{\sqrt{np(1-p)}} \in [- \lambda, \lambda] \right) \\ &= P_p \left(p \in \left[\bar{X}_n - \lambda \sqrt{\frac{p(1-p)}{n}}, \bar{X}_n + \lambda \sqrt{\frac{p(1-p)}{n}} \right] \right) \\ &\stackrel{p(1-p) \leq \frac{1}{4}}{\leq} P_p \left(p \in \left[\bar{X}_n - \frac{\lambda}{2\sqrt{n}}, \bar{X}_n + \frac{\lambda}{2\sqrt{n}} \right] \right). \end{aligned}$$

Also ist

$$\left(\left[\bar{X}_n(\omega) - \frac{\lambda}{2\sqrt{n}}, \bar{X}_n(\omega) + \frac{\lambda}{2\sqrt{n}} \right] \right)_{\omega \in \Omega}$$

ungefähr ein zweiseitiges Konfidenzintervall zum Niveau $1 - \alpha$.

Eine für die Praxis wichtige Frage bei Normalverteilungen ist:

Was tut man in 10.2, falls nicht nur μ , sondern auch σ^2 unbekannt sind?

Idee: Ersetze σ^2 durch den Standardvarianzschätzer

$$\hat{s}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Das führt dann zu den „approximativen“ Konfidenzintervallen

$$\left(\left[\bar{X}_n(\omega) - \lambda \sqrt{\frac{\hat{s}_n^2}{n}}, \bar{X}_n(\omega) + \lambda \sqrt{\frac{\hat{s}_n^2}{n}} \right] \right)_{\omega \in \Omega}.$$

Dies wird nun exakt untersucht. Dazu ist es nötig die Verteilung von

$$\frac{\bar{X}_n - n\mu}{\sqrt{n \cdot \hat{s}_n^2}}$$

zu untersuchen.

Dazu nun einige Vorbereitungen

10.4 Mehrdimensionale Normalverteilung:

a) Die d -dimensionale Standardnormalverteilung auf \mathbb{R}^d hat die Dichte

$$f(x) = \frac{1}{(2\pi)^{\frac{d}{2}}} \cdot e^{-\frac{\|x\|_2^2}{2}} \text{ mit der 2-Norm } \|x\|_2^2 = x^T x = x_1^2 + \dots + x_n^2.$$

b) Seien $X : \Omega \rightarrow \mathbb{R}^d$ eine d -dimensional standardnormalverteilte Zufallsvariable, $\mu \in \mathbb{R}^d$ und $A \in \text{GL}_d(\mathbb{R})$. Bestimme die Verteilung der \mathbb{R}^d -wertigen Zufallsvariable $Y(\omega) := (AX(\omega) + \mu)$:

Für jede Borelmenge $B \in \mathcal{B}(\mathbb{R}^d)$ gilt

$$P(AX + \mu \in B) = P(x \in A^{-1}(B - \mu)) = \int_{A^{-1}(B - \mu)} f(x) dx \stackrel{y:=Ax+\mu}{=} \frac{1}{|\det A|} \int_B f(A^{-1}(y - \mu)) dy$$

[Man definiert die positiv definite Matrix $\Sigma := A \cdot A^T$ mit $\det \Sigma = |\det A|^2$.]

$$= \int_B \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} \cdot e^{-\frac{1}{2}(y-\mu)^T (A^T)^{-1} A^{-1} (y-\mu)} dy$$

[Wegen $(A^T)^{-1} A^{-1} = (AA^T)^{-1}$ gilt]

$$= \int_B \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} \cdot e^{-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)} dy$$

Die Zufallsvariable Y hat also auf \mathbb{R}^d die Dichte

$$f_{\mu, \Sigma}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}} \cdot e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}.$$

Die assoziierte Verteilung auf \mathbb{R}^d heißt *d-dimensionale Normalverteilung* mit *Mittelwertvektor* $\mu \in \mathbb{R}^d$ und *Kovarianzmatrix* $\Sigma \in \mathcal{M}_d(\mathbb{R})$.

c) Beispiel: Sei $\mu = 0$ und

$$A \in \mathcal{O}_d(\mathbb{R}) := \{A \in \mathcal{M}_d(\mathbb{R}) \mid A \text{ ist orthogonale Matrix}\}.$$

Ferner sei X eine \mathbb{R}^d -wertige standardnormalverteilte Zufallsvariable. Dann ist die Kovarianzmatrix

$$\Sigma = AA^T = I_d$$

und die Zufallsvariable AX ist ebenfalls d -dimensional standardnormalverteilt.

