Stochastik
Prof. Reiß

Sommersemester 2012

Vorlesungsmitschriften[1] Paul Boeck[2]

Zuletzt geändert am 14. Oktober 2012

Fehler bitte an

boeck@math.hu-berlin.de

Inhaltsverzeichnis

---

[1] Nicht von Prof. Reiß korrigiert

[2] mit Hilfe von Stephanie Brandl, Jane Knöchel und Sorin Pascu

I. Wahrscheinlichkeitsräume

I.1. Ereignisse, Wahrscheinlichkeiten und Zufallsvariablen

Wir widmen uns zunächst der mathematischen Modellierung des Zufalls und werden erst später kurz auf Interpretationen eingehen.

Beispiel I.1

a) Das Würfeln mit zwei unterscheidbaren Würfeln. Jeder Würfel wird eine der Zahlen $\{1,2,3,4,5,6\}$ zeigen. Als Ergebnis *nach* dem Würfeln erhalten wir so ein Tupel (n_1,n_2) mit $n_1,n_2 \in \{1,2,3,4,5,6\}$. Interpretation: n_1 ist die Augenzahl von Würfel 1 und n_2 entsprechend von Würfel 2.

   Vor dem Würfelexperiment werden wir unter der Annahme zweier fairer Würfel, die sich nicht gegenseitig beeinflussen, sagen, dass jeder Versuchsausgang mit gleicher Wahrscheinlichkeit (W-keit) eintritt. Weitere Fragen sind dann beispielsweise Wahrscheinlichkeiten für folgende Ereignisse

   A_1 : Es wird ein Pasch gewürfelt.

   A_2 : Die Augensumme ist 7.

   A_3 : Es tritt keine 1 auf.

   Mathematisch werden Ereignisse als Teilmengen der Menge $\Omega = \{1,2,3,4,5,6\}^2$ aller Versuchsausgänge modelliert, also hier:

$$\begin{aligned}
A_1 &= \{(n_1,n_2) \in \Omega \mid n_1 = n_2\} = \{(1,1),\dots,(6,6)\} \\
A_2 &= \{(n_1,n_2) \in \Omega \mid n_1 + n_2 = 7\} = \{(1,6),(2,5),\dots,(6,1)\} \\
A_3 &= \{(n_1,n_2) \in \Omega \mid (n_1 \neq 1) \wedge (n_2 \neq 1)\}
\end{aligned}$$

   Sind alle Versuchsausgänge gleich wahrscheinlich, erhalten wir die Wahrscheinlichkeit für ein Ereignis A, indem wir die Zahl aller für A günstigen Ausgänge durch die Zahl aller möglichen Ausgänge teilen, also hier

$$P(A_1) = \frac{|A_1|}{|\Omega|} = \frac{6}{36} = \frac{1}{6} \quad P(A_2) = \frac{|A_2|}{|\Omega|} = \frac{6}{36} \quad P(A_3) = \frac{|A_3|}{|\Omega|} = \frac{25}{36}$$

b) n-facher Münzwurf. Wir werfen eine Münze n-mal hintereinander. Den Versuchsausgang codieren wir mit einem n-Tupel $(b_1,\dots,b_n) \in \{0,1\}^n$ mit der Interpretation: $b_k = 1$ heißt, „k-ter Wurf ist Kopf" und $b_k = 0$ heißt „k-ter Wurf ist Zahl". Unter fairen Bedingungen werden wir jeden Versuchsausgang wieder dieselbe Wahrscheinlichkeit zuweisen ($= 2^{-n}$).

   Betrachten zu $\Omega = \{0,1\}^n$.

   A_1 : „n-mal Kopf"

$$= \{(b_1,\dots,b_n) = b = \Omega \mid b_1 = b_2 = \dots = b_n = 1\} \implies P(A_1) = \frac{|A_1|}{|\Omega|} = \frac{1}{2^n} = 2^{-n}$$

A_2 : „$(n-1)$-mal Kopf“

$$= \{(0,1,1,\dots,1),(1,0,1,1,\dots,1),\dots,(1,1,\dots,1,0)\} \implies P(A_2) = \frac{|A_2|}{|\Omega|} = \frac{n}{2^n}$$

A_3 : „mindestens einmal Zahl“

$$= \{b \in \Omega \mid b \neq (1,1,\dots,1)\} = A_1^C \implies P(A_3) = \frac{|A_3|}{|\Omega|} = \frac{2^n - 1}{2^n} = 1 - 2^{-n} = 1 - P(A_3^C)$$

c) beliebig häufiger Münzwurf. Es ist natürlich zu fragen, was passiert, wenn eine faire Münze beliebig häufig („∞ oft“) geworfen wird. Dann erhalten wir

$$\Omega = \text{„Menge aller 0-1-Folgen“} = \{0,1\}^{\mathbb{N}}$$

Also $b \in \Omega$ ist eine Folge $b = (b_1, b_2, \dots) = (b_n)_{n\in\mathbb{N}}$ mit Werten in $\{0,1\}$. Betrachte nun die Ereignisse

A_1 : „immer nur Kopf“ $= \{b \in \Omega \mid \forall k \geq 1 : b_k = 1\} = \bigcap_{k\geq 1}\{b \in \Omega \mid b_k = 1\} = \{(1,1,1,\dots)\}$. Aufgrund von Beispiel b) würden wir aus „Stetigkeitsgründen“ A_1 die Wahrscheinlichkeit $\lim_{n\to\infty} 2^{-n} = 0$ zuordnen. Formal können wir hier nicht mit einer Gleichverteilung argumentieren, weil $|\Omega| = \infty$. Bereits intuitiv sind Wahrscheinlichkeiten für andere Ereignisse sehr unklar.

A_2: „für jedes $M \in \mathbb{N}$ gibt es einen „Kopfrun“ der Länge M

$$= \{b \in \underbrace{\{0,1\}^{\mathbb{N}}}_{\Omega} \mid \forall M \in \mathbb{N} \exists k \in \mathbb{N} : b_k = b_{k+1} = \cdots = b_{k+M-1} = 1\}$$

Beachte, dass hier $\Omega = \{0,1\}^{\mathbb{N}}$ die Kardinalität $\aleph_1$, d.h. die Kardinalität der reellen Zahlen besitzt, die $\mathcal{P}(\Omega) = \{A \mid A \subset \Omega\}$ sogar noch größere Kardinalität besitzt und es daher nicht einfach ist, eine Wahrscheinlichkeit jeder Teilmenge (jedem Ereignis) zuzuordnen. Dies ist aber meist auch gar nicht beabsichtigt und wir beschränken uns auf „interessierende“ Ereignisse. (siehe unten Satz von Vitali I.14).

Definition I.2

Mit Ω werde die (nichtleere) Menge der möglichen Versuchsausgänge bezeichnet. Ein Teilmengensystem $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ heißt σ-Algebra *oder* Menge der interessierenden Ereignisse, *falls gilt*

(A_1) $\Omega \in \mathcal{F}$

(A_2) $A \in \mathcal{F} \implies A^C = \Omega \setminus A \in \mathcal{F}$

(A_3) $A_n \in \mathcal{F}, n \in \mathbb{N} \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$

Jedes Element von $\mathcal{F}$ nennen wir Ereignis. *Ein* Wahrscheinlichkeitsmaß *(W-Maß) ist eine Abbildung $P : \mathcal{F} \to [0,1]$ mit*

(B_1) $P(\Omega) = 1$

(B_2) *Für $A_n \in \mathcal{F}, n \in \mathbb{N}$, die paarweise disjunkt sind ($\forall i \neq j : A_i \cap A_j = \emptyset$) gilt $P(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$* σ-Additivität.

Ein W-Raum *ist ein Tripel* $(\Omega, \mathcal{F}, P)$ *mit* Ω *nichtleere Menge,* $\mathcal{F}$ σ*-Algebra über* Ω *und* P *W-Maß auf* $\mathcal{F}$.

Bemerkung Axiome $(B_1), (B_2)$ heißen Kolmogorovsche Axiome (Kolmogorov, 1933).

Lemma I.3
Für jede σ*-Algebra* $\mathcal{F}$ *gilt:*

(a) $\varnothing \in \mathcal{F}$

(b) $A_1, A_2 \in \mathcal{F} \implies A_1 \cup A_2, A_1 \cap A_2 \in \mathcal{F}$

(c) $A_n \in \mathcal{F}, n \in \mathbb{N} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$

Beweis: klar (beachte $A_1 \cap A_2 = (A_1^C \cup A_2^C)^C$)

Beispiel I.4 Die trivialen σ-Algebren sind $\{\varnothing, \Omega\}$ und $\mathcal{P}(\Omega)$.

Lemma I.5
Für jedes W-Maß P *auf* $\mathcal{F}$ *gilt*

(a) $P(\varnothing) = 0$

(b) $A, B \in \mathcal{F}, A \subset B \implies P(A) \leq P(B)$

(c) $A, B \in \mathcal{F} : P(A \cup B) = P(A) + P(B) - P(A \cap B)$

(d) $A_n \in \mathcal{F}, n \geq 1$ *beliebig* $\implies P(\bigcup_{n \geq 1} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$ *(*Subadditivität, Bonferroni-Ungleichung*)*

(e) $A_n \in \mathcal{F}, n \geq 1$ *mit* $A_n \uparrow A \in \mathcal{F}$ *(d.h.* $A_n \subseteq A_{n+1} \forall n$ *und* $\bigcup_{n \geq 1} A_n = A$*):*

$$P(A) = \lim_{n \to \infty} P(A_n) \quad \sigma\text{-Stetigkeit}$$

Beweis: (a)-(d) und die Umkehrung von (e) als Übung.

(e) Setze $B_1 := A_1, B_2 := A_2 \setminus A_1, B_3 = A_3 \setminus A_2, \ldots B_n = A_n \setminus A_{n-1}$. Dann sind die $(B_n)_{n \geq 1}$ paarweise disjunkt mit $\bigcup_{n \geq 1} B_n = \bigcup_{n \geq 1} A_n$. Daher gilt

$$\begin{aligned} P(A) &= P(\bigcup_{n \geq 1} A_n) = P(\bigcup_{n \geq 1} B_n) \overset{\sigma\text{-Add.}}{=} \sum_{n \geq 1} P(B_n) \\ &= \lim_{N \to \infty} \sum_{n=1}^{N} P(B_n) \overset{\bigcup_{n=1}^{N} B_n = A_N}{=} \lim_{N \to \infty} P(A_N) \end{aligned}$$

□

Häufig interessieren uns nicht die Versuchsausgänge selbst, sondern begleitete Größen, z.B. nur die Augensumme beim Würfeln mit zwei Würfeln. In dem Beispiel würde man also modellieren:

$$\Omega = \{1, \ldots, 6\}^2, \quad \mathcal{F} = \mathcal{P}(\Omega), \quad P(A) = \frac{|A|}{36} \forall A \in \mathcal{F}, A \subseteq \Omega \qquad \text{(das ist ein W-Raum!)}$$

und Augensumme $S(\omega) = S(\omega_1, \omega_2) = \omega_1 + \omega_2, \omega \in \Omega$.

Wie groß ist die Wahrscheinlichkeit dafür, dass die Augensumme gleich k ist für k aus der Menge $\{2,3,\ldots,12\}$? $S(\omega)=k \iff \omega\in S^{-1}(\{k\})$. Also

$$P(\{\omega\in\Omega \mid S(\omega)=k\}) = P(S^{-1}(\{k\})) = \begin{cases} \frac{1}{36} & k=2\\ \frac{2}{36} & k=3\\ \frac{3}{36} & k=4\\ \frac{4}{36} & k=5\\ \frac{5}{36} & k=6\\ \frac{6}{36} & k=7\\ \frac{5}{36} & k=8\\ \vdots \end{cases}$$

Wir haben mittels S einen neuen W-Raum konstruiert! $\tilde\Omega = \{2,\ldots,12\}, \tilde{\mathcal F} := \mathcal P(\tilde\Omega)$ und $P^S(A) := P(S^{-1}(A)), A\subset\tilde\Omega$. In der Maßtheorie heißt P^X *Bildmaß* von X unter P.

Definition I.6
Es sei $(\Omega,\mathcal F,P)$ ein W-Raum und $(S,\mathcal S)$ ein Messraum. Dann heißt eine Abbildung $X:\Omega\to S$ S-wertige Zufallsvariable, *falls X messbar bezüglich $(\mathcal F,\mathcal S)$ ist (d.h. $\forall A\in\mathcal S: X^{-1}(A)\in\mathcal F$). Im Fall $(S,\mathcal S)=(\mathbb R,\mathcal B)$ (Borel-σ-Algebra) spricht man bloß von reellwertigen Zufallsvariable oder* Zufallsgröße, *im Fall $(S,\mathcal S)=(\mathbb R^d,\mathcal B_{\mathbb R^d})$ von einem* Zufallsvektor. *Das W-Maß (check!) $P^X(A) := P(X^{-1}(A)), A\in\mathcal S$ heißt* Verteilung *der Zufallsvariablen X.*

Man schreibt statt P^X auch $\mathcal L(X)$ („law of X") sowie

$$\{X\in A\} = \{\omega\in\Omega \mid X(\omega)\in A\} = X^{-1}(A),\ \{X=x\} = \{\omega\in\Omega\mid X(\omega)=x\} = X^{-1}(\{x\})$$
$$P(X\in A) = P(\{X\in A\}) = P(X^{-1}(A)) = P^X(A);\ P(X=x) = P(\{X=x\}) = P^X(\{x\}).$$

Beispiel I.7 (a) Ist $\mathcal F=\mathcal P(\Omega)$, so ist *jede* Abbildung $X:\Omega\to S$ messbar, also eine Zufallsvariable. Das gilt insbesondere also für die Augensumme S von oben sowie diskrete W-Räume (siehe unten).

(b) Für ein Ereignis $A\in\mathcal F$ ist $1_A(\omega) = \begin{cases}1, & \text{falls } \omega\in A\\ 0, & \text{falls } \omega\notin A\end{cases}$ (*Indikatorfunktion*) eine Zufallsvariable auf $(\Omega,\mathcal F,P)$ mit Werten in $(\{0,1\},\mathcal P(\{0,1\}))$ oder auch in $(\mathbb R,\mathcal B_{\mathbb R})$. Insbesondere ist jede konstante Funktion $X(\omega)=c$ für $c\in\mathbb R^d$ eine Zufallsvariable.

Beispiel I.8 (Geburtstagsproblem) Wie groß ist die Wahrscheinlichkeit, dass n ($n\le 365$) Personen an n verschiedenen Tagen Geburtstag haben? Wir gehen davon aus, dass das Jahr genau 365 Tage hat und jeder Tag mit gleicher Wahrscheinlichkeit Geburtstag einer Person ist (und diese Geburtstage unabhängig für alle Personen gewählt werden).

Die Menge aller Geburtstagskombinationen kann als Ziehen von n Kugeln einer Urne mit 365 Kugeln mit Zurücklegen und mit Beachtung der Reihenfolge modelliert werden. Es ergeben sich 365^n Kombinationen. Die „günstigen" Ereignisse (an verschiedenen Tagen Geburtstag) entsprechen dem Ziehen ohne Zurücklegen. Dafür ergeben sich $365\cdot 364\cdots(365-n+1)$ Kombinationen. Die gesuchte Wahrscheinlichkeit ist also

$$\frac{365!}{365^n(365-n)!} = 1\frac{364}{365}\cdots\frac{365-n+1}{365} = 1\cdot(1-\tfrac{1}{365})\cdots(1-\tfrac{n-1}{365}) = \exp\left(\sum_{k=1}^{n-1}\log\left(1-\tfrac{k}{365}\right)\right)$$

Für kleine x gilt $\log(1+x) \approx x$, stets gilt $\log(1+x) \leq x \lessapprox \exp(\sum_{k=1}^{n} -\frac{k}{365}) = \exp(-n(n-1)/2 \cdot 365)$

I.2. Diskrete Verteilungen

Definition I.9
Ist Ω eine endliche oder abzählbar unendliche Menge $\mathcal{F} = \mathcal{P}(\Omega)$ und ein W-Maß auf $\mathcal{F}$, so heißt $(\Omega, \mathcal{F}, P)$ diskreter W-Raum. *Eine S-wertige Zufallsvariable heißt* diskret verteilt, *falls $(S, \mathcal{S} = \mathcal{P}(S), P^S)$ ein diskreter W-Raum ist.*

Beispiel I.10 n-facher Münzwurf, Würfelwurf, Geburtstagsproblem, aber *nicht* der beliebig häufige Münzwurf.

Lemma I.11
(a) Ist $(\Omega, \mathcal{F}, P)$ diskreter W-Raum, so ist P eindeutig durch seine Zähldichte

$$p(\omega) := P(\{\omega\}), \omega \in \Omega$$

festgelegt. Ebenso legt bei einer diskret verteilten Zufallsvariable X (mit Werten in S)

$$p^X(s) = P^X(\{s\}) = P(X = s)$$

die Verteilung P^X von X eindeutig fest.

(b) Ist Ω andererseits endlich oder abzählbar unendlich und $p : \Omega \to [0,1]$, so dass $\sum_{\omega \in \Omega} p(\omega) = 1$ gilt, dann wird durch $P(A) := \sum_{\omega \in A} p(\omega)$, $A \subset \Omega$ ein W-Maß P auf $(\Omega, \mathcal{P}(\Omega))$ definiert, dessen Zähldichte p ist.

Beweis: (a) Seien P, Q W-Maße auf $(\Omega, \mathcal{F})$ mit $p(\omega) = P(\{\omega\}) = Q(\{\omega\})$ für alle $\omega \in \Omega$. Wegen σ-Additivität gilt dann für alle $A \subseteq \Omega$:

$$\begin{aligned} P(A) = P(\bigcup_{\omega \in A} \{\omega\}) &\overset{\sigma\text{-Add.}}{=} \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p(\omega) = \sum_{\omega \in A} Q(\{\omega\}) \\ &= Q(\bigcup_{\omega \in A} \{\omega\}) = Q(A) \implies P = Q \end{aligned}$$

Genauso für Zufallsvariablen.

(b) Es gilt $P(\Omega) \overset{\text{Def}}{=} \sum_{\omega \in \Omega} p(\omega) \overset{\text{Vor.}}{=} 1$ sowie für $A_n \subseteq \Omega$ paarweise disjunkt

$$\begin{aligned} P(\bigcup_{n \geq 1} A_n) &\overset{\text{Def}}{=} \sum_{\omega \in \bigcup_{n \geq 1} A_n} p(\omega) = \sum_{n \geq 1} \sum_{\omega \in A_n} p(\omega) \\ &\overset{\text{Def}}{=} \sum_{n \geq 1} P(A_n) \implies P \quad \text{ist } \sigma\text{-additiv} \end{aligned}$$

Also ist P ein W-Maß. Natürlich gilt $P(\{\omega\}) = p(\omega)$ nach Definition. □

I.3. Wichtige diskrete Verteilungen

Gleichverteilung/Laplaceverteilung Ist Ω endlich, so beschreibt die Zähldichte $p(\omega) = \frac{1}{|\Omega|}, \omega \in \Omega$, die Gleichverteilung auf Ω. Es gilt also

$$P(A) = \sum_{\omega \in A} p(\omega) = \frac{|A|}{|\Omega|}, A \subseteq \Omega$$

Das hatten wir bereits bei Münzwurf und Würfelwurf verwendet.

Hypergeometrische Verteilung

Bernoulli-Schema/Bernoulli-Kette : Wir spielen ein Spiel n-mal hintereinander, bei dem jeweils mit Wahrscheinlichkeit p Erfolg eintritt und die einzelnen Spielausgänge nicht voneinander abhängen. („Münzwurf“). Setze „1“ für Erfolg und „0“ für Misserfolg und modelliere $\Omega = \{0,1\}^n$ mit Zähldichte

$$\bar{p}(\omega) = \bar{p}(\omega_1, \ldots, \omega_n) = p^{\sum_{i=1}^n \omega_i}(1-p)^{\sum_{i=1}^n (1-\omega_i)}$$

Man spricht von einem Bernoulli-Schema der Länge n mit Erfolgswahrscheinlichkeit p.

Binomialverteilung: Es bezeichne N die Anzahl der Erfolge in einem Bernoulli-Schema der Länge n mit Erfolgswahrscheinlichkeit p. N ist also die Zufallsvariable

$$N : \Omega^{\text{Bernoulli}} \to \{0,1,\ldots,n\} \quad \text{mit} \quad N(\omega) = N(\omega_1,\ldots,\omega_n) = \sum_{i=1}^n \omega_i$$

Was ist die Verteilung P^N von N? Betrachte die Zähldichte p^N.

$$p^N(k) = P(N = k) = \binom{n}{k} \cdot p^k(1-p)^{n-k} =: \mathrm{Bin}_{n,p}(k)$$

Die durch $\mathrm{Bin}_{n,p}(\cdot)$ beschriebene Verteilung heißt *Binomialverteilung* der Länge n mit Wahrscheinlichkeit p. Beispiele sind „Anzahl Kopf“ beim Münzwurf, Gewinne beim regelmäßigen Lotto, Heilungserfolg eines Medikaments bei Patienten etc.