10.5 Satz:

Seien X_1, \dots, X_n \mathbb{R} -wertige unabhängige identisch $N_{0,1}$ -verteilte Zufallsvariablen. Dann gilt:

a) Die Standardschätzer

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \quad \text{und} \quad \hat{s}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

sind stochastisch unabhängig.

b) Der Schätzer $(n-1)\hat{s}_n^2$ ist χ_{n-1}^2 -verteilt.

Beweis:

Es sei $A \in \mathcal{O}_n(\mathbb{R})$ mit

$$A \cdot \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad [\text{z.B. eine Drehmatrix}]$$

und

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

eine in \mathbb{R}^n standardnormalverteilte Zufallsvariable.

Mit 10.4 c) folgt dann, dass auch die Zufallsvariable

$$Y := AX$$

eine in \mathbb{R}^n standardnormalverteilte Zufallsvariable ist. Wegen der Produktform sind andererseits aber auch die eindimensionale Zufallsvariablen

$$Y_1 := (AX)_1, \dots, Y_n := (AX)_n$$

unabhängige identisch $N_{0,1}$ -verteilte Zufallsvariablen.

Betrachte nun die Matrix

$$P := \frac{1}{n} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & & \vdots \\ 1 & \dots & 1 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R}).$$

Diese besitzt die folgenden Eigenschaften:

$$\boxed{1.} \quad PX = P \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \bar{X}_n \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n.$$

2. P ist orthogonale Projektion von \mathbb{R}^n auf den eindimensionalen Unterraum

$$\text{span} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbb{R} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \subseteq \mathbb{R}^n,$$

denn

$$\langle X - PX, \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \rangle = X_1 + \dots + X_n - \frac{1}{n}(X_1 + \dots + X_n) \cdot \underbrace{\left\langle \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right\rangle}_{=n} = 0.$$

Da $A \in \mathcal{O}_n(\mathbb{R})$ ist, gilt

$$A(I - P)X = \begin{pmatrix} 0 \\ \star \\ \vdots \\ \star \end{pmatrix}.$$

Also haben $AX = APX + A(I - P)X$ und APX in der ersten Komponente den gleichen Eintrag. Dieser ist

$$\frac{1}{\sqrt{n}}(X_1 + \dots + X_n),$$

denn

$$APX = AP \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \overline{X_n} A \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n).$$

Insgesamt hat man also, dass $Y_1 = (AX)_1 = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n) = \sqrt{n} \cdot \overline{X_n}$ von den

$$Y_1^2 + \dots + Y_n^2 = \left(\sum_{k=1}^n Y_k^2 \right) - Y_1^2 = \left(\sum_{k=1}^n X_k^2 \right) - n\overline{X_n}^2 = \sum_{k=1}^n (X_k - \overline{X_n})^2 = (n-1)\hat{s}_n^2.$$

stochastisch unabhängig ist. Daraus folgen die Behauptungen.²⁷ □

10.6 Lemma:

Es seien X, Y unabhängige \mathbb{R} -wertige Zufallsvariablen mit zugehörigen Dichtefunktionen f, g mit

$$g = 0 \text{ auf } (-\infty, 0] \text{ d.h. } P(Y > 0) = 1.$$

Dann ist $\frac{X}{Y}$ eine Zufallsvariable mit zugehöriger Dichtefunktion

$$h(t) := \int_0^{\infty} f(ty)g(y)y dy \quad t \in \mathbb{R}. \quad (45)$$

²⁷ Ende der siebenundzwanzigsten Vorlesung vom Fr, 11.07.14

Beweis:

Für die Verteilungsfunktion von $\frac{X}{Y}$ erhält man $\forall u \in \mathbb{R}$

$$\begin{aligned} P\left(\frac{X}{Y} \leq u\right) &= \int_{\{(x,y) \in \mathbb{R}^2 \mid y > 0, x \leq u \cdot y\}} f(x) \cdot g(y) d(x, y) = \int_0^\infty \int_{-\infty}^{uy} f(x) dx g(y) dy \\ &\stackrel{x=ty}{\stackrel{dx=y dt}{\equiv}} \int_0^\infty \int_{-\infty}^u f(ty) y dt g(y) dy \stackrel{\text{Fubini}}{\equiv} \int_{-\infty}^u \int_0^\infty f(ty) g(y) y dy dt \\ &= \int_{-\infty}^u h(t) dt. \end{aligned}$$

□

10.7 Die Student-Verteilung: (*t*-Verteilung)

Seien $\mathbb{N} \ni n \geq 2$ und X, S unabhängige \mathbb{R} -wertige Zufallsvariablen mit Verteilungen $P_X = N_{0,1}$ und $P_S = \chi_{n-1}^2$. Dann ist

$$T := \sqrt{n-1} \cdot \frac{X}{\sqrt{S}}$$

eine \mathbb{R} -wertige Zufallsvariable mit zugehöriger Dichtefunktion

$$b_{n-1}(x) := \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\sqrt{\pi(n-1)}} \left(1 + \frac{x^2}{n-1}\right)^{-\frac{n}{2}}.$$

Beweis:

Sei $P_S = \chi_{n-1}^2$ mit einer Dichtefunktion f_{n-1} wie im Abschnitt über die χ^2 -Verteilung definiert. Die transformierte Zufallsvariable

$$Y := \sqrt{\frac{S}{n-1}}$$

hat die Dichte

$$g_{n-1}(y) = f_{n-1}((n-1)y^2) \cdot 2y(n-1).$$

Lemma 10.6 liefert dann mit einiger Rechnung die Behauptung. □

Bemerkung: Für $n \rightarrow \infty$ konvergiert b_{n-1} punktweise gegen die Dichtefunktion $f_{0,1}$ der Standardnormalverteilung.

Definition: (Student-Verteilung)

Die Verteilung auf \mathbb{R} mit Dichte b_{n-1} ($n \geq 2$) heißt (Fisher'sche) *t-Verteilung* oder *Studentverteilung*.

10.8 Anwendung auf Konfidenzintervalle:

a) Es seien X_1, \dots, X_n unabhängige \mathbb{R} -wertige identisch $N_{0,1}$ -verteilte Zufallsvariablen. Mit Satz 10.5 folgt dann:

$$P_{\overline{X}_n} = N_{0, \frac{1}{n}} \text{ und } P_{(n-1)\hat{s}_n^2} = \chi_{n-1}^2$$

und die Schätzer $\overline{X}_n, \hat{s}_n^2$ sind unabhängig. Mit 10.7 und $X := \sqrt{n} \cdot \overline{X}_n$ folgt, dass die Zufallsvariable

$$T := \frac{\sqrt{n(n-1)}}{\sqrt{(n-1)\hat{s}_n^2}} \cdot \overline{X}_n = \frac{\sqrt{n}}{\sqrt{\hat{s}_n^2}} \cdot \overline{X}_n$$

ist t_{n-1} -verteilt (d.h. student-verteilt mit $n-1$ Freiheitsgraden).

b) Es seien X_1, \dots, X_n unabhängige \mathbb{R} -wertige identisch N_{μ, σ^2} -verteilte Zufallsvariablen. Es treten die folgenden Probleme auf:

1. σ^2 bekannt und μ unbekannt, μ zu schätzen. [vgl. 10.2]
2. μ bekannt und σ^2 unbekannt, σ^2 zu schätzen. [vgl. 10.9]
3. σ^2 unbekannt und μ unbekannt, μ zu schätzen. [vgl. 10.10]
4. μ unbekannt und σ^2 unbekannt, σ^2 zu schätzen. [vgl. 10.11]

10.9 Schätzen der Varianz bei bekanntem Mittelwert:

Man hat den erwartungstreuen Schätzer

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \mu)^2$$

für die Varianz σ^2 und $0 < \alpha \ll 1$.

a) Einseitige Konfidenzintervalle: (relevant für die Praxis)

Der Schätzer

$$\frac{n}{\sigma^2} \cdot \hat{\sigma}_n^2 = \sum_{k=1}^n \underbrace{\left(\frac{X_k - \mu}{\sqrt{\sigma^2}} \right)^2}_{=: Y_k^2 \text{ mit } P_{Y_k} = N_{0,1}}$$

ist χ_n^2 -verteilt. Man sucht nun in einer Tabelle für die χ_n^2 -Verteilung nach einer Zahl $r > 0$ zu vorgegebenem Niveau $1 - \alpha$ mit

$$1 - \alpha = \chi_n^2([r, \infty)) = P_{\mu, \sigma^2} \left(\frac{n}{\sigma^2} \cdot \hat{\sigma}_n^2 \geq r \right) = P_{\mu, \sigma^2} \left(\sigma^2 \leq \frac{n}{r} \cdot \hat{\sigma}_n^2 \right).$$