$\mathrm{Bin}_{n,p}$ kann durch ein sogenanntes Stabdiagramm visualisiert werden.

Geometrische Verteilung Es bezeichne K den Zeitpunkt des ersten Erfolgs in einem Bernoulli-Schema. Für $k \leq n$ gilt

$$P(K = k) = (1-p)^{k-1}p =: \mathrm{Geo}_p(k)$$

Man kann die Zähldichte $\mathrm{Geo}_p(\cdot)$ der geometrischen Verteilung mit Wahrscheinlichkeit p auf $\mathbb{N} = \{1,2,\ldots\}$ als Zähldichte der Zufallsvariablen K bei einem beliebig langen Bernoulli-Schema ($n = \infty$) auffassen. Beachte

$$\sum_{k=1}^\infty \mathrm{Geo}_p(k) = \sum_{k=1}^\infty (1-p)^{k-1} \cdot p = p \cdot \frac{1}{1-(1-p)} = 1 \quad \text{für } p \in (0,1]$$

Beispielanwendung ist z.B. die Anzahl von Tagen bis zum Ausfall eines technischen Geräts („Glühbirne“).

Multinomialverteilung : Oft gibt es nicht nur zwei Versuchsausgänge (Erfolg, Misserfolg), sondern $r > 2$. In der Genetik werden z.B. weiß-, rosa- und rotblühende Ausprägungen einer Pflanze modelliert und gezählt. Auf der Menge $\Omega = \{k \in \{0,\dots,n\}^r \mid \sum_{i=1}^r k_i := n\}$, wobei $k = (k_1,\dots,k_r)$ ist und k_i die Anzahl der Versuchsausgänge mit Merkmal $i \in \{1,\dots,r\}$ (oder in Klasse i) bezeichnet, betrachte die Zähldichte

$$\mathrm{Mult}_{n,p_1,\dots,p_r}(k) = \underbrace{\frac{n!}{k_1!k_2!\cdots k_r!}}_{\text{Multinomialkoeffizient}} p_1^{k_1}\cdot p_2^{k_2}\cdots p_r^{k_r} \quad \text{für} \quad p_i \geq 0, \sum_{i=1}^r p_i = 1$$

p_i gibt dabei die Wahrscheinlichkeit für Merkmal/Klasse i an. Bei zwei Klassen, d.h. $r = 2$, ergibt sich

$$\mathrm{Mult}_{n,p_1,p_2}(k) = \frac{n!}{k_1!k_2!}p_1^{k_1}p_2^{k_2} = \frac{n!}{k_1!(n-k_1)!}p_1^{k_1}\cdot(1-p_1)^{n-k_1}$$

dann in der ersten Koordinate $\mathrm{Bin}_{n,p_1}(k_1)$

Poissonverteilung Auf $\Omega = \mathbb{N}_0$ definiert für $\lambda > 0$

$$\mathrm{Poiss}_\lambda(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k \geq 0$$

die Zähldichte der Poissonverteilung. Die Wichtigkeit ergibt sich aus dem *Poissonschen Grenzwertsatz:*

Satz I.12 (Poissonscher Grenzwertsatz)
Es seien $p_n \in [0,1]$ mit $\lim_{n\to\infty} np_n = \lambda > 0$. Dann gilt für alle $k \in \mathbb{N}_0$

$$\lim_{n\to\infty} \mathrm{Bin}_{n,p_n}(k) = \mathrm{Poiss}_\lambda(k)$$

Beweis: Schreibe $A_n \sim B_n$ falls $\lim_{n\to\infty}\frac{A_n}{B_n} = 1$. Dann gilt

$$\begin{aligned}
\binom{n}{k} &= \frac{n(n-1)\cdots(n-k+1)}{n^k}\cdot\frac{n^k}{k!} \sim \frac{n^k}{k!}\\
\Longrightarrow \mathrm{Bin}_{n,p_n}(k) &= \binom{n}{k}p_n^k(1-p_n)^{n-k} \sim \frac{n^k}{k!}p_n^k(1-p_n)^{n-k}\\
&= \frac{n^k}{k!}p_n^k\cdot\frac{(1-p_n)^{\frac{1}{p_n}\cdot np_n}}{(1-p_n)^k} \overset{\text{Analysis}}{\sim} \frac{(np_n)^k}{k!}e^{-np_n}\\
&\sim \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}$$

□

Bemerkung Später werden wir mit stoch. Methoden folgende Konvergenzgeschwindigkeit erhalten:

$$\sum_{k=0}^n |\mathrm{Bin}_{n,p}(k) - \mathrm{Poiss}_\lambda(k)| \leq 2np^2 = 2\lambda p \quad \text{für } \lambda = np$$

Auf Grund dieses Grenzwertsatzes heißt die Poissonverteilung auch „Verteilung der seltenen Ereignisse“

Beispiel I.13 (a) In einem Land gibt es 730 Geburten pro 100.000 Einwohnern im Jahr. Wie ist die Anzahl X der Geburten pro Tag in einer Stadt mit 280.000 Einwohnern verteilt? (wichtig z.B. für Geburtsstation im Krankenhaus.) Approximative Modellierung

p: „Wahrscheinlichkeit, dass ein Einwohner an einem festen Tag ein Kind bekommt.“ $p = \frac{730}{100\,000 \cdot 365} = \frac{2}{100\,000}$

$n = 280\,000$ unabhängige Wiederholungen dieses „Experiments“

n groß, p klein, das heißt Poisson-Approximation mit $\lambda = n \cdot p = 5.6$. Die Zufallsvariable X ist also approximativ $\text{Poiss}(5,6)$ verteilt, d.h. $P(X = k) = \text{Poiss}_{5,6}(k)$ mit $k \in \mathbb{N}_0$.

(b) Radioaktiver Zerfall: Anzahl N der Zerfälle in einer festen Zeiteinheit wird kanonisch mit der Poissonverteilung modelliert (Atomzahl n groß, Wkeit für Zerfall eines Atoms sehr klein).

Folgendes Beispiel zeigt nun, dass wir bei überabzählbaren Mengen Ω nicht mehr vernünftig definieren können, weshalb wir uns dann auf σ-Algebren (wie Borel-σ-Algebra) beschränken werden.

Satz I.14 (Vitali, 1905)
Sei $\Omega = \{0,1\}^{\mathbb{N}}$ die Ergebnismenge des beliebig langen Münzwurfs. Dann gibt es keine *Abbildung $P : \mathcal{P}(\Omega) \to [0,1]$, die den Kolmogorovschen Axiomen (eines W-Maßes) genügt und folgende Invarianzeigenschaften besitzt:*

$$\forall A \subset \Omega, n \geq 1 : P(T_n(A)) = P(A)$$
$$\textit{wobei} \quad T_n(\omega) = T_n((\omega_1, \omega_2, \dots, \omega_n, \dots)) = (\omega_1, \omega_2, \dots, \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, \dots)$$

das Ergebnis des n-ten Wurfes umkehrt.

Beweis: Definiere Äquivalenzrelation $\sim$ auf Ω: $\omega \sim \omega' \iff \exists N \geq 1 \forall n \geq N : \omega_n = \omega'_n$ (ω und ω' unterscheiden sich nur in endlich vielen Gliedern). Nach dem Auswahlaxiom existiert eine Menge $A \subseteq \Omega$, die aus jeder Äquivalenzklasse genau einen Repräsentanten enthält. Nun sei

$$\mathcal{S} = \{S \subseteq \mathbb{N} \mid |S| \text{ endlich}\} = \bigcup_{m \geq 1} \underbrace{\{S \subset \mathbb{N} \mid \max S = m\}}_{\text{endliche Kardinalität}}$$

so dass $\mathcal{S}$ abzählbar ist. Setze $T_S := \prod_{n \in S} T_n = T_{n_1} \circ T_{n_2} \circ \dots \circ T_{n_k}$ für $S = \{n_1, n_2, \dots, n_k\}$ (T_S diskret, dreht also die Ergebnisse an allen Indizes in S um). Dann gilt:

(a) $\Omega = \bigcup_{S \in \mathcal{S}} T_S(A)$ (zu jedem $\omega \in \Omega$ existiert $\omega' \in A$ mit $\omega' \sim \omega$, also auch ein $S \subseteq \mathbb{N}$, $|S| < \infty$ mit $T_S(\omega') = \omega$)

(b) Die Mengen $(T_S(A))_{S \in \mathcal{S}}$ sind paarweise disjunkt, da sonst für $\omega, \omega' \in A$ und $S, S' \in \mathcal{S}$

$$T_S(\omega) = T_{S'}(\omega') \implies \omega \sim T_S(\omega) = T_{S'}(\omega') \sim \omega' \quad \text{aber} \quad \omega, \omega' \in A \implies \omega = \omega'$$

σ-Additivität zeigt: $1 = P(\Omega) = P(\bigcup_{S \in \mathcal{S}} T_S(A)) = \sum_{S \in \mathcal{S}} P(T_S(A)) \overset{\text{Ann.}}{=} \sum_{S \in \mathcal{S}} P(A) \in \{0, \infty\}$ ↯

Also kann ein solches P nicht existieren. □

I.4. Maßtheorie und W-Maße in $\mathbb{R}^d$

Zunächst rekapitulieren wir die Grundlagen der Maßtheorie.

- Für jedes $\mathrm{E} \subset \mathcal{P}(\Omega)$ gibt es eine kleinste σ-Algebra $\sigma(\mathrm{E})$, die E enthält.
- Ist (S,d) ein metrischer Raum (wie $(\mathbb{R}^d, |\cdot|)$), so heißt $\mathcal{B}_S = \sigma(\{\mathcal{O} \subseteq S \mid \mathcal{O} \text{ offen}\})$ *Borel-σ-Algebra* von S.
- $\mathcal{B}_\mathbb{R}$ wird z.B. erzeugt von $\mathrm{E} = \{(a,b] \mid a,b \in \mathbb{R}\}$ oder $\mathcal{F} = \{(-\infty,b] \mid b \in \mathbb{R}\}$. $\mathcal{B}_{\mathbb{R}^d}$ analog durch $\mathrm{E} = \{(a_1,b_1] + \ldots + (a_d,b_d] \mid a_1,\ldots,a_d,b_1,\ldots b_d \in \mathbb{R}\}$ (Quader, Rechtecke)
- $g : \Omega \to S$ ist $(\mathcal{F} \subseteq \mathcal{P}(\Omega), \mathcal{S} \subseteq \mathcal{P}(S))$-messbar, wenn für einen Erzeuger E von $\mathcal{S}$ gilt:
$$\forall A \in \mathrm{E} : g^{-1}(A) \in \mathcal{F}$$
- jede stetige Funktion zwischen metrischen Räumen ist Borel-messbar (d.h. messbar bzgl. der jeweiligen Borel-σ-Algebren).
- Jede Funktion $g : \Omega \to \mathbb{R}$ mit $\{g \leq y\} \in \mathcal{F}$ für alle $y \in \mathbb{R}$ ist $(\mathcal{F}, \mathcal{B}_\mathbb{R})$-messbar.
- Falls $g_n : \Omega \to \mathbb{R}$ $(\mathcal{F}, \mathcal{B}_\mathbb{R})$-messbar sind, so auch
$$\inf_n g_n, \sup_n g_n, \liminf_{n\to\infty} g_n, \limsup_{n\to\infty} g_n, \lim_{n\to\infty} g_n$$
(sofern existent)
- $g_1,\ldots,g_d : \Omega \to \mathbb{R}$, $(\mathcal{F}, \mathcal{B}_\mathbb{R})$-messbar, dann ist $\omega \mapsto (h(g_1(\omega),\ldots,g_d(\omega))$ mit $h : \mathbb{R}^d \to \mathbb{R}^k$ Borelmessbar ist $(\mathcal{F}, \mathcal{B}_{\mathbb{R}^k})$-messbar. Insbesondere gilt:
$$(g_1,\ldots,g_d), g_1 + g_2, g_1 - g_2, g_1 \cdot g_2, g_1/g_2, \max(g_1,\ldots,g_d), \min(g_1,\ldots,g_d)$$
sind dann messbar. [3]
- $g : \Omega \to S$, $(\mathcal{F}, \mathcal{S})$-messbar, $h : S \to T$ $(\mathcal{S}, \mathcal{T})$-messbar $\Longrightarrow$ $h \circ g : \Omega \to T$ ist $(\mathcal{F}, \mathcal{T})$-messbar.

Definition I.15
Sei Ω nichtleere Menge. Dann heißt $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ Algebra *über Ω, falls*

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Longrightarrow A^C \in \mathcal{A}$
- $A, B \in \mathcal{A} \Longrightarrow A \cup B \in \mathcal{A}$.

$\mu : \mathcal{A} \to [0,+\infty]$ *heißt* Prämaß *über $\mathcal{A}$, falls*

- $\mu(\varnothing) = 0$
- $\forall A_n \in \mathcal{A}$ *paarweise disjunkt mit* $\bigcup_{n\geq 1} A_n \in \mathcal{A} : \mu(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} \mu(A_n)$.

μ heißt Maß, *falls $\mathcal{A}$ bereits eine σ-Algebra ist. Ein Maß μ heißt σ-*endlich, *falls es $A_n \in \mathcal{A}$ gibt mit $\mu(A_n) < \infty$ und $\bigcup_n A_n = \Omega$. Ein Maß μ mit $\mu(\Omega) = 1$ ist ein W-Maß (s.o.).*

[3] g_1/g_2 nur falls sie wohldefiniert ist.

Beispiel I.16 Auf $\Omega = \mathbb{R}$ ist $\mathcal{A} := \{\bigcup_{k=1}^n (a_k, b_k] \mid n \geq 1, a_k \in \mathbb{R} \cup \{-\infty\}, b_k \in \mathbb{R} \cup \{\infty\}\}$ eine Algebra (wobei $(a_k, \infty] = (a_k, \infty)$. Setze $\lambda((a,b]) := b - a$ für $b \geq a$. Für paarweise disjunkte $(a_k, b_k] : \lambda(\bigcup_{k=1}^n (a_k, b_k]) = \sum_{k=1}^n (b_k - a_k)$. Unten werden wir zeigen, dass dieses λ ein Prämaß auf $\mathcal{A}$ definiert. Mit Hilfe des folgenden Satzes wird dies zu einem Maß auf $\mathcal{B}_\mathbb{R}$ fortgesetzt, dem *Lebesguemaß*:

Satz I.17 (von Caratheodory, 1917)
Jedes Prämaß μ auf $\mathcal{A}$ kann zu einem Maß $\tilde{\mu}$ auf $\sigma(\mathcal{A})$ fortgesetzt werden, d.h. $\tilde{\mu}(A) = \mu(A)$ für alle $A \in \mathcal{A}$.

Eindeutigkeitssatz: Es seien μ, ν σ-endliche Maße auf $(\Omega, \mathcal{F})$ mit $\exists (A_n) \in \mathcal{F} : \bigcup A_n = \Omega, \mu(A_n) = \nu(A_n) < \infty$. Ist dann E ein Erzeuger von $\mathcal{F}$, der $\cap$-stabil (d.h. $A, B \in \mathrm{E} \implies A \cap B \in \mathrm{E}$) und gilt $\mu(A) = \nu(A)$ für alle E, so folgt $\mu = \nu$ auf $\mathcal{F}$. (Insbesondere gilt Eindeutigkeit der Maßerweiterung bei W-Maßen).

Definition I.18
Für ein W-Maß P auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ ist seine Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ *gegeben durch $F(x) := P((-\infty, x])$. Für reellwertige Zufallsvariablen X setzt man entsprechend $F^X(x) = P^X((-\infty, x]) = P(X \leq x), x \in \mathbb{R}$.*

Lemma I.19
Jede Verteilungsfunktion F ist monoton wachsend, rechtsstetig und erfüllt $\lim_{x\to-\infty} F(x) = 0$ *und* $\lim_{x\to\infty} F(x) = 1$.

Beweis: Monotonie für $x < y$ gilt

$$F(y) - F(x) = P((-\infty, y]) - P((-\infty, x]) \overset{\text{Add. v. } P}{=} P((x,y]) \geq 0$$

Rechtsstetigkeit Es gelte $x_n \downarrow x$ ($x_n \to x, x_n \geq x, x_n$ monoton fallend) $F(x_n) - F(x) \overset{\text{s.o.}}{=} P((x, x_n])$. Nun gilt

$$(x, x_n] \supseteq (x, x_{n+1}] \supseteq \cdots \forall n \geq 1 \text{ mit } \bigcap (x, x_n] = \varnothing \text{ und } \lim_{n\to\infty} (x, x_n] = \varnothing$$

$$\overset{\sigma\text{-Stetigkeit}}{\Longrightarrow} \lim_{n\to\infty} P((x, x_n]) = P(\lim_{n\to\infty} (x, x_n]) = P(\varnothing) = 0$$

$$\Longrightarrow \lim_{n\to\infty} F(x_n) = F(x) \quad (F \text{ ist rechtsstetig})$$

σ-Stetigkeit: $A_n = (x, x_n]^\mathsf{C}$, $A_n \subseteq A_{n+1} \forall n, \bigcup_{n\geq 1} A_n = \mathbb{R} \overset{\text{s.o.}}{\Longrightarrow} P(A_n) \to P(\mathbb{R}) = 1$
Dann ist $P((x, x_n]) = P(A_n^\mathsf{C}) = 1 - P(A_n) \xrightarrow{n\to\infty} 0$.

$$\lim_{x\to-\infty} F(x) = \lim_{x\downarrow-\infty} P((-\infty, x]) \overset{\sigma-\text{Stet.}}{=} P(\lim_{x\downarrow-\infty} (-\infty, x]) = P(\varnothing) = 0$$

$$\lim_{x\to\infty} F(x) = \lim_{x\uparrow\infty} P((-\infty, x]) = P(\mathbb{R}) = 1$$

□

Satz I.20
Es sei $F : \mathbb{R} \to \mathbb{R}$ eine monoton wachsende, rechtsstetige Funktion. Dann existiert ein Maß μ auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ mit

$$\mu((a,b]) = F(b) - F(a) \quad \text{für alle } a < b \in \mathbb{R}$$

μ ist eindeutig durch F definiert und heißt Lebesgue-Stieltjes-Maß *zu F.*

Korollar I.21
Es gibt genau ein Maß λ auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ mit $\lambda((a,b]) = b - a$ ($\forall a < b \in \mathbb{R}$). λ heißt Lebesguemaß *(wähle $F(x) = x$)*

Korollar I.22
Zu jedem $F : \mathbb{R} \to [0,1]$ monoton wachsend, rechtsstetig und $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$ gibt es genau ein W-Maß P, dessen Verteilungsfunktion F ist.

Beispiel I.23 Für alle Intervalle berechnet man dann leicht $P((a,b]) = F(b) - F(a)$.

Beweis: (des Satzes) Eindeutigkeit folgt, weil die Menge $\{(a,b] \mid a < b \in \mathbb{R}\}$ ein $\cap$-stabiler Erzeuger von $\mathcal{B}_\mathbb{R}$ ist und $\bigcup_{n\geq 1}[-n,n] = \mathbb{R}$ gilt, aus dem Eindeutigkeitssatz. Auf der Algebra (s.o) $\mathcal{A} = \{\bigcup_{k=1}^K (a_k, b_k] \mid K \geq 1, -\infty \leq a_1 < b_1 \leq \dots \leq a_k < b_k \leq \infty\}$ definiere

$$\mu(\bigcup_{k=1}^K (a_k, b_k]) := \sum_{k=1}^K (F(b_k) - F(a_k))$$

Dann ist μ offensichtlich additiv mit $\mu(\emptyset) = 0$.

Zu zeigen ist σ-Additivität für μ Prämaß: $A_n = \bigcup_{k=1}^{K_n}(a_k^n, b_k^n]$ mit A_n paarweise disjunkt und $\bigcup_{n\geq 1} A_n \in \mathcal{A}$, d.h.

$$\bigcup_{n\geq 1} A_n = \bigcup_{k=1}^{K_\infty}(a_k^\infty, b_k^\infty], \quad K_\infty \in \mathbb{N}, \quad -\infty \leq a_1^\infty \leq b_1^\infty \leq \ldots \leq a_{K^\infty}^\infty \leq b_{K^\infty}^\infty \leq \infty$$

$$\implies \sum_{K\geq 1}\sum_{k=1}^{K_n}(F(b_k^n) - F(a_k^n)) \stackrel{!}{=} \sum_{k=1}^{K_\infty}(F(b_k^\infty) - F(a_k^\infty))$$

Dafür reicht es, für ein Intervall zu zeigen:

$$(a^\infty, b^\infty] = \bigcup_{l\geq 1}(a_l, b_l] \implies \mu((a^\infty, b^\infty]) = \sum_{l>1}\mu((a_l, b_l])$$

Natürlich gilt $\mu(\bigcup_{l=1}^L (a_l, b_l]) \leq \mu((a^\infty, b^\infty])$, also Subadditivität von μ, da $\bigcup_{l=1}^L (a_l, b_l] \subseteq (a^\infty, b^\infty])$

$$\forall L \geq 1 \sum_{l=1}^L \mu((a_l, b_l]) \leq \mu((a^\infty, b^\infty])$$
$$\implies \sum_{l=1}^\infty \mu((a_l, b_l]) \leq \mu((a^\infty, b^\infty])$$

Für „$\geq$" betrachte $\mathcal{O}_l = (a_l, b_l + \delta_l) \supset (a_l, b_l]$ mit $\mu((a_l, b_l + \delta_l]) \leq \mu((a_l, b_l]) + \varepsilon \cdot 2^{-l}$, was wegen Rechtsstetigkeit von F immer möglich ist. Dann ist $(\mathcal{O}_l)_{l\geq 1}$ eine offene Überdeckung von $[a^\infty + \delta^\infty, b^\infty]$, wobei $\mu((a^\infty, a^\infty + \delta^\infty]) \leq \varepsilon$ gelte (F rechtsstetig!).