Also sind

$$\left[0, \frac{n}{r} \cdot \hat{\sigma}_n^2(\omega) \right]$$

die einseitigen Konfidenzintervalle zum Niveau $1 - \alpha$.

b) Zweiseitige Konfidenzintervalle:

Man sucht nun in einer Tabelle für die χ_n^2 -Verteilung nach Zahlen $0 < a < b$ zu vorgegebenem Niveau $1 - \alpha$ mit

$$\frac{\alpha}{2} = \chi_n^2([0, 1]) = P_{\mu, \sigma^2} \left(\sigma^2 \geq \frac{n}{a} \cdot \hat{\sigma}_n^2 \right)$$

und

$$\frac{\alpha}{2} = \chi_n^2([0, 1]) = P_{\mu, \sigma^2} \left(\sigma^2 \leq \frac{n}{b} \cdot \hat{\sigma}_n^2 \right).$$

Also sind

$$\left[\frac{n}{b} \cdot \hat{\sigma}_n^2(\omega), \frac{n}{a} \cdot \hat{\sigma}_n^2(\omega) \right]$$

die zweiseitigen Konfidenzintervalle zum Niveau $1 - \alpha$.

10.10 Schätzen des Mittelwertes bei unbekannter Varianz:

Man hat den erwartungstreuen Schätzer

$$\hat{s}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

für die Varianz σ^2 und $0 < \alpha \ll 1$.

Nach 10.8 ist die zufallsvariable

$$T := \frac{\sqrt{n}}{\sqrt{\hat{s}_n^2}} \cdot (\bar{X}_n - \mu)$$

t_{n-1} -verteilt.

a) Zweiseitige Konfidenzintervalle: (relevant für die Praxis)

Man sucht nun in einer Tabelle für die t_{n-1} -Verteilung nach einer Zahl $\gamma > 0$ zu vorgegebenem Niveau $1 - \alpha$ mit

$$\begin{aligned} 1 - \alpha &= t_{n-1}([- \gamma, \gamma]) = P_{\mu, \sigma^2} (-\gamma \leq T \leq \gamma) = P \left(-\gamma \leq \sqrt{\frac{n}{\hat{s}_n^2}} (\bar{X}_n - \mu) \leq \gamma \right) \\ &= P \left(\bar{X}_n - \gamma \sqrt{\frac{\hat{s}_n^2}{n}} \leq \mu \leq \bar{X}_n + \gamma \sqrt{\frac{\hat{s}_n^2}{n}} \right) \end{aligned}$$

Also sind

$$\left[\bar{X}_n(\omega) - \gamma \sqrt{\frac{\hat{s}_n^2(\omega)}{n}} \leq \mu \leq \bar{X}_n(\omega) + \gamma \sqrt{\frac{\hat{s}_n^2(\omega)}{n}} \right]$$

die zweiseitigen Konfidenzintervalle zum Niveau $1 - \alpha$.

Die einseitigen Konfidenzintervalle berechnet man analog zu 10.2.

Bemerkungen:

a) Das Vorgehen ist völlig analog zu 10.2 mit bekanntem σ^2 , man ersetzt nur das echte σ^2 durch den Schätzer \hat{s}_n^2 und wählt γ mit Hilfe einer Tabelle für die t_{n-1} Verteilung statt der $N_{0,1}$ -Verteilung.

b) Man sieht leicht, dass t_{n-1} schwach gegen $N_{0,1}$ konvergiert für $n \rightarrow \infty$. In der Praxis ist t_{n-1} also für große n uninteressant. Man ersetzt dann t_{n-1} durch $N_{0,1}$. In der Praxis ist also die t_{n-1} -Verteilung nur für kleine n von Bedeutung.