Nach dem Satz von Heine-Borel gilt wegen Kompaktheit von $[a^\infty + \delta^\infty, b^\infty]$, dass es ein $L \geq 1$ existiert mit $[a^\infty + \delta^\infty, b^\infty] \subseteq \bigcup_{l=1}^L (a_l, b_l + \delta_l]$.

$$\implies \mu((a^\infty, b^\infty]) = \mu((a^\infty + \delta^\infty, b^\infty]) + \mu((a^\infty, a^\infty + \delta^\infty]) \leq \sum_{l=1}^L (\mu((a_l, b_l]) + \varepsilon \cdot 2^{-l}) + \varepsilon$$
$$\leq \sum_{l\geq 1}\mu((a_l, b_l]) + 2\varepsilon$$

Dies gilt für beliebiges $\varepsilon > 0$. Mit $\varepsilon \to 0$ folgt $\sum_{l\geq 1}\mu((a_l, b_l]) \geq \mu((a^\infty, b^\infty])$. □

Definition I.24
Ist $f : \mathbb{R}^d \to [0,\infty)$ *eine (Lebesgue)-integrierbare Funktion mit* $\int_{\mathbb{R}^d} f(x)\,dx = 1$*, so heißt* f Lebesgue-Dichte *oder auch* (Wahrscheinlichkeits)-Dichte *auf* $\mathbb{R}^d$*.*

Lemma I.25
Jede W-Dichte f *auf* $\mathbb{R}$ *erzeugt mittels* $P_f((a,b]) := \int_a^b f(x)\,dx$ *ein W-Maß* P_f *auf* $(\mathbb{R},\mathcal{B}_\mathbb{R})$*.*

Beweis: Betrachte $F(x) = \int_{-\infty}^x f(y)\,dy$. Dann ist F monoton wachsend, stetig und erfüllt

$$\lim_{x\to-\infty} F(x) = 0, \qquad \lim_{x\to\infty} F(x) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Der Satz/Korollar liefert, dass F die Verteilungsfunktion eines W-Maßes P_f ist, so dass gilt $P_f((a,b]) = F(b) - F(a) = \int_a^b f(x)\,dx$. □

Bemerkung

(a) Aus der Integrationstheorie folgt dann für jede Borelmenge $\mathcal{B}$, dass

$$P_f(\mathcal{B}) = \int_{\mathcal{B}} f(x)\,dx$$

(b) Man kann Dichten auch bezüglich beliebiger σ-endlicher Maße definieren, nicht nur bzgl. des Lebesguemaßes.

Lemma I.26

(a) Ist f *die Dichte eines W-Maßes* P *auf* $(\mathbb{R},\mathcal{B}_\mathbb{R})$ *mit Verteilungsfunktion* F*, so gilt* $\forall x \in \mathbb{R}$: $F(x) = \int_{-\infty}^x f(s)\,ds$

(b) Ist die Verteilungsfunktion F *auf* $\mathbb{R}$ *(schwach) differenzierbar, so ist* $f(x) = F'(x), x \in \mathbb{R}$*, die zugehörige W-Dichte.*

Bemerkung F heißt *schwach differenzierbar* (oder *absolut stetig*), falls es eine Funktion F gibt mit der Eigenschaft $F(b) - F(a) = \int_a^b F'(y)\,dy$ für alle $a < b$.

Beispiel I.27 $F(x) = |x|$ ist schwach differenzierbar mit $F'(x) = \mathbf{1}_{(0,\infty)}(x) - \mathbf{1}_{(-\infty,0)}(x)$.

Beispiel I.28 (a) W-Dichte $f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$, $x \in \mathbb{R}$ für $a < b$. Dann ist $P_f([a,b]) = 1$ und $P_f = \frac{1}{b-a} \cdot \lambda\big|_{[a,b]}$ (das Lebesguemaß auf $[a,b]$). P_f heißt *gleichmäßige Verteilung* auf $[a,b]$, Notation $\mathrm{U}([a,b])$ für „uniform".

(b) *Exponentialverteilung* für $\lambda > 0$ setze $f_\lambda(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0,\infty)}(x)$, $x \in \mathbb{R}$ ist W-Dichte. Das zugehörige W-Maß P_λ heißt Exponentialverteilung mit Parameter λ, Notation $\mathrm{Exp}(\lambda)$. Die Verteilung $\mathrm{Exp}(\lambda)$ wird häufig zur Modellierung von Wartezeiten (Bedienzeiten, Atomzerfall etc) verwendet. λ entspricht z.B. Stärke der Radioaktivität (siehe Übung)

(c) *Normalverteilung* Parameter $\mu \in \mathbb{R}$, $\sigma > 0$. W-Dichte

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

Das zugehörige W-Maß heißt Normalverteilung, Notation $\mathcal{N}(\mu, \sigma^2)$. Zeige

$$\int_{-\infty}^{\infty} \varphi_{\mu,\sigma^2}(x)\,dx \overset{!}{=} 1$$

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \underset{\text{subst.}}{\overset{z=\frac{x-\mu}{\sigma}}{=}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\cancel{\sigma}} e^{-\frac{z^2}{2}} \cdot \cancel{\sigma}\,dz$$

$$\implies \int \varphi_{\mu,\sigma^2} = \int \varphi_{0,1} \quad \text{Reduktion auf Fall } \mu = 0, \sigma = 1$$

$$\left(\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz\right)^2 \overset{\text{Fubini}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \cdot e^{-\frac{y^2}{2}} \underbrace{dz\,dy}_{=d\lambda_{\mathbb{R}^2}} = \int_{\mathbb{R}^2} e^{-\frac{y^2+z^2}{2}}\,dy\,dz$$

$$\overset{\text{Polar}}{=} \int_{\mathbb{R}^+ \times [0,2\pi]} e^{-\frac{r^2}{2}} r\,dr\,d\varphi$$

$$= 2\pi \int_0^{\infty} re^{-\frac{r^2}{2}}\,dr = 2\pi \left(-e^{-\frac{r^2}{2}}\right)\Bigg|_{r=0}^{\infty} = 2\pi \implies \int \varphi_{0,1} = 1$$

(d) *Produktdichte* im $\mathbb{R}^d$. Sind $f_1, \dots, f_d : \mathbb{R} \to [0, \infty)$ W-Dichten in $\mathbb{R}$, so ist

$$f(x_1, \dots, x_d) = \prod_{i=1}^{d} f_i(x_i) \qquad (x_1, \dots, x_d) \in \mathbb{R}^d$$

eine W-Dichte in $\mathbb{R}^d$ die Produktdichte von $f_1, \dots, f_d$. Beachte

$$\int_{\mathbb{R}^d} f(x)\,dx \overset{\text{Fubini}}{=} \underbrace{\int_{-\infty}^{\infty}\left(\cdots \int_{-\infty}^{\infty}}_{d\text{-mal}} f_1(x_1)\cdots f_d(x_d)\,dx_1 \dots\right) dx_d$$

$$= \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(d-1)\text{-mal}} f_2(x_2)\cdots f_d(x_d)\,dx_2 \dots dx_d = \cdots = 1$$

(e) *Mehrdimensionale Standard-Normalverteilung:* $f(x) = (2\pi)^{-\frac{d}{2}} e^{-\frac{|x|^2}{2}}$, $x \in \mathbb{R}^d$ mit $|x|^2 = x_1^2 + \cdots + x_d^2$ ist Produktdichte der eindimensionalen Standard-Normalverteilungen $f_i(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}}$. Notation: $\mathcal{N}(\mathcal{O}, \mathrm{E}_d)$ mit $\mathrm{E}_d = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \in \mathbb{R}^{d \times d}$. Beachte: f ist rotationsinvariant.

Bemerkung (Resultate für W-Maße in $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$) Betrachte die d-dim. Verteilungsfunktion

$$F(x) = F(x_1, \dots, x_d) = P((-\infty, x_1] \times \cdots \times (-\infty, x_d])$$

F hat wiederum Eigenschaften wie im Eindimensionalen, nur die Monotonie wird ersetzt durch die Nichtnegativität eines komplexeren Ausdrucks.

Im $\mathbb{R}^2$ z.B. gilt $F(x_1, x_2) - F(x_1, y_2) - F(y_1, x_2) + F(y_1, y_2) \geq 0$, $\forall y_1 \leq x_1, y_2 \leq x_2$. Dann kann man umgekehrt zeigen, dass eine d-dim. Verteilungsfunktion (d.h. Funktion mit diesen

Eigenschaften) eindeutig ein W-Maß auf $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ bestimmt. Ist $f : \mathbb{R}^d \to [0,\infty)$ eine W-Dichte, so ist $F(x_1,\dots,x_d) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} f(z)\, dz_1 \dots dz_d$ eine Verteilungsfunktion, so dass es ein entsprechendes W-Maß P_f gibt mit $P_f(B) = \int_B f(x)\,dx$ für alle $B \in \mathcal{B}$.

Satz I.29 (Dichtetransformation)
Es sei X eine reellwertige Zufallsvariable mit stetiger Dichte f^X (d.h. f^X ist W-Dichte von P^X) sowie $Y = \varphi(X)$ mit $\varphi : \mathbb{R} \to \mathbb{R}$ monoton wachsend, so dass die Linksinverse φ^{-1} stetig differenzierbar ist. Dann hat Y ebenfalls eine Dichte und zwar $f^Y(y) = f^X(\varphi^{-1}(y)) \cdot (\varphi^{-1})'(y)$, $y \in \mathbb{R}$. Beachte: Für beliebiges φ muss Y keine Dichte besitzen, z.B. falls $\varphi(x) \equiv$const.

Beweis: Es gilt

$$\begin{aligned} F^Y(y) &= P^Y((-\infty,y]) = P(Y \le y) = P(\varphi(X) \le y) \\ &= P(\varphi^{-1}\varphi(X) \le \varphi^{-1}(y)) = P(X \le \varphi^{-1}(y)) \\ &= F^X(\varphi^{-1}(y)) \\ \Longrightarrow f^Y(y) &= (F^Y)'(y) = f^X(\varphi^{-1}(Y)) \cdot (\varphi^{-1})'(y) \end{aligned}$$

□

Korollar I.30
Im vorangegangen Satz gilt für $Y = aX + b$ mit $a \neq 0, b \in \mathbb{R}$

$$f^Y(y) = f^X\left(\frac{y-b}{a}\right)\frac{1}{a}$$

Beweis: klar wegen $\varphi^{-1}(y) = \frac{y-b}{a}$

Beispiel I.31 (a) $X \sim \mathcal{N}(0,1)$ d.h. $(P^X = N(0,1)) \implies Y = \sigma X + \mu \sim \mathcal{N}(\mu,\sigma^2)$.

$$f^Y(y) = \varphi_{0,1}(\tfrac{y-\mu}{\sigma}) \cdot \frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \cdot \frac{1}{\sigma} = \varphi_{\mu,\sigma^2}(y)$$

(b) $X \sim \text{Exp}(1) \overset{\lambda>0}{\Longrightarrow} X = \lambda^{-1} \cdot X \sim \text{Exp}(\lambda)$.

$$f^Y(y) = f^X(\lambda y) \cdot \lambda = \lambda \cdot e^{-\lambda y} \mathbf{1}_{\mathbb{R}^+}(y) \quad \text{(gilt trotz Sprungs von } f^X \text{ bei Null)}$$

(c) $X \sim \mathcal{N}(0,1) \implies Y = X^2 \sim \chi^2(1)$ (chi-quadrat-Verteilung mit einem Freiheitsgrad) Problem: $\varphi(x) = x^2$ nicht monoton. Argument über Verteilungsfunktion $y \ge 0$

$$P(Y \le y) = P(X^2 \le y) = P(X \in [-\sqrt{y}, +\sqrt{y}]) = \int_{-\sqrt{y}}^{\sqrt{y}} \varphi_{0,1}(z)\,dz$$

$$\Longrightarrow f^Y(y) = \frac{d}{dy} P(Y \le y) = \frac{d}{dy} \int_{-\sqrt{y}}^{\sqrt{y}} \varphi_{0,1}(z)\,dz = \cancel{2}\varphi_{0,1}(\sqrt{2}) \cdot \frac{1}{\cancel{2}\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} \quad y \ge 0$$

Der Dichtetransformationssatz lässt sich entsprechend im Mehrdimensionalen formulieren (s. Krengel), insbesondere gilt für eine affine Abbildung $\varphi : \mathbb{R}^d \to \mathbb{R}^d$, $\varphi(x) = Ax + b$ mit

$A \in \mathbb{R}^{d\times d}$ invertierbar, $b \in \mathbb{R}^d$ und einem d-dim. Zufallsvektor X mit Dichte f^X, dass $Y = \varphi(X) = AX + b$ die Dichte

$$f^Y(y) = f^X(A^{-1}(y-b)) \cdot |\det(A^{-1})| \quad \text{besitzt.}$$

Anwendung: $X \sim \mathcal{N}(0, \mathrm{E}_d)$, $Y = AX + b$:

$$\begin{aligned} f^Y(y) = f^X(A^{-1}(y-b))|\det(A^{-1})| &= (2\pi)^{-\frac{d}{2}} \cdot \exp\left(-\frac{|A^{-1}|(y-b)^2}{2}\right) \cdot \frac{1}{|\det(A)|} \\ &= (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{\left\langle (A^{-1})^* A^{-1}(y-b), (y-b)\right\rangle}{2}\right) \cdot \frac{1}{|\det(AA^*)|^{\frac{1}{2}}} \\ &= \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{\left\langle \Sigma^{-1}(y-b), y-b\right\rangle}{2}\right) \quad \text{mit} \quad \Sigma = AA^* \end{aligned}$$

$\Longrightarrow Y$ ist $\mathcal{N}(b, \Sigma)$-verteilt mit $\Sigma = AA^*$ (siehe III.3)

II. Bedingte Wahrscheinlichkeiten und Unabhängigkeit

II.1. Bedingte Wahrscheinlichkeit und Bayesformel

Beispiel II.1 (Motivation) Ein neu entwickelter Test auf eine Krankheit ergibt bei 1000 Versuchspersonen, davon 900 gesund und 100 krank, folgende Ergebnisse:

	Test positiv	Test negativ	Summe
gesund	15	885	900
krank	92	8	100
Summe	107	893	1000

Kontingenztafel

Was würde der Arzt einem Patienten sagen, dessen Test positiv ausfällt. Antwort: $\frac{92}{107}$ Wkeit für „krank“, d.h. ca 80% der Fälle ist der Patient wirklich krank. Andererseits liegt bei einem negativem Testergebnis in nur $\frac{8}{893} \approx 0.9\%$ der Fälle doch eine Krankheit vor.

$$\frac{92}{107} = \frac{|K \cap P|}{|P|}, \qquad \frac{8}{893} = \frac{|K \cap N|}{|N|}$$

Dieser Ansatz führt auf den Begriff der bedingten Wahrscheinlichkeit; hier würde dies – im Sinne relativer Häufigkeiten – der bedingten Wahrscheinlichkeiten für „krank“ unter der Bedingung „Test positiv“ bzw. „Test negativ“ entsprechen.

Definition II.2
Es seien A und B Ereignisse mit $P(B) > 0$. Dann bezeichnet $P(A|B) := \frac{P(A\cap B)}{P(B)}$ die bedingte Wahrscheinlichkeit *von A unter B (oder A gegeben B).*

Beispiel II.3 (Wurf zweier fairer Würfel) $A =$„Pasch“, $B =$„beide Augenzahlen ungerade“

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{3}{36}}{\frac{9}{36}} = \frac{1}{3} > \frac{1}{6} = P(A)$$

Satz II.4
Es sei B ein Ereignis mit $P(B) > 0$. Dann gilt

(a) *$A \mapsto Q(A) := P(A|B)$ ist ein W-Maß.*

(b) *Formel von der* totalen Wahrscheinlichkeit*: Für $B = \dot\bigcup_{i=1}^N B_i$ die Vereinigung paarweise disjunkter Ereignisse B_i mit $P(B_i) > 0$ und Ereignis A gilt*

$$P(A \cap B) = \sum_{i=1}^N P(B_i)P(A|B_i)$$

(c) Bayesformel *Für Ereignisse A mit $P(A) > 0$ und paarweise disjunkte Ereignisse $B_1, \dots, B_N$ mit $P(B_i) > 0 \forall i$ und $\dot\bigcup_{i=1}^N B_i = \Omega$ gilt*

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{j=1}^N P(B_j) \cdot P(A|B_j)}$$

In (b) und (c) kann $N = \infty$ gesetzt werden.

Beweis: (a) $Q(\Omega) = P(\Omega|B) = \frac{P(\overbrace{\Omega \cap B}^{B})}{P(B)} = 1$, A_i paarweise disjunkt:

$$Q(\dot\bigcup_{i \geq 1} A_i) = \frac{P((\dot\bigcup_i A_i) \cap B_i)}{P(B)} = \frac{P(\dot\bigcup(A_i \cap B))}{P(B)} = \sum_i \frac{P(A_i \cap B)}{P(B)} = \sum_i Q(A_i)$$

(b) $P(A \cap B) = P(\dot\bigcup_{i=1}^N (A \cap B_i)) = \sum_{i=1}^N (A \cap B_i) = \sum_{i=1}^N P(A|B_i) \cdot P(B_i)$

(c) Nach (b) folgt $\frac{P(B_i) \cdot P(A|B_i)}{\sum_{j=1}^N P(B_j)P(A|B_j)} = \frac{P(A \cap B_i)}{P(A)} = P(B_i|A)$ □

Bemerkung (a) Beachte den Unterschied zwischen $P(A|B)$ und $P(A \cap B)$. Es gilt stets $P(A|B) \geq P(A \cap B)$.

(b) Interpretation:

a) *frequentistisch:* Bei häufiger Wiederholung ist $P(A|B)$ der Anteil der Fälle, in denen A eintritt, in der Gesamtheit der Fälle, in denen B eintritt.

b) *subjektiv:* Ist P meine Einschätzung der Lage vor Beginn des Experiments, so ist $P(\cdot|B)$ meine Einschätzung, nachdem ich vom Eintreten von B erfahren habe.

Beispiel II.5 (Test auf eine seltene Krankheit) Eine Krankheit komm bei ca 0.5% der Bevölkerung vor. Ein Test führt bei 99% der Kranken zu einer Reaktion, aber auch bei 2% der Gesunden. Wie groß ist die Wahrscheinlichkeit, dass bei positiver Reaktion eine Person wirklich krank ist. A = „Reaktion positiv“, B = „krank“

$$P(B|A) \stackrel{\text{Bayes}}{=} \frac{P(B) \cdot P(A|B)}{P(B)P(A|B) + P(B^C)P(A|B^C)} = \frac{0.005 \cdot 0.99}{0.005 \cdot 0.99 + 0.995 \cdot 0.02} = \frac{495}{2485} \approx 0.2$$

Diese Wahrscheinlichkeit beträgt also nur ca. 20%!