10.11 Schätzen der Varianz bei unbekanntem Mittelwert:

Man hat den erwartungstreuen Schätzer

$$\hat{s}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$$

für die Varianz σ^2 und $0 < \alpha \ll 1$. Nach Satz 10.5 ist die Zufallsvariable

$$\frac{n-1}{\sigma^2} \hat{s}_n^2$$

χ_{n-1}^2 -verteilt. (Statt χ_n^2 -verteilt wie $\hat{\sigma}_n^2$ bei bekanntem μ .) Nun geht man mit $n-1$ statt n und \hat{s}_n^2 statt $\hat{\sigma}_n^2$ analog zu 10.9 vor und erhält die einseitigen Konfidenzintervalle

$$\left[0, \frac{n-1}{r} \cdot \hat{s}_n^2(\omega) \right]$$

und die zweiseitigen Konfidenzintervalle

$$\left[\frac{n-1}{b} \cdot \hat{s}_n^2(\omega), \frac{n-1}{a} \cdot \hat{s}_n^2(\omega) \right]$$

zum Niveau $1 - \alpha$.²⁸

²⁸ Ende der achtundzwanzigsten Vorlesung vom Mi, 16.07.14

11 Optimale Schätzer (Einführung)

11.1 Wiederholung:

Es seien (Ω, \mathcal{A}) ein Messraum ($\hat{=}$ Experiment) und

$$\mathcal{M}^\Theta := \{P_\vartheta | \vartheta \in \Theta\} \subseteq \mathcal{M}^1(\Omega, \mathcal{A}).$$

Ferner seien $g : \Theta \rightarrow \mathbb{R}$ eine zu schätzende messbare Kenngröße (z.B. Erwartungswert oder Varianz) und $\hat{g} : \Omega \rightarrow \mathbb{R}$ ein möglicher Schätzer für das echte unbekannte $g(\vartheta)$. Dann gilt:

1. \hat{g} soll erwartungstreu sein, d.h.

$$\forall \vartheta \in \Theta : E_\vartheta(\hat{g}) = g(\vartheta).$$

2. Der mittlere quadratische Fehler

$$R(\vartheta, \hat{g}) := E_\vartheta((\hat{g} - g(\vartheta))^2)$$

soll nicht zu groß sein. Dabei gilt

$$R(\vartheta, \hat{g}) = E_\vartheta(((\hat{g} - E_\vartheta(\hat{g})) + (E_\vartheta(\hat{g}) - g(\vartheta)))^2) = \text{Var}_\vartheta(\hat{g}) + (E_\vartheta(\hat{g}) - g(\vartheta))^2.$$

Ist \hat{g} erwartungstreu, so gilt

$$R(\vartheta, \hat{g}) = \text{Var}_\vartheta(\hat{g}).$$

3. Ein Schätzer \hat{g} für g heißt glm bester Schätzer, falls für alle $\vartheta \in \Theta$ und alle Schätzer \tilde{g} von g stets

$$R(\vartheta, \hat{g}) \leq R(\vartheta, \tilde{g})$$

gilt.

4. Ein Schätzer \hat{g} für g heißt glm bester erwartungstreuer Schätzer, falls für alle $\vartheta \in \Theta$ und alle erwartungstreuen Schätzer \tilde{g} von g stets

$$R(\vartheta, \hat{g}) \leq R(\vartheta, \tilde{g})$$

gilt.

11.2 Informationsungleichung:

Es seien Ω höchstens abzählbar, $\mathcal{A} := \mathcal{P}(\Omega)$, $\Theta \subseteq \mathbb{R}$ ein Intervall und

$$\mathcal{M}^\Theta := \{P_\vartheta | \vartheta \in \Theta\} \subseteq \mathcal{M}^1(\Omega, \mathcal{A}).$$

Annahme 1: Für alle $\omega \in \Omega$ und $\vartheta \in \Theta$ gelte

$$f_\omega(\vartheta) := P_\vartheta(\{\omega\}) > 0$$

und $f_\omega(\vartheta)$ sei für alle $\omega \in \Omega$ diffbar.

Setze $L_\omega(\vartheta) := \ln f_\omega(\vartheta)$ mit $L'_\omega(\vartheta) = \frac{f'_\omega(\vartheta)}{f_\omega(\vartheta)}$.