Lemma II.6 (Multiplikationsformel)
Für Ereignisse $A_1, \dots A_n$ mit $P(A_1 \cap \dots \cap A_n) > 0$ gilt

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Beweis: rechte Seite: $\cancel{P(A_1)} \cdot \frac{\cancel{P(A_1 \cap A_2)}}{\cancel{P(A_1)}} \cdots \frac{P(A_1 \cap A_2 \cap \dots \cap A_n)}{\cancel{P(A_1 \cap \dots \cap A_{n-1})}} = P(A_1 \cap \dots \cap A_n)$ □

Dies führt auf die sogenannte *Pfadregel*. Betrachte z.B. ein Urnenmodell mit Ziehen ohne Zurücklegen und Beachtung der Reihenfolge sowie W weißen und S schwarzen Kugeln, $N = S + W$

II.2. Unabhängigkeit

Falls Kenntnis über das Ereignis B die Wahrscheinlichkeit des Ereignisses A nicht beeinflusst (und umgekehrt), so bezeichnet man A und B als (stochastisch) *unabhängig*

$$P(A|B) = P(A) \quad \text{und} \quad P(B|A) = P(B) \iff \frac{P(A \cap B)}{P(B)} = P(A) \iff \frac{P(A \cap B)}{P(A)} = P(B)$$

Allgemeiner definiert man auch im Fall von $P(A) = 0$ oder $P(B) = 0$

Definition II.7

(a) *Zwei Ereignisse A und B sind (stochastisch)* unabhängig, *falls $P(A \cap B) = P(A) \cdot P(B)$ gilt.*

(b) *Eine Familie $(A_i)_{i \in I}$, I beliebige Indexmenge von Ereignissen heißt* unabhängig, *falls für jede endliche Teilmenge $J \subseteq I$ gilt $P(\bigcap_{j \in J} A_j) = \prod_{j \in J} P(A_j)$.*

Satz II.8 (Lemma von Borel-Cantelli)
Sei $(A_n)_{n \geq 1}$ eine Folge von Ereignissen in $(\Omega, \mathcal{F}, P)$ sowie

$$A := \limsup_{n \to \infty} A_n := \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n = \{\omega \in \Omega \mid \omega \in A_n \text{ gilt für unendlich vielen } n \geq 1\}$$

Dann gilt

(a) *Aus $\sum_{n \geq 1} P(A_n) < \infty$ folgt $P(A) = 0$.*

(b) *Gilt $\sum_{n \geq 1} P(A_n) = \infty$ und ist $(A_n)_{n \geq 1}$ unabhängig, so folgt $P(A) = 1$.*

Beweis: (a) Es gilt $A \subseteq \bigcup_{n \geq m} A_n$ für alle $m \geq 1$, weshalb

$$P(A) \leq \inf_{m \geq 1} P(\bigcup_{n \geq m} A_n) \leq \inf_{m \geq 1} \sum_{n \geq m} P(A_n) = \lim_{m \to \infty} \sum_{n \geq m} P(A_n) \overset{\sum_{n \geq 1} P(A_n) < \infty}{=} 0$$

(b) Betrachte $A^C := \bigcup_{m\geq 1}\bigcap_{n\geq m} A_n^C$:

$$\begin{aligned}
P(A^C) &\leq \sum_{m\geq 1} P(\bigcap_{n\geq m} A_n^C) \overset{\sigma\text{-stet.}}{=} \sum_{m\geq 1}\lim_{N\to\infty} P(\bigcap_{n=m}^N A_n^C) \\
&\overset{\text{UE}}{=} \sum_{m\geq 1}\lim_{N\to\infty}\prod_{n=m}^N \underbrace{P(A_n^C)}_{1-P(A_n)} \qquad 1-x\leq e^{-x}, x\geq 0 \\
&\leq \sum_{m\geq 1}\lim_{N\to\infty}\prod_{n=m}^N \exp(-P(A_n)) \\
&= \sum_{m\geq 1}\exp(-\lim_{N\to\infty}\sum_{n=m}^N P(A_n)) = \sum_{m\geq 1} 0 = 0
\end{aligned}$$

□

Beispiel II.9 (a) In Teil (b) kann nicht auf die Unabhängigkeit verzichtet werden: Betrachte z.B. ein Ereignis A_1 mit $P(A_1)\in(0,1)$ und setze $A_n := A_1$, $n\geq 2$. Dann ist $\limsup_{n\to\infty} A_n = A_1$ und somit $P(\limsup_{n\to\infty} A_n) < 1$, obwohl $\sum_{n\geq 1} P(A_n) = +\infty$ ist.

(b) Runs im beliebig langen Münzwurf: Es sei A_n^M das Ereignis „Kopf in den Würfen $n, n+1, \dots, n+M-1$" (M-Run ab n-ten Münzwurf). Frage: $P(\limsup_{n\to\infty} A_n^{(M)})$ =?.

Da $(A_n^{(M)})_{n\geq 1}$ *nicht* unabhängig ist, betrachten wir die Teilfolge $(A_{kM+1}^{(M)})_{k\geq 1}$. Diese ist unabhängig, weil für $k_1 < k_2 < \dots < k_n$ gilt

$$\begin{aligned}
P\left(A_{k_1M+1}^{(M)}\cap\ldots\cap A_{k_nM+1}^{(M)}\right) &= \sum_{\omega\in A_{k_1M+1}^{(M)}\cap\ldots\cap A_{k_nM+1}^{(M)}} \left(\frac{1}{2}\right)^{\sum_{i=1}^{(k_n+1)\cdot M}(1-\omega_i)} = 2^{-M\cdot n} \\
&= \underbrace{P\left(A_{k_1M+1}^{(M)}\right)}_{2^{-M}}\cdots\underbrace{P\left(A_{k_nM+1}^{(M)}\right)}_{2^{-M}}
\end{aligned}$$

$$\underset{\sum_{k\geq 1} P(A_{kM+1}^{(M)})=\infty}{\overset{\text{Borel-Cantelli (b)}}{\Longrightarrow}} P\left(\limsup_{n\to\infty} A_n^{(M)}\right) \geq \lim_{k\to\infty} P\left(\limsup_{k\to\infty} A_{kM+1}^{(M)}\right) \overset{\text{B-C}}{=} 1$$

Es gilt sogar noch mehr:

$$P(\text{“für jedes } M\geq 1 \text{ gibt es } \infty\text{-viele } M\text{-Runs“}) = P\left(\bigcap_{M\geq 1}\limsup_{n\to\infty} A_n^{(M)}\right) = 1$$

weil der Schnitt von 1-Mengen wieder eine 1-Menge ist, d.h.:

$$P(E_n) = 1,\ n\geq 1 \implies P(\bigcap_{n\geq 1} E_n) = 1 - P(\bigcup_{n\geq 1} E_n^C) \geq 1 - \sum_{n\geq 1} P(E_n^C) = 1$$

Definition II.10
Seien $\mathcal{M}_i \subset \mathcal{F}_i, i\in I$ Mengen von Ereignissen. Dann heißt $(\mathcal{M}_i)_{i\in I}$ unabhängig, *falls für jede beliebige Auswahl von $A_i\in\mathcal{M}_i$, die Familie $(A_i)_{i\in I}$ unabhängig ist.*

II.3. Unabhängige Zufallsvariablen

Definition II.11
Eine Familie $(X_i)_{i\in I}$ von $(S_i, \mathcal{S}_i)$-wertige Zufallsvariablen heißt unabhängig, *falls für beliebige Ereignisse $A_i \in \mathcal{S}_i$ die Ereignisse $(\{X_i \in A_i\})_{i\in I}$ unabhängig sind. Äquivalent ist $(X_i)_{i\in I}$ unabhängig genau dann, wenn die* von den X_i erzeugten σ-Algebren $\mathcal{F}^{X_i} := \{X_i^{-1}(A) \mid A \in \mathcal{S}_i\}, i \in I$ *unabhängig sind.*

Satz II.12
Es seien $(X_i, i \in I)$ Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$ mit Werten in $(S_i, \mathcal{S}_i)$ und $\mathcal{E}_i$ seien $\cap$-stabile Erzeuger der $\mathcal{S}_i$. Damit ist $(X_i)_{i\in I}$ bereits unabhängig, falls $(\{X_i \in A_i\})_{i\in I}$ unabhängig sind für beliebige Wahlen von A_i in $\mathcal{E}_i$.

Beweis: Es reicht, diesen Satz für beliebige *endliche* Indexmengen $J \subseteq I$ zu beweisen. Wegen $|J| < \infty$ gibt es dann ein n mit $n \geq |\{j \in J \mid A_j \notin \mathcal{E}_j\}|$ bei einer beliebigen Wahl von $A_i \in \mathcal{S}_i$.

Induktionsbehauptung: Es gilt $P\left(\bigcap_{j\in J}\{X_j \in A_j\}\right) \overset{(*)}{=} \prod_{j\in J} P\left(\{X_j \in A_j\}\right)$ für beliebige $A_j \in \mathcal{S}_j$, sofern $|\{j \in J \mid A_j \notin \mathcal{E}_j\}| \leq n$.

Induktionsanfang ($n = 0$): Das ist gerade die Annahme.

Induktionsschritt ($n \to n+1$): Sei o.B.d.A. $|J| \geq 2$ und seien die A_j so, dass $|\{j \in J \mid A_j \notin \mathcal{E}_j\}| = n+1$ gilt. Setze $J' = J \setminus \{j\}$ für ein j mit $A_j \notin \mathcal{E}_j$. Nach Annahme gilt $(*)$ für J'. O.B.d.A. sei dabei $P\left(\bigcap_{j\in J'}\{X_j \in A_j\}\right) > 0$ (sonst gilt $(*)$ trivialerweise auch für J : $0 = 0$). Betrachte nun auf $\mathcal{S}_j$ die W-Maße (!!)

$$A \mapsto P\left(\{X_j \in A\} \mid \bigcap_{i\in J'}\{X_i \in A_i\}\right) =: Q_1(A) \qquad A \mapsto P(X_j \in A) =: Q_2$$

Für $A \in \mathcal{E}_j$ gilt nach Induktionsannahme

$$Q_1(A) = \frac{P\left(\{X_j \in A\} \cap \bigcap_{i\in J'}\{X_i \in A_i\}\right)}{P\left(\bigcap_{i\in J'}\{X_i \in A_i\}\right)} \overset{\text{I.A.}}{=} \frac{P(X_j \in A) \cdot \cancel{\prod_{i\in J'} P(X_i \in A_i)}}{\cancel{\prod_{i\in J'} P(X_i \in A_i)}} = Q_2$$

$$\underset{\text{keitssatz}}{\overset{\text{Eindeutig-}}{\Longrightarrow}} Q_1 = Q_2 \implies (*) \quad \text{gilt für } J$$ □

Korollar II.13
Es seien $X_1, \ldots, X_n$ Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$.

(a) *Sind alle X_k diskret verteilt, so sind $X_1, \ldots, X_n$ genau dann unabhängig, wenn gilt*

$$P(X_1 = s_1, \ldots, X_n = s_n) = p^X(s_1, \ldots, s_n) \overset{!}{=} p^{X_1}(s_1) \cdots p^{X_n}(s_n) = P(X_1 = s_1) \cdots P(X_n = s_n)$$

falls alle $s_i \in S_i$, $i = 1 \ldots, n$

(b) *Sind die X_k reellwertig, so sind $X_1, \ldots, X_n$ genau dann unabhängig, wenn gilt*

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = F^X(x_1, \ldots, x_n) = F^{x_1}(x_1) \cdots F^{x_n}(x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n)$$

Beweis: (a) Die Richtung „⇒“ ist klar (wähle $A_i = \{s_i\}$ in der Definition). Andererseits ist $\mathcal{E}_k = \{\{s_j\} \mid s_k \in S_k\} \cup \{\varnothing\}$ ein $\cap$-stabiler Erzeuger von $\mathcal{S}_k = \mathcal{P}(S_k)$. Die trivialen Ereignisse $\{X_k \in \varnothing\} = \varnothing$ erfüllen stets die Unabhängigkeitsbedingung $P(\varnothing) = 0$). Nach Voraussetzung gilt nun

$$P(X_1 \in A_1, \dots X_n \in A_n) = \prod_{k=1}^n P(X_k \in A_k)$$

für die $\cap$-stabilen Erzeuger $\mathcal{E}_k$ und alle $A_k \in \mathcal{E}_k \overset{\text{Satz 2.12}}{\Longrightarrow} X_1, \dots, X_n$ sind unabhängig.

(b) Wir schließen wie in (a) mit den $\cap$-stabilen Erzeugen $\mathcal{E}_k = \{(-\infty, x] \mid x \in \mathbb{R}\}$. □

Satz II.14
Es sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor auf $(\Omega, \mathcal{F}, P)$ mit W-Dichte $f^X : \mathbb{R}^n \to [0, \infty)$. Dann gilt:

(a) Jedes X_k besitzt eine Dichte, die sogenannte Randdichte

$$f^{X_k}(x_k) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{n-1} f^X(x_1, \dots, x_n)\, dx_1 \dots dx_{k-1}\, dx_{k+1} \dots dx_n$$

(b) Die Zufallsvariablen $X_1, \dots, X_n$ sind unabhängig genau dann, wenn

$$f^X(x_1, \dots, x_n) = f^{X_1}(x_1) \cdots f^{X_n}(x_n)$$

für Lebesgue-fast alle $x_1, \dots, x_n \in \mathbb{R}$.

Beweis: (a) Es gilt

$$P(X_k \le b_k) = \int_{\{(\xi_1 \dots, \xi_n) \in \mathbb{R}^n \mid \xi_k \le b_k\}} P(dx) \overset{\text{Fubini}}{=} \int_{-\infty}^{b_k} f^{X_k}(\xi_k)\, d\xi_k$$

für alle $b_k \in \mathbb{R}$ und daher f^{X_k} ist Dichte von X_k.

(b) Es bleibt zu zeigen (wegen Satzes):

$$P(X_1 \le b_1, \dots, X_n \le b_n) = \prod_{k=1}^n P(X_k \le b_k) \qquad \forall b_1, \dots, b_n$$
$$\Longleftrightarrow f^X(x_1, \dots, x_n) = \prod_{k=1}^n f^{X_k}(x_k) \qquad \lambda\text{-f.ü.}$$

„⇐“ folgt durch Integration

$$P(X_1 \le b_1, \dots, X_n \le b_n) = \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_n} f^X(x_1, \dots, x_n)\, dx_n \dots dx_1$$
$$\overset{\text{Fubini}}{=} \prod_{k=1}^n \int_{-\infty}^{b_k} f^{X_k}(x_k)\, dx_k = \prod_{k=1}^n P(X_k \le b_k)$$

„⇒“: Die selbe Rechnung zeigt hier

$$\int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_n} f^X(x_1, \ldots, x_n)\, dx_n \ldots dx_1 = \prod_{k=1}^n \int_{\infty}^{b_k} f^{X_k}(x_k)\, dx_k \qquad \forall b_1, \ldots, b_n \in \mathbb{R}$$

$$\underset{\substack{\text{Hauptsatz der} \\ \text{Diff/Int-Rechnung}}}{\overset{\text{Maßtheorie}}{\Longrightarrow}} \quad f^X(x_1, \ldots, x_n) = \prod_{k=1}^n f^{X_k}(x_k) \qquad \lambda\text{-f.ü.}$$

□

Beispiel II.15 (a) Besitzt $X = (X_1, \ldots, X_n)$ eine Produktdichte $f^X : \mathbb{R}^n \to [0, \infty)$ (d.h. $f^X(x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i)$), so sind $X_1, \ldots, X_n$, unabhängig mit den Randdichten $f^{X_i}(x) = f_i(x)$:

$$f^{X_i}(x_i) = \underbrace{\int \cdots \int}_{(n-1)\text{-mal}} \underbrace{f^X(x_1, \ldots, x_n)}_{(\prod_{j \neq i} f_j(x_j)) \cdot f_i(x_i)}\, dx_1 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_n = 1 \cdot f_i(x_i)$$

und obiger Satz zur Unabhängigkeit findet Anwendung. Bei der n-dimensionalen Standardnormalverteilung $X \sim \mathcal{N}(O, E_n)$ mit Dichte $f^X(x) = (2\pi)^{-\frac{n}{2}} e^{-\frac{|x|^2}{2}}, x \in \mathbb{R}^n$ hat eindimensionale Randverteilungen

$$\mathcal{N}(0,1): \qquad f^X(x) = \prod_{i=1}^n \frac{1}{2\pi} e^{-\frac{x_i^2}{2}}$$

Außerdem sind die Koordinaten $X_1, \ldots X_n$ unabhängig.

(b) Es sei $X = (X_1, X_2)$ gleichverteilt auf $G = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$, d.h. $f^X(x) = \frac{1}{|G|} \mathbf{1}_G(x), x \in \mathbb{R}^2$. Schreibe nun X in Polarkoordinaten $R = \sqrt{X_1^2 + X_2^2}$, $\Phi = \arctan(\frac{X_2}{X_1}) \in [0, 2\pi)$. Es gilt für beliebige $r \in [0,1], \varphi \in [0, 2\pi]$:

$$P(R \leq r, \Phi \leq \varphi) = \varphi \frac{r^2}{2} \cdot \frac{1}{\pi} = \underbrace{P(R \leq r)}_{r^2} \underbrace{P(\Phi \leq \varphi)}_{\frac{\varphi}{2\pi}} \implies \quad R, \Phi \text{ unabhängig.}$$

Also ist $f^R(r) = 2r \cdot \mathbf{1}_{[0,1]}(r)$ und $f^\Phi(\varphi) = \frac{1}{2\pi} \mathbf{1}_{[0,2\pi)}(\varphi)$. Mit dem Satz ergibt sich die gemeinsame Dichte als Produktdichte:

$$f^{(R,\Phi)}(r, \varphi) = \frac{r}{\pi} \mathbf{1}_{[0,1]}(R) \mathbf{1}_{[0,2\pi)}(\varphi)$$

Alternative Herleitung: Dichtetransformation im $\mathbb{R}^2$ $(X_1, X_2) \mapsto (R, \Phi)$.

(c) Falls $X_1, \ldots, X_n$ Dichten $f^{X_i} : \mathbb{R} \to [0, \infty)$ besitzen, so braucht der Zufallsvektor $X = (X_1, \ldots X_n)$ keine Dichte zu besitzen. Einfachstes Beispiel ist der Fall $X_1(\omega) = X_2(\omega) = \ldots = X_n(\omega) \forall \omega \in \Omega$. Dann gilt für die Diagonale $D = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1 = x_2 = \ldots = x_n\}$, dass $P(X \in D) = P(X_1 = X_2 = \cdots = X_n) = 1$ und andererseits besitzt D Lebesguemaß Null im $\mathbb{R}^n, n \geq 2$. Hätte X eine Dichte f^X, so wäre $P(X \in D) = \int_D f^X(x)\, dx = 0 \neq 1$. Also gibt es keine Dichte f^X. Dies heißt insbesondere, dass für $X_1, \ldots, X_n$ $\mathcal{N}(0,1)$-verteilt nicht notwendigerweise $X = (X_1, \ldots, X_n)$ als Vektor normalverteilt ist ($P = \mathcal{N}(0, E_n)$ gilt *nur* bei Unabhängigkeit).

Zur Existenz unabhängiger Zufallsvariablen

Frage: Gegeben Verteilungen P_i auf $(S_i, \mathcal{S}_i)$, gibt es unabhängige Zufallsvariablen $(X_i)_{i\in I}$ mit $P^{X_i} = P_i$? Gesucht ist also ein W-Raum $(\Omega, \mathcal{F}, P)$ und messbare Abbildungen $X_i : \Omega \to S_i$.

Definition II.16
Es seien $(\Omega_i, \mathcal{F}_i, P_i)_{1\le i\le n}$ W-Räume. Setze $\Omega = \Omega_1 \times \cdots \times \Omega_n$,

$$\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 \cdots \otimes \mathcal{F}_n = \sigma\left(\{A_1 \times A_2 \times \cdots \times A_n \mid A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n\}\right)$$

Dann heißt $\mathcal{F}$ Produkt σ-Algebra *auf Ω der $\mathcal{F}_1, \ldots \mathcal{F}_n$. Gilt für ein W-Maß P auf $\mathcal{F}$ $P(A_1 \times \cdots \times A_n) = P(A_1)\cdots P(A_n)$ für alle $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$, s heißt P* Produktmaß *der $P_1, \ldots, P_n$ Notation $P = P_1 \otimes P_2 \otimes \cdots \otimes P_n$.*

Bemerkung Da die Rechteckmengen $A_1 \times A_2 \times \cdots A_n$ ein $\cap$-stabiler Erzeuger von $\mathcal{F}$ sind, ist das Produktmaß auf jeden Fall eindeutig bestimmt. In der Maßtheorie wird gezeigt, dass ein solches Produktmaß stets existiert („Prinzip von Cavalieri").

Beispiel II.17 Sind $f_1, \ldots, f_d : \mathbb{R} \to [0, \infty)$ W-Dichten von W-Maßen $P_1, \ldots, P_d$, auf $\mathcal{B}_\mathbb{R}$, so hat das Produktmaß $P = P_1 \otimes \cdots \otimes P_d$ die Produktdichte $f(x_1, \ldots, x_d) = f_1(x_1)\cdots f_d(x_d)$ denn:

$$P(A_1 \times \cdots \times A_d) = \int_{A_1\times\cdots\times A_d} f(x)\,dx \overset{\text{Fubini}}{=} \int_{A_1}\int_{A_2} \cdots \int_{A_d} f_1(x_1)\cdots f_d(x_d)\,dx_d \cdots dx_2\,dx_1 = \prod_{k=1}^{d} \underbrace{\int_{A_k} f_k(x)\,dx}_{P_k(A_k)}$$

$$= \prod_{k=1}^{d} P_k(A_k)$$

Ein anderes Beispiel ist das Bernoulli-Schema, wofür für die Zähldichte gilt

$$p(\underbrace{\omega_1}_{\in\{0,1\}}, \ldots \omega_n) = p^{\sum \omega_i}(1-p)^{\sum(1-\omega_i)} = \prod_{i=1}^{n} p^{\omega_i}(1-p)^{1-\omega_i}$$

Wir erhalten also ein Produktmaß auf dem Produktraum $\{0,1\}^n$, $\mathcal{F} = \mathcal{P}(\Omega) \overset{!}{=} \mathcal{P}(\{0,1\})^{\otimes n}$.

Lemma II.18
Ist $(\Omega_1 \times \cdots \times \Omega_n, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n, P_1 \otimes \cdots \otimes P_n)$ ein Produkt-W-Raum*, so sind die Koordinatenabbildung $X_i(\omega_1, \ldots, \omega_n) := \omega_i$ von $\Omega_1 \times \cdots \times \Omega_n$ auf Ω_i unabhängige $(\Omega_i, \mathcal{F}_i)$-wertige Zufallsvariablen mit Verteilung P_i.*

Beweis: X_i ist $(\mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n, \mathcal{F}_i)$-messbar wegen

$$X_i^{-1}(A) \overset{(*)}{=} \Omega_1 \times \cdots \times \Omega_{i-1} \times A \times \Omega_{i+1} \times \cdots \times \Omega_n \in \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$$

für alle $A \in \mathcal{F}_i$. Außerdem gilt

$$\begin{aligned}(P_1 \otimes \cdots P_n)^{X_i}(A) &= P_1 \otimes \cdots \otimes P_n(X_i^{-1}(A)) \\ &\underset{\text{und }(*)}{\overset{\text{Produktmaß}}{=}} \underbrace{P_1(\Omega_1)}_{=1} \cdots P_{i-1}(\Omega_{i-1}) \cdot P_i(A) \cdot P_{i+1}(\Omega_{i+1}) \cdots P_n(\Omega_n) = P_i(A)\end{aligned}$$

Also besitzt X_i die Verteilung P_i.

Unabhängigkeit folgt aus

$$\begin{aligned}&(P_1 \otimes \cdots \otimes P_n)(X_1 \in A_1, \ldots, X_n \in A_n) = P_1 \otimes \cdots \otimes P_n(A_1 \times \cdots \times A_n)\\ &= P_1(A_1)\cdots P_n(A_n) \overset{\text{s.o.}}{=} (P_1 \otimes \cdots \otimes P_n)(X_1 \in A_1)\cdots(P_1 \otimes \cdots \otimes P_n)(X_n \in A_n) \qquad \square\end{aligned}$$

Für unendliche Produkträume gelten bei W-Maßen analoge Resultate, für deren Beweis wir auf Stochastik II verweisen (dort allg. für stochastische Prozesse).

Definition II.19
Es seien $(\Omega_i, \mathcal{F}_i, P_i)_{i\in I}, I \neq \varnothing$ beliebige Indexmenge, eine Familie von W-Räumen. Setze $\Omega := \prod_{i\in I} \Omega_i = \times_{i\in I} \Omega_i$ (kartesisches Produkt), $\pi_j((\omega_i)_{i\in I}) = \omega_j$ (j-te Koordiantenprojektion), $\mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i := \sigma(\{\pi_i^{-1}(A_i) \mid i \in I, A_i \in \mathcal{F}_i\})$. (Produkt-$\sigma$-Algebra) Dann heißt P Produktmaß *der $(P_i)_{i\in I}$ auf F, falls*

$$P\left(\bigcap_{i\in J} \pi_i^{-1}(A_i)\right) = \prod_{i\in J} P_i(A_i)$$

gilt für alle $J \subseteq I$ endlich und $A_i \in \mathcal{F}_i$.

Beachte: Ist $I = \{1, \ldots, n\}$ so ist die Produkt-σ-Algebra gerade gleich $\sigma(\{A_1 \times \cdots \times A_n \mid A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n\})$, weil $A_1 \times \cdots \times A_n = \pi_1^{-1}(A_n) \cap \ldots \cap \pi_n^{-1}(A_n)$ gilt und σ-Algebren $\cap$-stabil sind. Die Definition des Produktmaßes ist ebenso wie vorher.

Satz II.20
Es solches Produktmaß existiert stets und ist eindeutig.

Korollar II.21
Zu gegebenen W-Maßen P_i auf $(\Omega_i, \mathcal{F}_i)$, $i \in I$, existiert stets eine Familie unabhängiger Zufallsvariablen X_i mit $P^{X_i} = P_i$, $i \in I$, nämlich

$$\Omega = \prod_{i\in I} \Omega_i, \quad \mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i, \quad P = \bigotimes_{i\in I} P_i \quad \textit{und} \quad X_i(\omega) = \pi_i(\omega) = \omega_i$$

Beweis: wie oben

Beispiel II.22 Man kann ein unendlich langes Münzwurfexperiment auf

$$\Omega = \{0,1\}^{\mathbb{N}}, \mathcal{F} = \mathcal{P}(\{0,1\})^{\otimes\mathbb{N}} \subsetneq \mathcal{P}(\{0,1\}^{\mathbb{N}})$$

konstruieren (Siehe Vitali).

Satz II.23 (0-1-Gesetz von Kolmogorov)
Seien $(X_k)_{k\geq 1}$ unabhängige Zufallsvariablen auf $(\Omega, \mathcal{F}, P)$. Dann gilt für jedes bezüglich $(X_k)_{k\geq 1}$ asymptotische Ereignis A:

$$P(A) = 0 \quad \textit{oder} \quad P(A) = 1$$

Definition II.24
Ein Ereignis A heißt asymptotisch *bezüglich $(X_k)_{k\geq 1}$, falls es für alle $n \in \mathbb{N}$ nur von $(X_k, k \geq n)$ abhängt, genauer falls A ist der asymptotischen σ-Algebra $\mathfrak{A}^X = \bigcap_{n\geq 1} \sigma(\bigcup_{k\geq n} \mathcal{F}^{X_k})$ (mit $\mathcal{F}^{X_k} = \{X_k^{-1}(B) \mid B \in \mathcal{S}\}$) liegt.*

Lemma II.25
Sind $(X_i)_{i\in I}$ $(S_i,\mathcal{S}_i)$-wertige unabhängige Zufallsvariablen und $\varphi_i : S_i \to T_i$ $(\mathcal{S}_i,\mathcal{T}_i)$-messbar, so sind auch die $(T_i,\mathcal{T}_i)$-wertige Zufallsvariablen $(\varphi_i(X_i))_{i\in I}$ unabhängig.

Beweis: Zu zeigen ist, dass für $\forall A_i \in \mathcal{T}_i$ die $(\{\varphi_i(X_i) \in A_i\})_{i\in I}$ unabhängig sind. Wegen

$$\begin{aligned}\{\varphi_i(X_i) \in A_i\} &= \{\omega \in \Omega \mid \varphi_i(X_i(\omega)) \in A_i\} = \{\omega \in \Omega \mid X_i(\omega) \in \varphi^{-1}(A_i)\}\\ &= \{X_i \in \underbrace{\varphi_i^{-1}(A_i)}_{=:B_i}\}\end{aligned}$$

und $B_i \in \mathcal{S}_i$ (da φ_i messbar) sind die Ereignisse $\{\varphi_i(X_i) \in A_i\} = \{X_i \in B_i\}$ unabhängig auf Grund der Unabhängigkeit der $(X_i)_{i\in I}$. □

Beispiel II.26 In obigen Beispiel der Gleichverteilung auf dem Einheitskreis waren R,Φ unabhängig. Dann sind zum Beispiel auch (R^2,Φ) oder $(R,\cos\Phi)$ unabhängig, aber *nicht* $(R\cos\Phi, R\sin\Phi)$.

Frage: Betrachte die harmonische Reihe $\sum_{k=1}^{\infty}\frac{1}{k}$, wähle aber die Vorzeichen jedes Summanden rein zufällig. Wie groß ist die Wahrscheinlichkeit, dass diese zufällige Reihe konvergiert?

Wir wissen:

$$\sum_{k=1}^{\infty}\frac{1}{k} = 1+\frac{1}{2}+\frac{1}{3}+\cdots = +\infty \qquad \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k} = \log 2 < +\infty$$

Nun betrachte $\sum_{k=1}^{\infty} X_k\frac{1}{k}$ mit $P(X_k = 1) = P(X_k = -1)$ und $(X_k)_{k\geq 1}$ unabhängig.

Zur Erinnerung: A ist asymptotisches Ereignis bzgl. $(X_k)_{k\geq 1}$ falls $A \in \mathfrak{A}_X = \bigcap_{n\geq 1}\sigma\left(\bigcup_{k\geq n}\mathcal{F}^{X_k}\right)$

Hier:

$$A := \left\{\sum_{k=1}^{\infty}\frac{X_k}{k} \text{ konvergiert}\right\} = \bigcap_{\substack{\varepsilon>0\\ \varepsilon\in\mathbb{Q}}}\bigcup_{N\in\mathbb{N}}\bigcap_{m,n\geq N}\underbrace{\left\{\left|\sum_{k=n}^{m}\frac{X_k}{k}\right| < \varepsilon\right\}}_{\in\sigma\left(\bigcup_{k=n}^{m}\mathcal{F}^{X_k}\right)=\mathcal{F}^{X_n,\dots,X_m}\subseteq\sigma\left(\bigcup_{k\geq N}\mathcal{F}^{X_k}\right)}$$

$$\Longrightarrow \left(\bigcup_{N\in\mathbb{N}}\bigcap_{m,n\geq N}\left\{\left|\sum_{k=n}^{m}\frac{X_k}{k}\right| < \varepsilon\right\}\right)^C = \bigcap_{N\geq 1}\overbrace{\bigcup_{m,n\geq N}\underbrace{\left\{\left|\sum_{k=n}^{m}\frac{X_k}{k}\right| \geq \varepsilon\right\}}_{\in\sigma\left(\bigcup_{k\geq N}\mathcal{F}^{X_k}\right)}}^{\in\sigma\left(\bigcup_{k\geq N}\mathcal{F}^{X_k}\right)} \in \mathfrak{A}_X$$

$A^C \in \mathfrak{A}_X \implies A \in \mathfrak{A}_X$, also ist A asymptotisches Ereignis. Das 0-1-Gesetz von Kolmogorov impliziert daher $P(A) \in \{0,1\}$ d.h. mit Wahrscheinlichkeit 1 konvergiert diese zufällige Reihe oder divergiert. Es kann also nicht sein, dass z.B. beides mit Wahrscheinlichkeit 0.5 eintritt.

Zum Beweis des 0-1-Gesetzes benötigen wir

Lemma II.27
Es seien $(X_i)_{i\in I}$ eine Familie unabhängiger Zufallsvariablen mit Werten in $(S_i,\mathcal{S}_i)$ und $I = I_1 \uplus I_2$ eine Zerlegung der Indexmenge I. Dann sind die σ-Algebren $\mathcal{F}_1 = \sigma\left(\bigcup_{i\in I_1}\mathcal{F}^{X_i}\right)$ und $\mathcal{F}_2 = \sigma\left(\bigcup_{i\in I_2}\mathcal{F}^{X_i}\right)$ unabhängig.

Beweis: Man betrachte die Zufallsvariablen $Y_1 = (X_i)_{i \in I_1}$, $Y_2 = (X_i)_{i \in I_2}$ mit Werten in $(\prod_{i \in I_1} S_i, \bigotimes_{i \in I_1} \mathcal{S}_i)$ beziehungsweise $(\prod_{i \in I_2} S_i, \bigotimes_{i \in I_2} \mathcal{S}_i)$. [Messbarkeit folgt, weil $\bigotimes_{i \in I_1} \mathcal{S}_i$ von $\{\pi_i^{-1}(A) \mid i \in I_1, A \in \mathcal{S}_i\}$ erzeugt wird.] Wegen dieses Erzeugers ist die Unabhängigkeit der σ-Algebren $\mathcal{F}_1, \mathcal{F}_2$ auch äquivalent zur Unabhängigkeit der Zufallsvariablen Y_1, Y_2 (d.h. $\mathcal{F}_1 = \mathcal{F}^{Y_1}, \mathcal{F}_2 = \mathcal{F}^{Y_2}$). Nach obigem Satz reicht es, für die $\cap$-stabilen Erzeuger $\mathcal{E}_j = \left\{\bigcup_{i \in J} \pi_i^{-1}(A_i) \mid J \subseteq I_j, \text{endlich}, A_i \in \mathcal{S}_i\right\}$ für $j = 1, 2$ zu zeigen.

$$\forall B_1 \in \mathcal{E}_1, B_2 \in \mathcal{E}_2 : P(Y_1 \in B_1, Y_2 \in B_2) = P(Y_1 \in B_2) \cdot P(Y_2 \in B_2)$$

Dies folgt nun aus der Unabhängigkeit der $(X_i)_{i \in I}$: Betrachte $B_1 = \bigcap_{i \in J_1} \pi_i^{-1}(A_i), B_2 = \bigcap_{i \in J_2} \pi_i^{-1}(A_i)$ mit $J_1, J_2 \subseteq I$ endlich, $J_1 \cap J_2 = \varnothing$, $A_i \in \mathcal{S}_i$. Dann gilt

$$\begin{aligned} P(Y_1 \in B_1, Y_2 \in B_2) &= P(\forall i \in J_1, X_i \in A_i \text{ und } \forall i \in J_2 : X_i \in A_i) \underset{\text{der } X_i}{\overset{\text{Unabh.}}{=}} \prod_{i \in J_1 \cup J_2} P(X_i \in A_i) \\ &= \left(\prod_{i \in J_1} P(X_i \in A_i)\right)\left(\prod_{i \in J_2} P(X_i \in A_i)\right) \\ &\underset{\text{unabh.}}{\overset{X_i}{=}} P(\forall i \in J_1 : X_i \in A_i) P(\forall i \in J_2 : X_i \in A_i) \\ &= P(Y_1 \in B_1) \cdot P(Y_2 \in B_2) \end{aligned}$$

□

Beweis: (des 0-1-Gesetzes) Betrachte die σ-Algebren $\mathcal{F}_1^n = \sigma\left(\bigcup_{k=1}^n \mathcal{F}^{X_k}\right)$ sowie $\mathcal{F}_{n+1}^\infty = \sigma\left(\bigcup_{k=n+1}^\infty \mathcal{F}^{X_k}\right)$. Nach dem Lemma sind $\mathcal{F}_1^n, \mathcal{F}_{n+1}^n$ unabhängig. Für die Zufallsvariable $X = (X_k)_{k \geq 1}$ mit Werten in $(\prod_{k \geq 1} S_k, \bigotimes_{k \geq 1} \mathcal{S}_k)$ und für $\mathbf{1}_A(\omega)$ gilt wegen

$$\forall n \geq 1, B_i \in \mathcal{S}_i : P\Bigg(\underbrace{X \in \bigcap_{i=1}^n \pi_i^{-1}(B_i)}_{=\{X_1 \in B_1, \dots, X_n \in B_n\} \in \mathcal{F}_1^n}, \underbrace{\mathbf{1}_A = 1}_{\in \mathcal{F}_{n+1}^n}\Bigg) = P\left(X \in \bigcap_{i=1}^n \pi_i^{-1}(B_i)\right) P(\mathbf{1}_A = 1)$$

Nun ist $\{\bigcap_{i=1}^n \pi_i^{-1}(B_i) \mid n \in \mathbb{N}, B_i \in \mathcal{S}_i\}$ ein σ-stabiler Erzeuger von $\bigotimes_{i \geq 1} \mathcal{S}_i$. Also folgt nach obigem Satz, dass X und $\mathbf{1}_A$ unabhängige Zufallsvariablen sind [beachte $\mathcal{F}^{\mathbf{1}_A} = \{A, A^C, \Omega, \varnothing\}$ + Übung]. Andererseits gilt $\mathfrak{A}_X \subseteq \sigma\left(\bigcup_{k \geq 1} \mathcal{F}^{X_k}\right) = \mathcal{F}^X$. Also gibt es für $A \in \mathfrak{A}_X$ ein $B \in \bigotimes_{k \geq 1} \mathcal{S}_k$ mit $A = X^{-1}(B)$. Sind $\mathbf{1}_A$ und X unabhängig, so sind auch $\mathbf{1}_B(X)$ und $\mathbf{1}_A$ nach obigem Lemma unabhängig und daher ist $\mathbf{1}_A$ von sich selbst unabhängig.

$$\begin{aligned} &\implies P(A) = P(\mathbf{1}_A = 1) \implies P(\{\mathbf{1}_A = 1\} \cap \{\mathbf{1}_A = 1\}) = P(\mathbf{1}_A = 1) P(\mathbf{1}_A = 1) = P(A)^2 \\ &\implies P(A) \in \{0, 1\} \end{aligned}$$

□

II.4. Faltung

Definition II.28
Sind P und Q W-Maße auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, so definiere die Faltung *$P * Q$ von P und Q als das W-Maß auf*

$$(\mathbb{R}, \mathcal{B}_\mathbb{R}) \quad \textit{mit} \quad (P * Q)(B) = \int P(B - \{x\}) Q(dx), \qquad B \in \mathcal{B}_\mathbb{R}$$

Bemerkung (a) Dabei ist $B - \{x\} = \{b - x \mid b \in B\}$, z.B. $[0,1] - \{x\} = [-x, 1-x]$.

(b) $\int_{\mathbb{R}} f(x)Q(\,dx)$ bezeichnet das *Maß-Integral* (Lebesgue-Integral) von f bzgl. dem Maß Q. Insbesondere ist im Fall, dass Q eine W-Dichte q auf $\mathbb{R}$ besitzt,

$$\int_{\mathbb{R}} f(x)Q(\,dx) = \int_{\mathbb{R}} f(x)q(x)\,dx$$

Alternative Schreibweisen sind $\int f(x)Q(\,dx) = \int f(x)\,dQ(x) = \int f\,dQ$

(c) Dass $P * Q$ in der Tat ein W-Maß ist (und dass $x \mapsto P(B - \{x\})$ messbar ist), wird sich aus dem folgenden Lemma ergeben:

Lemma II.29
Es seien X und Y unabhängige *reellwertige Zufallsvariablen. Dann besitzt $X+Y$ die Verteilung $P^{X+Y} = P^X * P^Y$.*

Beweis: Wir zeigen, dass die Verteilungsfunktionen übereinstimmen, d.h.

$$\forall b \in \mathbb{R} : \underbrace{P(X+Y \le b)}_{P^{X+Y}((-\infty,b])} = P^X * P^Y((-\infty,b])$$

$$\begin{aligned} P(X+Y \le b) &= \int_{\{(x,y)\mid x+y\le b\}} P^{(X,Y)}(\,dx,\,dy) = \int_{\{(x,y)\mid x+y\le b\}} (P^X \oplus P^Y)(\,dx,\,dy) \\ &\overset{\text{Fubini}}{=} \int \left(\int \mathbf{1}(x+y \le b) P^X(\,dx) \right) P^Y(\,dy) \\ &= \int P^X\big((-\infty, b-y]\big) P^Y(\,dy) = \int P^X\big((-\infty,b] - \{y\}\big) P^Y(\,dy) \\ &= P^X * P^Y\big((-\infty,b]\big) \end{aligned}$$

□

Korollar II.30
Die Faltung ist kommutativ, also $P * Q = Q * P$ und assoziativ $P * (Q * R) = (P * Q) * R$.

Beweis: Das folgt sofort aus den Eigenschaften der Zufallsvariablen:

$$X + Y = Y + X \quad \text{bzw.} \quad (X+Y)+Z = X+(Y+Z)$$

für unabhängige Zufallsvariablen mit $P^X = P, P^Y = Q, P^Z = R$. □

Bemerkung Die Faltung ist allgemein auch für (signierte) Maße definiert, womit man sogar eine Faltungs-Algebra im Sinne der Algebra erhält.

Lemma II.31
Es seien X und Y unabhängige reellwertige Zufallsvariablen und X besitze eine W-Dichte f^X. Dann besitzt $X+Y$ die Dichte $f^{X+Y}(z) = \int_{\mathbb{R}} f^X(z-y)P^Y(\,dy)$. Falls auch Y eine Dichte f^Y besitzt, so gilt $f^{X+Y} = f^X * f^Y$, wobei die Faltung zweier integrierbarer Funktionen f,g definiert ist als $(f*g)(y) = \int_{\mathbb{R}} f(y-x)g(x)\,dx$.