Dann gilt

$$\sum_{\omega \in \Omega} f_\omega(\vartheta) = 1. \quad (46)$$

Für einen erwartungstreuen Schätzer \hat{g} von g gilt dann

$$g(\vartheta) = E_\vartheta(\hat{g}) = \sum_{\omega \in \Omega} \hat{g}(\omega) f_\omega(\vartheta). \quad (47)$$

Annahme 2: Die Summationen in den Gleichungen (46) und (47) dürfen mit Ableitungen nach ϑ vertauscht werden. Dann folgt für die Zufallsvariable

$$L'_\circ(\vartheta) : \omega \mapsto L'_\omega(\vartheta)$$

sofort

$$0 \stackrel{(46)}{=} \sum_{\omega \in \Omega} f'_\omega(\vartheta) = \sum_{\omega \in \Omega} \frac{f'_\omega(\vartheta)}{f_\omega(\vartheta)} f_\omega(\vartheta) = E_\vartheta(L'_\circ(\vartheta)). \quad (48)$$

Ferner gilt

$$g'(\vartheta) \stackrel{(47)}{=} \sum_{\omega \in \Omega} \hat{g}(\omega) \frac{f'_\omega(\vartheta)}{f_\omega(\vartheta)} f_\omega(\vartheta) = E_\vartheta(\hat{g} \cdot L'_\circ(\vartheta)). \quad (49)$$

Damit ergibt sich wegen (48) sofort

$$g'(\vartheta) = E_\vartheta(\hat{g} \cdot L'_\circ(\vartheta)) - E_\vartheta(\hat{g}) E_\vartheta(L'_\circ(\vartheta)) = E_\vartheta(L'_\circ(\vartheta) \cdot (\hat{g} - E_\vartheta(\hat{g}))).$$

Daraus folgt mit der Cauchy-Schwarz Ungleichung

$$g'(\vartheta)^2 = E_\vartheta(L'_\circ(\vartheta) \cdot (\hat{g} - E_\vartheta(\hat{g})))^2 \leq E_\vartheta(L'_\circ(\vartheta)^2) \cdot \text{Var}_\vartheta(\hat{g}). \quad (50)$$

Dies motiviert die

Definition: (Fisher-Information)

Die Zahl

$$I(\vartheta) := E_\vartheta(L'_\circ(\vartheta)^2) = \sum_{\omega \in \Omega} \frac{f'_\omega(\vartheta)^2}{f_\omega(\vartheta)} \in [0, \infty]$$

heißt *Fisher-Information* des zu Grunde liegenden Experiments zum Parameter $\vartheta \in \Theta$.

Eine Umformulierung von (50) liefert die *Informationsungleichung*

$$\text{Var}_\vartheta(\hat{g}) = R(\vartheta, \hat{g}) \geq \frac{g'(\vartheta)^2}{I(\vartheta)} \quad (51)$$

für jeden erwartungstreuen Schätzer \hat{g} von g unter den Annahmen 1 und 2.

Deutung: Wenn $I(\vartheta)$ groß ist, dann liefert eine Realisierung des Experiments eine gute Schätzung für $g(\vartheta)$.

11.3 Satz: (Additivität der Fisher-Information)

Es seien $(\Omega_1, \mathcal{A}_1)$, $(\Omega_2, \mathcal{A}_2)$ diskrete Messräume und $\Theta \subseteq \mathbb{R}$ ein Intervall. Für $i = 1, 2$ seien

$$\mathcal{M}^\Theta := \{P_\vartheta^{(i)} \in \mathcal{M}^1(\Omega_i, \mathcal{A}_i) \mid \vartheta \in \Theta\}$$

die Mengen aller möglichen Verteilungen bei diesen beiden „Teilexperimenten“ mit den Fisherinformationen $I_i(\vartheta)$. Betrachte das Produktexperiment $(\Omega := \Omega_1 \times \Omega_2, \mathcal{A} := \mathcal{P}(\Omega))$ mit

$$\mathcal{M}^\Theta := \{P_\vartheta := P_\vartheta^{(1)} \times P_\vartheta^{(2)} \mid \vartheta \in \Theta\}.$$

Die beiden Teilexperimente sind dann für alle $\vartheta \in \Theta$ unabhängig voneinander. In dieser Situation gilt

$$I(\vartheta) = I_1(\vartheta) + I_2(\vartheta). \quad (52)$$

Beweis:

Gleichung (52) folgt mit kurzer Rechnung aus der Produktregel der Differentialrechnung und den Gleichungen (46) und (48). \square