Beweis:

$$
\begin{aligned}
\forall b \in \mathbb{R} : (P^X * P^Y)((-\infty, b]) &\overset{\text{Def.}}{=} \int_{\mathbb{R}} P^X((-\infty, b-y])P^Y(\,dy) = \int_{\mathbb{R}} \left(\int_{-\infty}^{b-y} f^X(x)\,dx \right) P^Y(\,dy) \\
&\overset{\text{Fubini}}{=} \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}(x \le b-y) f^X(x) P^Y(\,dy)\,dx \\
&\underset{z=x+y}{\overset{\text{Substitution}}{=}} \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}(z \le b) f^X(z-y) P^Y(\,dy)\,dz \\
&= \int_{-\infty}^{b} \left(\int_{\mathbb{R}} f^X(z-y) P^Y(\,dy) \right) dz
\end{aligned}
$$

Wegen $P^{X+Y} = P^X * P^Y$ besitzt $X+Y$ also die Verteilungsfunktion

$$
F^{X+Y}(b) = \int_{-\infty}^{b} \left(\int_{\mathbb{R}} f^X(z-y) P^X(\,dy) \right) dz
$$

und somit Dichte $f^{X+Y}(z) = \int_{\mathbb{R}} f^X(z-y)P^Y(\,dy)$. Im Fall, dass Y die Dichte f^Y besitzt, gilt nach Integrationstheorie „$P^Y(\,dy) = f^Y(y)\,dy$“ d.h.

$$
f^{X+Y}(z) = \int f^X(z-y) f^Y(y)\,dy = f^X * f^Y(z)
$$

□

Beispiel II.32 Es seien X und Y unabhängig mit $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$, $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$. Wie ist $X+Y$ verteilt? Es ist nun klar, dass $X+Y$ wieder eine Dichte hat, nämlich $f^{X+Y} = f^X * f^Y$.

$$
\begin{aligned}
&f^X * f^Y(z) \\
&= \int \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{(z-y-\mu_x)^2}{2\sigma_x^2} \right) \cdot \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y^2} \right) dy \\
&\underset{\text{konst.}}{\overset{c_i>0}{=}} c_1 \int \exp\left(-\frac{((z-\mu_x)-y)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2} \right) dy \\
&= c_1 \exp\left(-\frac{(z-\mu_x-\mu_y)^2}{2\sigma_x^2} \right) \cdot \exp\left(\frac{(z-\mu_x-\mu_y)^2\sigma_y^2}{(\sigma_x^2+\sigma_y^2)\cdot\sigma_x^2} \right) \int \exp\left(-\frac{\left(y - \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}(z-\mu_x-\mu_y) \right)^2}{2\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2+\sigma_y^2}} \right) dy \\
&= c_2 \cdot \exp\left(-\frac{1}{2}(z-\mu_x-\mu_y)^2 \cdot \left(\frac{1}{\sigma_x^2} - \frac{\sigma_y^2}{\sigma_x^2(\sigma_x^2+\sigma_y^2)} \right) \right) = c_2 \exp\left(-\frac{1}{2} \frac{(z-\mu_x-\mu_y)^2}{\sigma_x^2+\sigma_y^2} \right)
\end{aligned}
$$

Da $f^X * f^Y$ wieder eine W-Dichte sein muss, ist die Normierungskonstante $c_2 = \frac{1}{\sqrt{2\pi(\sigma_x^2+\sigma_y^2)}}$ und es gilt $X+Y \sim \mathcal{N}(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$.

III. Erwartungswert, Varianz und Kovarianz

III.1. Der Erwartungswert und Momente

Wir wollen für reellwertige Zufallsvariablen einen Mittelwert als Kenngröße angeben. Dafür gibt es mehrere Möglichkeiten, aber wie bei deterministischen Größen ist das Analogon des arithmetischen Mittels, nämlich der Erwartungswert, die gängigste Wahl.

Beispiel III.1 Beim Würfeln von zwei Würfeln erwarten wir als Augensumme X im Mittel

$$\mathbb{E}[X] = P(X = 2) \cdot 2 + P(X = 3) \cdot 3 + \cdots + P(X = 12) \cdot 12 = 7$$

gewichtetes Mittel der möglichen Werte von X.

Definition III.2
Eine reellwertige Zufallsvariable X heißt einfach, *falls sie nur endlich viele Werte annimmt, d.h. X ist von der Form $X(\omega) = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i}(\omega)$ mit $m \in \mathbb{N}, \alpha_i \in \mathbb{R}, A_i \in \mathcal{F}$ geeignet. Für eine solche Zufallsvariable definieren wir ihren* Erwartungswert

$$\mathbb{E}[X] := \mathbb{E}X := \sum_{i=1}^{m} \alpha_i P(A_i)$$

Lemma III.3
Für einfache Zufallsvariablen X auf $(\Omega, \mathcal{F}, P)$ gilt:

(a) $\mathbb{E}[X] = \sum\limits_{x \in X(\Omega)} xP(X = x)$*; insbesondere hängt der Erwartungswert* nur *von der Verteilung P^X von X ab.*

(b) Der Erwartungswert ist linear und monoton: Ist Y eine weitere einfache Zufallsvariable und sind $\alpha, \beta \in \mathbb{R}$, so gilt $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$ und aus $X \leq Y$ (d.h. $\forall \omega \in \Omega : X(w) \leq Y(\omega)$) folgt $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

(c) Falls X und Y unabhängige einfache Zufallsvariablen sind, so gilt $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(d) $\forall A \in \mathcal{F} : P(A) = \mathbb{E}[\mathbf{1}_A]$

Beweis: (a) o.B.d.A. A_i paarweise disjunkt

$$\begin{aligned}\mathbb{E}[X] &\overset{\text{Def.}}{=} \sum_{i=1}^{m} \alpha_i P(A_i) = \sum_{i=1}^{m} \alpha_i P(A_i \cap X = \alpha_i) = \sum_{x \in X(\Omega)} x \sum_{i=1}^{m} \mathbf{1}_{\{x=\alpha_i\}} P(A_i \cap X = \alpha_i) \\ &= \sum_{x \in X(\Omega)} xP(X = x) = P^X(x)\end{aligned}$$

hängt dies nur von der Verteilung P^X ab.

(b) Zu $X = \sum_{i=1}^{m} \alpha_i \mathbf{1}_{A_i}, Y = \sum_{j=1}^{n} \beta_j \mathbf{1}_{B_j}$ wähle eine gemeinsame Verteilung $(C_k)_{1 \leq k \leq r}$ der $(A_i), (B_i)$, sodass $X = \sum_{k=1}^{r} \tilde{\alpha}_k \mathbf{1}_{C_k}, Y = \sum_{k=1}^{r} \tilde{\beta}_k \mathbf{1}_{C_k}$. Da nach a) der Erwartungswert nicht von der Darstellung von X abhängt, gilt:

$$\begin{aligned}\mathbb{E}[\alpha X+\beta Y] =&\mathbb{E}\left[\sum_{k=1}^r(\alpha\tilde{\alpha}_k+\beta\tilde{\beta}_k)\mathbf{1}_{C_k}\right]=\sum_{k=1}^r(\alpha\tilde{\alpha}_k+\beta\tilde{\beta}_k)P(C_k)\\ =&\alpha\sum_{k=1}^r\tilde{\alpha}_kP(C_k)+\beta\sum_{k=1}^r\tilde{\beta}_kP(C_k)=\alpha\mathbb{E}[X]+\beta\mathbb{E}[Y]\end{aligned}$$

Genauso folgt $\mathbb{E}[X]=\sum_{k=1}^r\tilde{\alpha}_kP(C_k)\leq\sum_{k=1}^r\tilde{\beta}_kP(C_k)=\mathbb{E}[Y]$ für $X\leq Y$ und o.B.d.A (C_k) paarweise disjunkt $\implies\tilde{\alpha}_k\leq\tilde{\beta}_k$

(c) $\mathbb{E}[XY]\overset{(a)}{=}\sum_{z\in(XY)(\Omega)}zP(XY=z)=\sum_{\substack{z\in(XY)(\Omega)\\ z\neq0}}z\sum_{\substack{x\in X(\Omega)\\ x\neq0}}P(XY=z,X=x)$

$$\begin{aligned}&\overset{\text{Unabhängig}}{=}\sum_{\substack{x\in X(\Omega)\\ x\neq0}}\sum_{\substack{z\in(XY)(\Omega)\\ z\neq0}}zP(X=x)P(Y=\tfrac{z}{x})\\ &=\sum_{\substack{x\in X(\Omega)\\ x\neq0}}\sum_{\substack{y\in Y(\Omega)\\ y\neq0}}xyP(X=x)P(Y=y)\\ &=\sum_{\substack{x\in X(\Omega)\\ x\neq0}}xP(X=x)\sum_{\substack{y\in Y(\Omega)\\ y\neq0}}yP(Y=y)=\mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$

(d) klar nach Def. □

Beispiel III.4 Sei X eine $\mathrm{Bin}(n,p)$-verteilte Zufallsvariable

$$\begin{aligned}\mathbb{E}[X]\overset{\text{Def.}}{=}&\sum_{k=0}^nkP(X=k)=\sum_{k=0}^nk\binom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^n\frac{n!}{(n-k)!(k-1)!)}p^k(1-p)^{n-k}\\ =&n\sum_{k=1}^n\binom{n-1}{k-1}p^k(1-p)^{n-k}\overset{l=k-1}{=}n\sum_{l=0}^{n-1}\binom{n-1}{l}p^{l+1}(1-p)^{n-1-l}\\ =&np\sum_{l=0}^{n-1}\binom{n-1}{l}p^l(1-p)^{(n-1)-l}=np\end{aligned}$$

Zweite Möglichkeit: es gilt ja $X=\sum_{i=1}^nZ_i$ mit unabhängigen $Z_i\sim\mathrm{Bin}(1,p)$. Linearität liefert $\mathbb{E}[X]=\mathbb{E}[\sum_{i=1}^nZ_i]=\sum_{i=1}^n\mathbb{E}[Z_i]=\sum\sum_{k=0}^1kP(Z_i=k)=\sum_{i=1}^n1p=np$

Definition III.5
Für eine nicht negative Zufallsvariable X setze mittels approximierender Folge (X_n) einfacher Zufallsvariablen mit $X_n\uparrow X$

$$\mathbb{E}[X]:=\lim_{n\to\infty}\mathbb{E}[X_n]\in[0,\infty]$$

Betrachte für einen Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},P)$ die Menge

$$\mathcal{L}^1=\mathcal{L}^1(\Omega,\mathcal{F},P)=\{X:\omega\to\mathbb{R}\textit{ messbar},\ \mathbb{E}[|X|]<\infty\}$$

Definiere für $X \in \mathcal{L}^1$ mit $X^+(\omega) = \max(X(\omega), 0)$ und $X^-(\omega) = \max(-X(\omega), 0)$ den Erwartungswert

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-] \in \mathbb{R}$$

Wegen $\mathbb{E}[X^+] \leq \mathbb{E}[|X|] < \infty$ und $\mathbb{E}[X^-] \leq \mathbb{E}[|X|] < \infty$ gilt für $X \in \mathcal{L}^1$ stets $|\mathbb{E}[X]| < \infty$. Man schreibt auch

$$\mathbb{E}[X] = \int_\Omega X dP = \int_\Omega X(\omega) P(d\omega) = \int_\Omega X(\omega) dP(\omega)$$

und setzt $\int_A X dP = \int_\Omega X(\mathbf{1}_A) dP$ für $A \in \mathcal{F}$.

Satz III.6
Für $X \in \mathcal{L}^1(\Omega, F, P)$ gilt

(a) *$\mathbb{E}[X] = \int_\Omega x P^X(dx)$ und $\mathbb{E}[X]$ hängt nur von der Verteilung von X ab.*

(b) *Der Erwartungswert ist linear und monoton auf $\mathcal{L}^1$.*

(c) *Falls $X, Y \in \mathcal{L}^1$ unabhängig sind, so gilt $XY \in \mathcal{L}^1$ und $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$*

Beweis: (a),(b) folgen aus der Maßtheorie (Lemma + Approximation). Für (c) betrachte

$$\begin{aligned}
\mathbb{E}[|XY|] &\overset{\text{(a)}}{=} \int_{\mathbb{R}^2} |xy| P^{(XY)}(dx, dy) \overset{XY}{=} \int_{\mathbb{R}^2} |x||y| (P^X \otimes P^Y)(dx, dy) \\
&\overset{\text{Tonelli}}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} |x||y| P^X(dx) P^Y(dy) = \left(\int_{\mathbb{R}} |x| P^X(dx) \right) \left(\int_{\mathbb{R}} |y| P^Y(dy) \right) \\
&= \mathbb{E}[|X|]\mathbb{E}[|Y|] \overset{X,Y \in \mathcal{L}^1}{<} \infty \implies XY \in \mathcal{L}^1
\end{aligned}$$

Dieselbe Rechnung ohne Betrag zeigt dann $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ (Fubini!) □

Korollar III.7

(a) *Ist X eine Zufallsvariable mit Wahrscheinlichkeitsdichte $f^X : \mathbb{R} \to [0, \infty)$, so gilt*

$$X \in \mathcal{L}^1 \iff \int_{-\infty}^{\infty} |x| f^X dx < \infty \quad \text{und in diesem Fall} \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} x f^X dx$$

(b) *Ist X eine Zufallsvariable mit Zähldichte p^X, so gilt*

$$X \in \mathcal{L}^1 \iff \sum_{x \in X(\Omega)} |x| p^X(x) < \infty \quad \text{und in diesem Fall} \quad \mathbb{E}[X] = \sum_{x \in X(\Omega)} x p^X(x)$$

Beweis: (a) Sei $X \geq 0$, d.h. $f^X(x) = 0$ für $x < 0$. Dann gilt für

$$X_n = \begin{cases} k2^{-n}, & \text{falls } k2^{-n} \leq X < (k+1)2^{-n}, 0 \leq k \leq n2^n - 1 \\ n, & \text{falls } X \geq n \end{cases}$$

mit $X_n \uparrow X$, X_n einfach und

$$\begin{aligned}
\mathbb{E}[X_n] &= \sum_{k=0}^{n2^n-1} k2^{-n} P\big(k2^{-n} \le X \le (k+1)2^{-n}\big) + nP(X \ge n) \\
&= \sum_{k=0}^{n2^n-1} \int_{k2^{-n}}^{(k+1)2^{-n}} k2^{-n} f^X(x)dx + \int_n^\infty n f^X(x)dx \\
&= \int_0^\infty \left(\sum_{k=0}^{n2^n-1} k2^{-n} \mathbf{1}_{[k2^{-n},(k+1)2^{-n})}(x) + n\mathbf{1}_{[n,\infty)}(x) \right) f^X(x)dx \\
\xRightarrow[\text{Beppo-Levi}]{\text{mon. Konv.}} \mathbb{E}[X] &= \int_0^\infty x f^X(x)dx
\end{aligned}$$

Dies zeigt insbesondere $\mathbb{E}[|X|] = \int_0^\infty x f^{|X|}(x)dx = \int_0^\infty x(f^X(x) + f^X(-x))dx$ und dies ist endlich genau dann, wenn $\int_{-\infty}^\infty |x| f^X(x)dx < \infty$.
Ist nun $X \in \mathcal{L}^1$, so folgt

$$\begin{aligned}
\mathbb{E}[X^+] &= \int_0^\infty x f^{X^+}(x)dx = \int_0^\infty x f^X(x)dx \\
\mathbb{E}[X^-] &= \int_0^\infty x f^{X^-}(x)dx = \int_0^\infty x f^X(-x)dx = -\int_{-\infty}^0 x f^X(x)dx \\
\Longrightarrow \mathbb{E}[X] &= \mathbb{E}[X^+] - \mathbb{E}[X^-] = \int_{-\infty}^\infty x f^X(x)dx
\end{aligned}$$

(b) Setze

$$X_n = \begin{cases} k2^{-n} & , k2^{-n} \le X < (k+1)2^{-n}, k = 0, \dots, n2^n - 1 \\ n & , X \ge n \\ 0 & , X < 0 \end{cases}$$

also eine einfache Zufallsvariable. Dann ist

$$\begin{aligned}
\mathbb{E}[X_n^+] &= \left(\sum_{k=0}^{n2^n-1} k2^{-n} P(X \in [k2^{-n}, (k+1)2^{-n})) \right) + nP(X \ge n) + 0 \\
&= \left(\sum_{k=0}^{n2^n-1} k2^{-n} \cdot \sum_{x \in [k2^{-n},(k+1)2^{-n}) \cap X(\Omega)} p^X(x) \right) + n \sum_{x \in [n,\infty) \cap X(\Omega)} p^X(x) \\
&= \sum_{x \in X(\Omega)} p^X(x) \underbrace{\left[\left(\sum_{k=0}^{n2^n-1} k2^{-n} \mathbf{1}(k2^{-n} \le x < (k+1)2^{-n}) \right) + n\mathbf{1}(x \ge n) \right]}_{\text{monoton wachsend in } n \text{ mit Limes } \max\{x,0\}} \\
&\xrightarrow{n \to \infty} \sum_{x \in X(\Omega)} \max\{x, 0\} \cdot p^X(x)
\end{aligned}$$

Also gilt $\mathbb{E}[X_n^+] \to \sum_{x \in X(\Omega)} \max(x,0) p^X(x)$ und wegen $X_n^+ \uparrow \max\{X, 0\} = X^+$

nach Definition

$$\mathbb{E}[X^+] = \sum_{x\in X(\Omega)} \max\{x,0\}p^X(x) \quad (\text{analog } \mathbb{E}[X^-] = \sum \max\{-x,0\}p^X(x))$$

$$\Longrightarrow \mathbb{E}[|X|] = \mathbb{E}[X^+ + X^-] = \sum_{x\in X(\Omega)} |x|p^X(x)$$

und es gilt $\mathbb{E}[|X|] < \infty \iff \sum_{x\in X(\Omega)} |x|p^X(x) < \infty$. In dem Fall sind auch $\mathbb{E}[X^+], \mathbb{E}[X^-]$ endlich und somit

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-] = \sum_{x\in X(\Omega)} (\max\{x,0\} - \max\{-x,0\})p^X(x) = \sum_{x\in X(\Omega)} xp^X(x)$$

□

Satz III.8
Es seien X ein Zufallsvektor mit Dichte $f^X : \mathbb{R}^d \to [0,\infty)$ sowie $h : \mathbb{R}^d \to \mathbb{R}$ Borel-messbar. Dann gilt

$$h(X) \in \mathcal{L}^1 \iff \int_{\mathbb{R}^d} |h(x)|f^X(x)\,dx < \infty$$

In dem Fall gilt $\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x)f^X(x)\,dx$.

Beweis: (maßtheoretische Induktion)

(a) h einfache Funktion: $h = \sum_{i=1}^m \alpha_i \mathbf{1}_{A_i} \Longrightarrow h \circ X = \sum_{i=1}^m \alpha_i \mathbf{1}_{X^{-1}(A_i)}$

$$\begin{aligned}\mathbb{E}[h(X)] &\overset{\text{linear}}{=} \sum_{i=1}^m \alpha_i \mathbb{E}[\mathbf{1}_{X^{-1}(A_i)}] = \sum_{i=1}^m \alpha_i P(X\in A_i) = \sum_{i=1}^m \alpha_i \int_{\mathbb{R}^d} \mathbf{1}_{A_i}(x)f^X(x)\,dx \\ &= \int_{\mathbb{R}^d} h(x)f^X(x)\,dx\end{aligned}$$

(b) $h \geq 0$ messbar: Dann gibt es einfache h_n mit $h_n \uparrow h$:

$$\mathbb{E}[h_n(X)] \overset{(a)}{=} \int h_n(x)f^X(x)\,dx \underset{\text{Konv.}}{\overset{\text{mon.}}{\uparrow}} \int h(X)f^X(x)\,dx \overset{\text{Def}}{=} \mathbb{E}[h(X)] \in [0,\infty]$$

Wir schließen daraus, dass $\mathbb{E}[|h(X)|] < \infty \iff \int |h(x)|f^X(x)\,dx < \infty$ für beliebige messbare h.

(c) h beliebig messbar mit $\int |h|f^X(x)\,dx < \infty$:

$$\begin{aligned}\mathbb{E}[h(X)] &\overset{\text{Def}}{=} \mathbb{E}[h^+(X)] - \mathbb{E}[h^-(X)] \overset{(b)}{=} \int h^+(x)f^X(x)\,dx - \int h^-(x)f^X(x)\,dx \\ &= \int h(x)f^X(x)\,dx\end{aligned}$$

□

Beispiel III.9 (a) $X \sim \mathrm{U}[0,1] \Longrightarrow \mathbb{E}[X^k] \overset{k\in\mathbb{N}}{=} \int_{\mathbb{R}} x^k \cdot f^X(x)\,d\,x = \int_0^1 x^k\,d\,x = \frac{1}{k+1}$, insbesondere $\mathbb{E}[X] = \frac{1}{2}$

(b) Zufallsvariable X mit Zähldichte auf $\mathbb{Z}$: $p^X(m) = \frac{3}{\pi^2}\cdot\frac{1}{m^2}$ für $m \in \mathbb{Z}\setminus\{0\}$ und $p^X(0) = 0$. Was ist $\mathbb{E}[X]$? Existiert nicht, weil $\mathbb{E}[|X|] = \sum_{m\in\mathbb{Z}} |m|\cdot p^X(x) = 2\sum_{m=1}^\infty m\frac{3}{\pi^2}\cdot\frac{1}{m^2} = \infty$.

Definition III.10
Eine Zufallsvariable X liegt im $\mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, P)$, für $p > 0$, falls $|X|^p \in \mathcal{L}^1$, also $\mathbb{E}[|X|^p] < \infty$ ist. Für $X \in \mathcal{L}^p$ und $p \in \mathbb{N}$ heißt $\mathbb{E}[X^p]$ p-tes Moment *von X; für $X \in \mathcal{L}^p$ und $p > 0$ heißt $\mathbb{E}[|X|^p]$* p-tes absolutes Moment.

Lemma III.11
Für $0 < p \le q$ gilt $\mathcal{L}^q \subseteq \mathcal{L}^p$.

Beweis: $\mathbb{E}[|X|^p] \overset{\text{Mon.}}{\le} \mathbb{E}[\max\{1, |X|^q\}] \overset{\text{Mon.}}{\le} \mathbb{E}[1 + |X|^q] \overset{\text{Lin.}}{=} 1 + \mathbb{E}[|X|^q] < \infty$ für $X \in \mathcal{L}^q, p \le q$. □

Lemma III.12
Für eine Zufallsvariable $X \in \mathcal{L}^2$ nimmt die Funktion $\varphi : \mathbb{R} \to \mathbb{R}$ mit $\varphi(x) = \mathbb{E}[(X - x)^2]$ ihr Minimum bei $x = \mathbb{E}[X]$ an.

Beweis: Wegen $\mathcal{L}^2 \subset \mathcal{L}^1$ ist $\mathbb{E}[X]$ wohldefiniert und

$$\mathbb{E}[(X - x)^2] = \mathbb{E}[X^2 - 2xX + x^2] = \mathbb{E}[X^2] - 2x\mathbb{E}[X] + x^2$$

endlich. Wir erhalten

$$\begin{aligned}
\varphi(x) &= \mathbb{E}[((X - \mathbb{E}[X]) + \underbrace{(\mathbb{E}[X] - x)}_{\text{deterministisch}})^2] \\
&= \mathbb{E}[(X - \mathbb{E}[X])^2] + 2(\mathbb{E}[X] - x)\underbrace{\mathbb{E}[X - \mathbb{E}[X]]}_{=0} + (\mathbb{E}[X] - x)^2 \\
&= \varphi(\mathbb{E}[X]) + \underbrace{(\mathbb{E}[X] - x])^2}_{\ge 0} \ge \varphi(\mathbb{E}[X])
\end{aligned}$$

□

Bemerkung (a) Das Minimum ist sogar eindeutig.

(b) Man versteht $\mathbb{E}[(X - x)^2]$ als den mittleren quadratischen Fehler beider Vorhersagen von X durch den reellen Wert x. Dieser Fehler ist minimal für den Erwartungswert.

(c) In der Statistik heißt die Identität

$$\varphi(x) = \underbrace{(x - \mathbb{E}[X])^2}_{\text{Bias}^2} + \underbrace{\mathbb{E}[(X - \mathbb{E}[X])^2]}_{\text{Varianz}}$$

Bias-Varianz-Zerlegung.

III.2. Varianz, Kovarianz und Korrelation

Das Minimum von φ im vorigen Lemma misst gerade die *Streuung* der Zufallsvariablen X um den (besten) deterministischen Wert $\mathbb{E}[X]$.

Definition III.13
Für eine Zufallsvariable $X \in \mathcal{L}^2$ bezeichnet $\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$ *die* Varianz *und* $\sigma(X) = \sqrt{\mathrm{Var}(X)}$ *die* Standardabweichung *von X.*

Satz III.14 (Eigenschaften)
Für $X, Y \in \mathcal{L}^2$ gilt

(a) $\operatorname{Var}(X) = 0 \iff P(X = c) = 1$ *für ein* $c \in \mathbb{R}$.

(b) $\forall \alpha, \beta \in \mathbb{R} : \operatorname{Var}(\alpha X + \beta) = \alpha^2 \cdot \operatorname{Var}(X)$

(c) $\operatorname{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$

(d) $\operatorname{Var}(X + Y) \le 2(\operatorname{Var}(X) + \operatorname{Var}(Y))$

(e) Falls X, Y unabhängig *sind, gilt* $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

Beweis: (a) $\operatorname{Var}(X) = 0 \iff \mathbb{E}[(X - \mathbb{E}[X])^2] = 0 \iff P(X - \mathbb{E}[X] = 0) = 1 \iff P(X = \mathbb{E}[X]) = 1$ und damit $\mathbb{E}[X] = c$.

(b) $\operatorname{Var}(\alpha X + \beta) = \mathbb{E}[(\alpha X + \cancel{\beta} - (\alpha \mathbb{E}[X] + \cancel{\beta}))^2] = \alpha^2 \mathbb{E}[(X - \mathbb{E}[X])^2] = \alpha^2 \operatorname{Var}(X)$

(c) $\operatorname{Var}(X) = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \cancel{2}\mathbb{E}[X]^2 + \cancel{\mathbb{E}[X]^2} = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

(d) $\operatorname{Var}(X + Y) = \mathbb{E}[(X - \mathbb{E}[X] + Y - \mathbb{E}[Y])]^2] \le \mathbb{E}[2(X - \mathbb{E}[X])^2 + 2(Y - \mathbb{E}[Y])^2] = 2(\operatorname{Var}(X) + \operatorname{Var}(Y)$.

(e)
$$
\begin{aligned}
\operatorname{Var}(X + Y) &= \mathbb{E}[(X - \mathbb{E}[X])^2 + 2(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) + (Y - \mathbb{E}[Y])^2] \\
&\overset{X,Y \text{ unabh.}}{=} \mathbb{E}[(X - \mathbb{E}X])^2] + 2\underbrace{\mathbb{E}[X - \mathbb{E}[X]]}_{=0} \cdot \mathbb{E}[Y - \mathbb{E}[Y]] + \mathbb{E}[(Y - \mathbb{E}[Y])^2] \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y)
\end{aligned}
$$
□

Gegeben seien nun zwei Zufallsvariablen $X, Y \in \mathcal{L}^2$. Wir fragen nach der besten in X linearen Vorhersage von Y.

Satz III.15
Für $X, Y \in \mathcal{L}^2$ sei $L_X := \{aX + b \mid a, b \in \mathbb{R}\} \subseteq \mathcal{L}^2$ die Menge der linearen affinen Funktionen in X. Dann nimmt der mittlere quadratische Fehler $\varphi : L_X \to [0, \infty)$ mit $\varphi(Z) = \mathbb{E}[(Y - Z)^2]$ sein Minimum an bei $Z = a^* X + b^*$ mit

$$a^* = \frac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\operatorname{Var}(X)}$$

bzw. beliebig falls $\operatorname{Var}(X) = 0$

$$b^* = \mathbb{E}[Y - a^* \mathbb{E}[X]]$$

an. Für $\operatorname{Var}(X) > 0$ *gilt* $\varphi(a^* X + b^*) = \operatorname{Var}(Y) - \dfrac{(\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])])^2}{\operatorname{Var}(X)}$.

Beweis: Setze $\tilde{X} = X - \mathbb{E}[X]$, $\tilde{Y} = Y - \mathbb{E}[Y]$. Dann gilt für jedes $Z = aX + b$,

$$\varphi(Z) = \mathbb{E}[((\tilde{Y} - a\tilde{X}) + \mathbb{E}[Y] - a\mathbb{E}[X] - b)^2] = \mathbb{E}[(\tilde{Y} - a\tilde{X})^2] + (\mathbb{E}[Y] - a\mathbb{E}[X] - b)^2 + 2 \cdot 0$$

$\implies b^* = \mathbb{E}[Y] - a^* \mathbb{E}[X]$. Damit ist $\varphi(a^* X + b^*) = \mathbb{E}[\tilde{Y}^2] + (a^*)^2 \mathbb{E}[\tilde{X}^2] - 2a^* \mathbb{E}[\tilde{X}\tilde{Y}]$

$\implies$ Minimieren in a^* gibt $a^* = \dfrac{\mathbb{E}[\tilde{X}\tilde{Y}]}{\mathbb{E}[\tilde{X}^2]} = \dfrac{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}{\operatorname{Var}(X)}$ falls $\operatorname{Var}(X) > 0$.

Im Fall das $\operatorname{Var}(X) = 0$ gilt $\varphi(a^* X + b^*) = \mathbb{E}[\tilde{Y}^2]$ für alle a^*.

Als Minimum erhalten wir $\varphi(a^*X+b^*) = \mathbb{E}[\tilde{Y}^2] - \frac{\mathbb{E}[\tilde{X}\tilde{Y}]}{\mathbb{E}[\tilde{X}^2]}$ wie behauptet. □

Definition III.16
Für Zufallsvariablen $X, Y \in \mathcal{L}^2$ definiert $\mathrm{Cov}(X,Y) := \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$ *die* Kovarianz *von X und Y. Falls außerdem $\sigma(X) > 0, \sigma(Y) > 0$ gilt, so heißt*

$$\varrho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$$

die Korrelation *von X und Y. Im Fall* $\mathrm{Cov}(X,Y) = 0$ *heißen X und Y* unkorreliert.

Bemerkung Im Satz gilt also $a^* = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$ und das Minimum ist $\mathrm{Var}(Y)(1-\varrho^2(X,Y))$.

Satz III.17 (Eigenschaften der Kovarianz)
Für $X, Y, Z \in \mathcal{L}^2$ gilt

(a) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$; $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

(b) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

(c) $\forall a, b \in \mathbb{R} : \mathrm{Cov}(aX+b, Y) = a\,\mathrm{Cov}(X,Y)$

(d) $\mathrm{Cov}(X+Y, Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z)$

(e) Falls X, Y unabhängig sind, so sind X, Y unkorreliert

(f) $|\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y)$ *und* $|\varrho(X,Y)| \leq 1$ *(d.h. $\varrho(X,Y) \in [-1,1]$)*

Beweis:

(a) $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ und $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$ klar nach Definition

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X\mathbb{E}[Y]] - \mathbb{E}[\mathbb{E}[X]Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned}$$

(b) $\mathrm{Var}(X+Y) = \mathbb{E}[(X-\mathbb{E}[X]+Y-\mathbb{E}[Y])^2] = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$

(c) $\mathrm{Cov}(aX+b, Y) = \mathbb{E}[(aX+b-(a\mathbb{E}[X]+b))(Y-\mathbb{E}[Y])] = a\,\mathrm{Cov}(X,Y)$

(d) $$\begin{aligned} \mathrm{Cov}(X+Y,Z) &= \mathbb{E}[((X-\mathbb{E}[X])+(Y-\mathbb{E}[Y]))(Z-\mathbb{E}[Z])] \\ &= \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z) \end{aligned}$$

(e) X, Y unabhängig $\overset{\text{s.o.}}{\Longrightarrow} \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \overset{\text{(b)}}{\Longrightarrow} \mathrm{Cov}(X,Y) = 0$

(f) Im Satz haben wir $\varphi(a^*X+b^*) = \mathrm{Var}(Y)(1-\varrho^2(X,Y))$ gezeigt, was wegen $\varphi(Z) > 0$ für alle Z impliziert $\varrho^2(X,Y) \leq 1$, d.h. $\varrho \in [-1,1]$ und $|\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y)$. Beachte nur, dass im Fall $\mathrm{Var}(X) = 0$ oder $\mathrm{Var}(Y) = 0$ gilt $P(X = \mathbb{E}(X)) = 1$ bzw. $P(Y = \mathbb{E}(Y)) = 1$, sodass $P((X-\mathbb{E}[X])(Y-\mathbb{E}[Y]) = 0) = 1$ gilt und somit $\mathrm{Cov}(X,Y) = 0$ folgt. □

Bemerkung Teil f) ist äquivalent zur Cauchy-Schwarz-Ungleichung für $X, Y \in \mathcal{L}^2$

$$|\mathbb{E}[XY]| \leq \mathbb{E}[X^2]^{\frac{1}{2}}\mathbb{E}[Y^2]^{\frac{1}{2}}$$

Beispiel III.18 Es seien $X_1, X_2 \in \mathcal{L}^2$ unabhängig mit $\mathrm{Var}(X_1) = \mathrm{Var}(X_2)$. Setze $S = X_1 + X_2$, $D = X_1 - X_2$. Dann gilt

$$\begin{aligned}\mathrm{Cov}(S, D) &= \mathrm{Cov}(X_1 + X_2, X_1 - X_2)\\ &= \mathrm{Cov}(X_1, X_1) + \mathrm{Cov}(X_2, X_1) - \mathrm{Cov}(X_1, X_2) - \mathrm{Cov}(X_2, X_2)\\ &= \mathrm{Var}(X_1) - \mathrm{Var}(X_2) = 0 \implies \text{S und D sind unkorreliert.}\end{aligned}$$

Beim Würfelwurf sind aber S und D nicht unabhängig: X_i Augensumme des i-ten Würfels. Beste Vorhersage von D, im Fall das $S = s$ bekannt ist gerade Null $(= \mathbb{E}[D])(\implies a^* = 0, b^* = 0)$. S und D sind aber abhängig, z.B. gilt S ungerade $\iff$ D ungerade oder aber $S = 2 \implies D = 0$

$$\left(P(S = 2, D = 0) = P(S = 2) = \frac{1}{36} \neq P(S = 2)P(D = 0) < \frac{1}{36}\right)$$

III.3. Mehrdimensionale Normalverteilung

Zur Erinnerung: Eine Matrix $M \in \mathbb{R}^{d\times d}$ ist *positiv semi-definit*, falls $\forall v \in \mathbb{R}^d : \langle Mv, v\rangle \geq 0$. Die Wurzel $M^{\frac{1}{2}}$ aus M ist diejenige positiv semi-definite Matrix N mit $M = N^2$. Die Existenz wird via Diagonalisierung gezeigt: $M = ODO^\top$, O orthogonal, D diagonal $\implies M^{\frac{1}{2}} = OD^{\frac{1}{2}}O^\top$ mit $D = \mathrm{diag}(\lambda_1, \dots, \lambda_d)$, $D^{\frac{1}{2}} = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_d})$.

Definition III.19
Es seien $\mu \in \mathbb{R}^d$ und $\Sigma \in \mathbb{R}^{d\times d}$ symmetrisch und positiv semi-definit. Ein Zufallsvektor $X \in \mathbb{R}^d$ heißt $\mathcal{N}(\mu, \Sigma)$-verteilt, falls $X = \mu + \Sigma^{\frac{1}{2}}Y$ gilt mit einem Standard-normalverteilten Zufallsvektor $Y \sim \mathcal{N}(0, E_\lambda)$. $\mathcal{N}(\mu, \Sigma)$ heißt d-dimensionale Normalverteilung *mit* Mittelwert μ *und* Kovarianzmatrix Σ.

Bemerkung Für $d = 1$ und $\Sigma = \sigma^2 > 0$ liefert die Dichtetransformation , dass X die Dichte $f^X(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ besitzt. Im Fall $\Sigma = \sigma^2 = 0$ gilt $X = \mu$ und $\mathcal{N}(\mu, 0) = \delta_\mu$ (Einpunktmaß in μ).

Lemma III.20
Für $X \sim \mathcal{N}(\mu, \Sigma)$ und $1 \leq l, k \leq d$ gilt:

$$\mathbb{E}[X_k] = \mu_k, \mathrm{Cov}(X_k, X_l) = \Sigma_{kl} \qquad \left(\mu = (\mu_k)_{1\leq k\leq d}, \Sigma = (\Sigma_{kl})_{1\leq l,k\leq d}\right)$$

Beweis:

$$\mathbb{E}[X_k] \overset{\text{Def}}{=} \mathbb{E}[(\mu_k + (\Sigma^{\frac{1}{2}}Y)_k)] = \mu_k + \sum_{l=1}^{d}(\Sigma_{kl}^{\frac{1}{2}})\mathbb{E}[Y_l] = \mu_k$$

$$\begin{aligned}\mathrm{Cov}(X_k, X_l) &= \mathbb{E}[(X_k - \mu_k)(X_l - \mu_l)] \overset{\text{Def}}{=} \mathbb{E}[(\Sigma^{\frac{1}{2}}Y)_k * (\Sigma^{\frac{1}{2}}Y)_l]\\ &= \sum_{i,j=1}^{d}\Sigma_{k,i}^{\frac{1}{2}}\Sigma_{l,j}^{\frac{1}{2}}\mathbb{E}[Y_iY_j] = \sum_{i=1}^{d}\Sigma_{k,i}^{\frac{1}{2}}\Sigma_{l,i}^{\frac{1}{2}} = (\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}})_{k,l}\\ &= \Sigma_{k,l}\end{aligned}$$

□

Lemma III.21
Ist $\Sigma \in \mathbb{R}^{d\times d}$ symmetrisch und strikt positiv definit, so besitzt die $\mathcal{N}(\mu,\Sigma)$-Verteilung eine Dichte, nämlich

$$\varphi_{\mu,\Sigma}(x) = (2\pi)^{-\frac{d}{2}} \cdot \det(\Sigma)^{-\frac{1}{2}} \cdot \exp\left(-\tfrac{1}{2}\left\langle \Sigma^{-1}(x-\mu),(x-\mu)\right\rangle\right), \quad x \in \mathbb{R}^d$$

Beweis: Dichtetransformation für $X = \varphi(Y)$, $Y \sim \mathcal{N}(0,E_d)$, $\psi(y) = \Sigma^{\frac{1}{2}} y + \mu$, $\psi^{-1}(x) = \Sigma^{-\frac{1}{2}}(x-\mu)$

$$\begin{aligned}\implies \varphi_{\mu,\Sigma}(x) &= \varphi_{0,E_d}(\psi^{-1}(x))(\det D_{\psi^{-1}}(x))\\ &= |\det(\Sigma^{-\frac{1}{2}})| \cdot (2\pi)^{-\frac{d}{2}} \exp(-\tfrac{1}{2}|\Sigma^{-\frac{1}{2}}(x-\mu)|^2)\end{aligned}$$

□

Lemma III.22 (Rotationsinvarianz der Normalverteilung)
Ist $\mathcal{O} \in \mathbb{R}^{d\times d}$ orthogonal, d.h. $\mathcal{O}^\top\mathcal{O} = E_d$, so gilt für $X \sim \mathcal{N}(0,\mathrm{E}_d)$, dass auch $\mathcal{O}X \sim \mathcal{N}(0,E_d)$ gilt.

Beweis: Dichtetransformation: $Y = \mathcal{O}X$:

$$f^Y(y) = f^X(\mathcal{O}^{-1}y)|\det(\mathcal{O}^{-1})| = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{|\mathcal{O}^{-1}y|^2}{2}\right)\cdot 1 = \varphi_{0,E_d}(y)$$

□

Satz III.23
Ist X ein $\mathcal{N}(\mu,\Sigma)$-verteilter Zufallsvektor im $\mathbb{R}^d$ und $A \in \mathbb{R}^{m\times d}$ eine deterministische Matrix, so ist $Y = A\cdot X$ ein $\mathcal{N}(A\cdot\mu, A\Sigma A^\top)$-verteilter Zufallsvektor im $\mathbb{R}^m$.

Beweis: Wir müssen zeigen, dass sich Y darstellen lässt als $Y = A\mu + (A\Sigma A^\top)^{\frac{1}{2}}Z$ mit einer geeigneten Zufallsvariablen $Z \sim \mathcal{N}(0,E_m)$. Aus der Darstellung $X = \mu + \Sigma^{\frac{1}{2}}W$ mit $W \sim \mathcal{N}(0,E_d)$ ergibt sich die zu erfüllende Bedingung als

$$A(\mu + \Sigma^{\frac{1}{2}}W) = A\mu + (A\Sigma A^\top)^{\frac{1}{2}}Z \quad \text{d.h.} \quad A\Sigma^{\frac{1}{2}}W = (A\Sigma A^\top)^{\frac{1}{2}}Z$$

Der

Satz III.24 (über orthogonale Transformationen)
zeigt, dass es orthogonale Matrizen $T_1 \in \mathbb{R}^{m\times m}, T_2 \in \mathbb{R}^{d\times d}$ und eine Diagonalmatrix $D \in \mathbb{R}^{r\times r}, r \leq \min\{m,d\}$ mit strikt positiven Diagonaleinträgen gibt, so dass in Blockmatrixnotation (beachte jeweils die Dimension!) $A\Sigma^{\frac{1}{2}} = T_1\left(\begin{smallmatrix} D & 0\\ 0 & 0\end{smallmatrix}\right)T_2$ gilt. Für einen Beweis, siehe [5], Seite 199.