11.4 Beispiele:

a) Es sei $\Omega := \{0, 1\}$ und $P_\vartheta = B_{1,\vartheta} = \vartheta\delta_1 + (1 - \vartheta)\delta_0$ für $\vartheta \in (0, 1)$. Dann ist

$$f_0(\vartheta) = \vartheta \text{ und } f_1(\vartheta) = 1 - \vartheta.$$

Damit ergibt sich für die Fisher-Information

$$I(\vartheta) = \frac{1}{\vartheta(1 - \vartheta)}.$$

b) Es sei $\Omega := \{0, 1\}^n$ und $P_\vartheta = B_{1,\vartheta} \times \cdots \times B_{1,\vartheta}$ für $\vartheta \in (0, 1)$. Teil a) und Gleichung (52) liefern induktiv

$$I(\vartheta) = \frac{n}{\vartheta(1 - \vartheta)}.$$

c) Es sei $\Omega := \{0, 1, \dots, n\}$ und $P_\vartheta = B_{n,\vartheta}$ für $\vartheta \in (0, 1)$. Nach einiger Rechnung erhält man

$$I(\vartheta) = \frac{n}{\vartheta(1 - \vartheta)}.$$

d) Es sei $\Omega := \mathbb{N}_0$ und $P_\vartheta = \Pi_\vartheta$ (Poisson-Verteilung) für $\vartheta \in (0, \infty)$. Dann gilt für alle $k \in \Omega$

$$f_k(\vartheta) = e^{-\vartheta} \frac{\vartheta^k}{k!} \Rightarrow f'_k(\vartheta) = e^{-\vartheta} \frac{\vartheta^{k-1}}{k!} (k - \vartheta).$$

Eine kurze Rechnung liefert dann

$$I(\vartheta) = \frac{1}{\vartheta}.$$

e) Es sei $\Omega := \mathbb{N}_0^n$ und $P_\vartheta = \Pi_\vartheta \times \cdots \times \Pi_\vartheta$ (Poisson-Verteilung) für $\vartheta \in (0, \infty)$. Teil d) und Gleichung (52) liefern induktiv

$$I(\vartheta) = \frac{n}{\vartheta}.$$

Wende nun diese Ergebnisse auf die Informationsungleichung an, um $g(\vartheta) := \vartheta$ zu schätzen. Dann gilt für alle erwartungstreuen Schätzer \hat{g} von $g(\vartheta) = \vartheta$ mit $g'(\vartheta)$:

a) $Var_\vartheta(\hat{g}) \geq \frac{1}{I(\vartheta)} = \vartheta(1 - \vartheta).$

b) $Var_\vartheta(\hat{g}) \geq \frac{1}{I(\vartheta)} = \frac{\vartheta(1 - \vartheta)}{n}.$

c) analog zu b)

d) $Var_\vartheta(\hat{g}) \geq \frac{1}{I(\vartheta)} = \vartheta.$

e) $Var_\vartheta(\hat{g}) \geq \frac{1}{I(\vartheta)} = \frac{\vartheta}{n}.$

Vergleiche diese unteren Schranken nun mit den konkreten Schätzern:

a) ML-Schätzer $\hat{g}(\omega) = \omega$ mit Varianz $\vartheta(1 - \vartheta)$. ✓

b) ML-Schätzer $\hat{g}(\omega) = \frac{1}{n} \sum_{k=1}^n \omega_k$ mit Varianz $\frac{\vartheta(1 - \vartheta)}{n}$. ✓

c) analog zu b) ✓

d) ML-Schätzer $\hat{g}(\omega) = \omega$ mit Varianz ϑ . ✓

e) ML-Schätzer $\hat{g}(\omega) = \frac{1}{n} \sum_{k=1}^n \omega_k$ mit Varianz $\frac{\vartheta}{n}$. ✓

Daraus folgt der abschließende

11.5 Satz:

Bei $B_{1,\vartheta}$ - und Π_ϑ -verteilten Einzelexperimenten und n -facher unabhängiger Wiederholung ist der Standard-Mittelwertschätzer

$$\hat{g}(\omega) = \frac{1}{n} \sum_{k=1}^n \omega_k$$

der gleichmäßig beste erwartungstreue Schätzer für den Mittelwert.²⁹

²⁹ Ende der neunundzwanzigsten Vorlesung vom Fr, 18.07.14