Dies impliziert

$$\begin{aligned}(A\Sigma A^\top)^{\frac{1}{2}} &= \left(A\Sigma^{\frac{1}{2}}\left(A\Sigma^{\frac{1}{2}}\right)^\top\right)^{\frac{1}{2}} = \left(T_1\left(\begin{smallmatrix} D & 0\\ 0 & 0\end{smallmatrix}\right)T_2T_2^\top\left(\begin{smallmatrix} D^\top & 0\\ 0 & 0\end{smallmatrix}\right)T_1^\top\right)^{\frac{1}{2}}\\ &= \left(T_1\left(\begin{smallmatrix} D^2 & 0\\ 0 & 0\end{smallmatrix}\right)T_1^\top\right)^{\frac{1}{2}} = T_1\left(\begin{smallmatrix} D & 0\\ 0 & 0\end{smallmatrix}\right)T_1^\top\end{aligned}$$

Wir müssen also $Z \sim \mathcal{N}(0,E_m)$ finden mit $\left(\begin{smallmatrix} D & 0\\ 0 & 0\end{smallmatrix}\right)T_1^\top Z = \left(\begin{smallmatrix} D & 0\\ 0 & 0\end{smallmatrix}\right)T_2W$. Setze

$$Z := T_1\left(\left(\begin{smallmatrix} E_r & 0\\ 0 & 0\end{smallmatrix}\right)T_2W + W'\right) \quad \text{mit} \quad W' \sim \mathcal{N}\left(0,\left(\begin{smallmatrix} 0 & 0\\ 0 & E_{m-r}\end{smallmatrix}\right)\right)$$

unabhängig von W, ggf. definiert auf einem größeren Wahrscheinlichkeitsraum (bzw. $W' = 0$ falls $m = r$). Aus dem Lemma 3.22 folgt $T_2 W \sim \mathcal{N}(0, E_d)$, weil T_2 eine orthogonale Matrix ist, und weiter, dass $\left(\begin{smallmatrix} E_r & 0 \\ 0 & 0 \end{smallmatrix}\right) T_2 W$ ein $\mathcal{N}(0, \left(\begin{smallmatrix} E_r & 0 \\ 0 & 0 \end{smallmatrix}\right))$-verteilter Vektor im $\mathbb{R}^m$ ist (Projektion auf die ersten r Koordinaten). Daher gilt $\left(\begin{smallmatrix} E_r & 0 \\ 0 & 0 \end{smallmatrix}\right) T_2 W + W' \sim \mathcal{N}(0, E_m)$ und es folgt wieder nach Lemma 3.22 $Z \sim \mathcal{N}(0, E_m)$.

Schließlich ergibt sich für $(A\Sigma A^\top)^{\frac{1}{2}} Z$

$$T_1 \left(\begin{smallmatrix} D & 0 \\ 0 & 0 \end{smallmatrix}\right) \left(\left(\begin{smallmatrix} E_r & 0 \\ 0 & 0 \end{smallmatrix}\right) T_2 W + W' \right) = T_1 \left(\begin{smallmatrix} D & 0 \\ 0 & 0 \end{smallmatrix}\right) T_2 W = A \Sigma^{\frac{1}{2}} W$$

was zu zeigen war. □

Korollar III.25
Ist X $\mathcal{N}(\mu_x, \sigma_x^2)$-verteilt und Y $\mathcal{N}(\mu_y, \sigma_y^2)$-verteilt mit $\mu_x, \mu_y \in \mathbb{R}$ und $\sigma_x, \sigma_y \geq 0$, so ist – falls X und Y unabhängig sind – $X + Y$ $\mathcal{N}(\mu_x + \mu_Y, \sigma_x^2 + \sigma_y^2)$-verteilt.

Bemerkung vergl. obiges Beispiel zur Faltung.

Beweis: Da X, Y unabhängig sind, besitzt der Zufallsvektor (X, Y) die Produktdichte

$$f^{(X,Y)}(x,y) = f^X(x) \cdot f^Y(y) = \frac{1}{2\pi} \cdot \frac{1}{\sigma_x \sigma_y} \exp\left(\frac{1}{2} \left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right) \right)$$

d.h. (X, Y) ist ein normalverteilter Zufallsvektor mit Parametern $\mu = \left(\begin{smallmatrix} \mu_x \\ \mu_y \end{smallmatrix}\right)$, $\Sigma = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$.
Wende nun die lineare Abbildung $(x, y) \mapsto x + y$ auf (X, Y) an und erhalte aus obigem Satz (d.h. $A = (\,1\ 1\,) \in \mathbb{R}^{1\times 2}$):

$$X + Y \sim \mathcal{N}\left(\mu_x + \mu_y, (\,1\ 1\,) \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix} \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) \right) = \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

□

Lemma III.26
Sind $X_1, \ldots, X_n$ gemeinsam normalverteilt, d.h. $(X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \Sigma)$ und sind alle X_i paarweise unkorreliert, d.h. $\Sigma_{ij} = 0$ für $i \neq j$, so sind $X_1, \ldots, X_d$ sogar unabhängig und es gilt $X_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$.

Beweis: Wegen $\Sigma = \operatorname{diag}(\Sigma_{11}, \ldots, \Sigma_{dd})$ mit $\Sigma_{ii} = \operatorname{Var}(X_i)$ folgt, dass $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \Sigma^{\frac{1}{2}} Y + \mu$ mit $Y \sim \mathcal{N}(0, E_d)$, also

$$X = \begin{pmatrix} \sqrt{\Sigma_{11}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\Sigma_{dd}} \end{pmatrix} Y + \mu = \begin{pmatrix} \sqrt{\Sigma_{11}} Y_1 + \mu_1 \\ \vdots \\ \sqrt{\Sigma_{dd}} Y_d + \mu_d \end{pmatrix}$$

Da $Y_1, \ldots, Y_d$ unabhängig sind und $X_i = \sqrt{\Sigma_{ii}} Y_i + \mu_i$ gilt, sind auch $X_1, \ldots, X_d$ unabhängig (s.o. Lemma). Außerdem impliziert $Y_i \sim \mathcal{N}(0,1)$, dass $X_i \sim \mathcal{N}(\mu_i, \Sigma_{ii})$ gilt. □

Bemerkung Im großen *Ausnahmefall* der gemeinsamen Normalverteilung sind also die Koordinaten genau dann unabhängig, wenn sie unkorreliert sind.

IV. Statistische Tests

IV.1. Hypothesentests

Beispiel IV.1 Wir wollen testen, ob die 1-Euro-Münze fair ist. Selbst wenn sie fair ist, ist bei 300 Würfen die Wahrscheinlichkeit, 150 mal Kopf und 150 mal Zahl zu bekommen $\mathrm{Bin}_{300,0.5}(150) \approx 0.065$. Man wird also die Annahme einer fairen Münze verwerfen, wenn die Anzahl Kopf außerhalb eines Intervalls $[150 - r, 150 + r]$ liegt.

Definition IV.2
Ein statistisches Modell *ist ein Tupel* $(\mathcal{X}, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ *bestehend aus einer Menge* $\mathcal{X}$ *mit einer* σ*-Algebra* $\mathcal{F}$ *(dem* Stichprobenraum*) und einer Familie* $(P_\vartheta)_{\vartheta\in\Theta}$ *von Wahrscheinlichkeitsmaßen auf* $\mathcal{F}$*. Die mindestens zweielementige Menge* Θ *heißt* Parametermenge *und jedes* $\vartheta \in \Theta$ Parameter*.*

Aufbau eines Testverfahrens:

1. Schritt Wahl eines statistischen Modells

Beispiel $\mathcal{X} = \{K, Z\}^{300}, \mathcal{F} = \mathcal{P}(\mathcal{X}), \Theta = [0,1], P_\vartheta(\{x\}) = \vartheta^{\#K}(1-\vartheta)^{\#Z}, x \in \mathcal{X}$.

2. Schritt Formulierung von Hypothese und Alternative: $\Theta = \Theta_0 \,\dot\cup\, \Theta_1$ disjunkte Zerlegung mit

$\vartheta \in \Theta_0 : \vartheta$ entspricht der Hypothese

$\vartheta \in \Theta_1 : \vartheta$ entspricht der Alternative

Beispiel $\Theta_0 = \{\frac{1}{2}\}, \Theta_1 = [0,1] \setminus \{\frac{1}{2}\}$

Man sagt, dass die (Null)-Hypothese $H_0 : \vartheta \in \Theta_0$ gegen die Alternative $H_1 : \vartheta \in \Theta_1$ getestet werden soll.

3. Schritt Wahl des *Irrtumsniveaus*: Wähle $\alpha \in (0,1)$ und fordern, dass die Wahrscheinlichkeit einer Entscheidung für die Alternative obwohl die Hypothese vorliegt (*Fehler 1. Art*) maximal α beträgt.

4. Schritt Wahl der *Entscheidungsregel*. Man wählt eine Zufallsvariable (messbare Funktion), *Teststatistik* genannt, $\varphi : \mathcal{X} \to [0,1]$ mit folgender Bedeutung: Tritt $x \in \mathcal{X}$ ein, so impliziert

$\varphi(x) = 0$: Entscheidung für H_0 („akzeptieren der Hypothese“)

$\varphi(x) = 1$: Entscheidung für H_1 („ablehnen der Hypothese“)

$\varphi(x) \in (0,1)$: Die Entscheidung wird mittels eines zusätzlichen unabhängigen Zufallsexperiments bestimmt, wobei mit Wahrscheinlichkeit $\varphi(x)$ die Wahl auf H_1 fällt.

Die letzte Möglichkeit (*randomisierter Test*) ist in der Anwendung oft nicht gewollt, aber erlaubt eine klarere mathematische Formulierung. Forderung: $\forall \vartheta \in \Theta_0 : \mathbb{E}_\vartheta[\varphi] \le \alpha$, dann heißt φ *Test zum Niveau* α (bei nicht randomisierten Test $P_\vartheta(\varphi = 1) \le \alpha$).

5. Schritt Durchführung des Experiments: Das passiert zum Schluss, weil sonst Täuschung und Selbsttäuschung fast unvermeidbar sind.

Definition IV.3
Für $\vartheta_1 \in \Theta_1$ und einen Test φ gibt $1 - \mathbb{E}_{\vartheta_1}[\varphi]$ die Wahrscheinlichkeit für den Fehler 2. Art, *der Entscheidung für H_0 obwohl $\vartheta_1 \in \Theta_1$ vorliegt, an. (φ nicht randomisiert. $1 - \mathbb{E}_{\vartheta_1}[\varphi] = 1 - P_{\vartheta_1}(\varphi = 1) = P_{\vartheta_1}(\varphi = 0)$)*

Definition IV.4
Ein Test φ von $H_0 : \vartheta \in \Theta_1$ heißt gleichmäßig bester Test zum Niveau α *(UMP = uniformly most powerful), falls*

(a) mit φ ist Test zum Niveau α gilt, dass $\sup\limits_{\vartheta_0 \in \Theta_0} \mathbb{E}_{\vartheta_0}[\varphi] \le \alpha$ und

(b) Für jeden anderen Test ψ zum Niveau α gilt $\forall \vartheta_1 \in \Theta_1 : \mathbb{E}_{\vartheta_1}[\varphi] \ge \mathbb{E}_{\vartheta_1}[\psi]$. („Fehlerwahrscheinlichkeit 2. Art ist minimal unter allen Niveau-α-Test und zwar gleichmäßig $\forall \vartheta_1 \in \Theta_1$“)

Ziel der Testtheorie ist es, beste Tests zu konstruieren.

IV.2. Neyman-Pearson-Lemma

Definition IV.5
Für Wahrscheinlichkeitsmaße P_0, P_1 auf einem diskreten Raum $(\mathcal{X}, \mathcal{P}(\mathcal{X}))$ mit Zähldichten p_0, p_1 heißt

$$R(x) = \begin{cases} \frac{p_1(x)}{p_0(x)} & \text{, falls } p_0(x) > 0 \\ +\infty & \text{, falls } p_0(x) = 0, p_1(x) > 0 \\ \text{beliebig} & \text{, falls } p_0(x) = p_1(x) = 0 \end{cases} \qquad x \in \mathcal{X}$$

Likelihood-Quotient *(oder* Dichtequotient *von p_1 bzgl. p_0.)*

Für Wahrscheinlichkeitsmaße p_0, p_1 mit $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ mit Dichten f_0, f_1 definieren wir

$$R(x) = \begin{cases} \frac{f_1(x)}{f_0(x)} & \text{, falls } f_0(x) > 0 \\ +\infty & \text{, falls } f_0(x) = 0, f_1(x) > 0 \\ \text{beliebig} & \text{, sonst} \end{cases} \quad \text{mit} \quad x \in \mathbb{R}^d$$

Jeder Test $\varphi : \mathcal{X} \to [0,1]$ der Form

$$\varphi(x) = \begin{cases} 1 & \text{, falls } R(x) > c \\ 0 & \text{, falls } R(x) < c \\ \gamma & \text{, falls } R(x) = c \end{cases}$$

mit beliebigen $c \ge 0, \gamma \in [0,1]$ heißt Neyman-Pearson-Test.

Bemerkung Mit dem Satz von Radon-Nykodym kann die Definition von $R(x)$ verallgemeinert werden.

Satz IV.6 (Neyman-Pearson-Lemma, 1932)
Für das Testen von $H_0 : \vartheta = 0$ gegen $H_1 : \vartheta = 1$ gilt

(a) *Ist φ^* ein Neyman-Pearson-Test, so gilt $\mathbb{E}_1[\varphi^*] \geq \mathbb{E}_1[\varphi]$ für jeden beliebigen Test φ mit $\mathbb{E}_0[\varphi] \leq \mathbb{E}_0[\varphi^*]$.*

(b) *Für jedes Niveau $\alpha \in (0,1)$ existiert ein Neyman-Pearson-Test φ^* mit exakt $\mathbb{E}_0[\varphi^*] = \alpha$.*

(c) *Ein gleichmäßig bester Test zum Niveau α ist gegeben durch einen Neyman-Pearson-Test φ^* mit $\mathbb{E}_0[\varphi^*] = \alpha$.*

Beweis: (nur für Zähldichten, Dichten analog) (a) Für $x \in A = \{x \in \mathcal{X} \mid \varphi^*(x) > \varphi(x)\}$ gilt $p_1(x) \geq cp_0(x)$ wegen $\varphi^*(x) > 0$. Für $x \in B = \{x \in \mathcal{X} \mid \varphi^*(x) < \varphi(x)\}$ gilt $p_1(x) \leq cp_0(x)$ wegen $\varphi^*(x) < 1$. Daher erhalten wir

$$\begin{aligned}\mathbb{E}_1[\varphi^*] - \mathbb{E}_1[\varphi] = \mathbb{E}_1[\varphi^* - \varphi] &= \sum_{x\in\mathcal{X}} (\varphi^*(x) - \varphi(x))p_1(x) \\ &\geq \sum_{x\in A} \underbrace{(\varphi^*(x) - \varphi(x))}_{>0} cp_0 + \sum_{x\in B} \underbrace{(\varphi^*(x) - \varphi(x))}_{<0} cp_0(x) + \sum_{x\in(A\cup B)^C} 0 \\ &= c\sum_{x\in\mathcal{X}} (\varphi^*(X) - \varphi(x))p_0(x) = c(\mathbb{E}_0[\varphi^*] - \mathbb{E}_0[\varphi]) \geq 0\end{aligned}$$

(b) Behauptung: Es gibt ein $c \geq 0$ mit $P_0(R \geq c) \geq \alpha$ und $P_0(R > c) \leq \alpha$. Setze $c := \inf\{k \geq 0 \mid P_0(R > k) \leq \alpha\}$. Wegen $P_0(R = \infty) = 0$ ist $c \in [0,\infty)$ wohldefiniert. Setze $\varrho(k) = P_0(R > k)$. Dann ist ϱ auf $[0,\infty)$ monoton fallend und rechtsstetig (beachte $\varrho(k) = 1 - P_0(R \leq k)$). Daher erfüllt c sowohl $\varrho(c) \leq \alpha$ als auch $\varrho(k) > \alpha$ für $k < c$. Außerdem erfüllt

$$P_0(R \geq c) \overset{\sigma\text{-Stetigkeit}}{=} 1 - \lim_{k\uparrow c} P_0\underbrace{(R \leq k)}_{<1-\alpha} \geq \alpha$$

Setze nun

$$\gamma = \begin{cases} 0 & \text{, falls } P_0(R = c) = 0 \\ \frac{\alpha - P_0(R>c)}{P_0(R=c)} & \text{, sonst} \end{cases} \qquad \text{sowie} \qquad \varphi^*(x) = \begin{cases} 1 & \text{, falls } R(x) > c \\ 0 & \text{, falls } R(x) < c \\ \gamma & \text{, falls } R(x) = c \end{cases}$$

Dann gilt $\gamma \subset [0,1]$ und $\mathbb{E}_0[\varphi^*] = 1P_0(R(x) > c) + \gamma P_0(R(x) = c) = \alpha$.

(c) folgt direkt aus (a) und (b). □

Beispiel IV.7 (Einfacher Gaußtest) Gegeben sei das statistisches Modell

$$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, (P_\vartheta)_{\vartheta\in[0,1]}) \quad \text{mit} \quad P_\vartheta \sim \mathcal{N}(\vartheta, 1)$$

Teste $H_0 : \vartheta = 0$ gegen $H_1 : \vartheta = 1$ zum Niveau $\alpha = 0.025$ Likelihood-Quotient:

$$R(x) = \frac{f_1(x)}{f_0(x)} = \exp\left(-\tfrac{1}{2}((x-1)^2 - x^2)\right) = \exp\left(x - \tfrac{1}{2}\right)$$

Wähle $c \geq 0$ so dass

$$\alpha \overset{!}{=} P_0(R(x) \geq c) = P_0\left(e^{x-\frac{1}{2}} \geq c\right) = P_0\left(x \geq \log c + \tfrac{1}{2}\right)$$

$$\implies \log c + \frac{1}{2} \approx 1.96 \implies c \approx 4.3$$

$$\implies \varphi^x(x) = \begin{cases} 1 & \text{, falls } e^{x-\frac{1}{2}} \geq 4.3 \\ 0 & \text{sonst} \end{cases}$$

ist gleichmäßig bester Test.

V. Grenzwertsätze

Wirft man einen Würfel unendlich oft, und zählt die „6“-Würfe, so wird bei fairem Würfel mit zunehmender Zahl der Versuche die Proportion der „6“-Würfe zur Anzahl der Versuche, d.h. die relative Häufigkeit der „6“ gegen deren Wahrscheinlichkeit $\frac{1}{6}$ konvergieren.

In welchem Sinne ist diese heuristische Feststellung wichtig?

V.1. Gesetze der großen Zahlen

Satz V.1 (Markov-Ungleichung)
Sei Y eine reelle Zufallsvariable und $f : [0,\infty) \to [0,\infty)$ monoton wachsend. Dann gilt für jedes $\varepsilon > 0$:

$$P(|Y| \geq \varepsilon) \leq \frac{\mathbb{E}[f(|Y|)]}{f(\varepsilon)} \quad \textit{falls } f(\varepsilon) > 0$$

Beweis: $|Y|$ ist Zufallsvariable, also auch $f(|Y|)$ (da f monoton wachsend $\implies$ f messbar, denn $f^{-1}((c,\infty))$ ist Intervall). Sei $\varepsilon > 0$ mit $f(\varepsilon) > 0$:

$$\mathbf{1}_{\{|Y|\geq\varepsilon\}} \leq \frac{f(|Y|)}{f(\varepsilon)} \implies \underbrace{\mathbb{E}[\mathbf{1}_{\{|Y|\geq\varepsilon\}}]}_{P(|Y|\geq\varepsilon)} \leq \mathbb{E}[\frac{f(|Y|)}{f(\varepsilon)}] \qquad \square$$

Beispiel V.2 Hat die Zufallsvariable X ein endliches p-tes Moment (d.h. $\mathbb{E}[|X|^p] < \infty$) mit $p > 0$, so gilt $P(|X| \geq k) \leq \frac{\mathbb{E}[|X|^p]}{k^p}$, d.h. für große k ist $P(|X| > k)$ maximal von der Ordnung k^{-p}.

Korollar V.3 (Tschebyscheff[4])
Sei $Y \in \mathcal{L}^2(P)$, $\varepsilon > 0$. Dann gilt $P(|Y - \mathbb{E}[Y]| \geq \varepsilon) \leq \dfrac{\mathrm{Var}(Y)}{\varepsilon^2}$.

Beweis: $X := Y - \mathbb{E}[Y]$, $f(x) := x^2$ für alle $x \geq 0$. Dann gilt

$$P(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[|X|^2]}{\varepsilon^2} = \frac{\mathrm{Var}(Y)}{\varepsilon^2} \qquad \square$$

[4] Weitere Umschriften für П. Л. Чебышёв sind Čebyšev oder Tschebyschow.

Satz V.4 (schwaches Gesetz der großen Zahlen)
Es seien $(X_i)_{i\geq 1}$ *unkorrelierte Zufallsvariablen in* $\mathcal{L}^2(P)$ *mit demselben Erwartungswert* $\mu \in \mathbb{R}$ *so dass* $\sup_{i\geq 1} \operatorname{Var}(X_i) < \infty$ *mit* $\sup_{i\geq 1} \operatorname{Var}(X_i) = c$*. Dann erfüllt das arithmetische Mittel* $A_n := \frac{1}{n}\sum_{i=1}^n X_i$ *für jedes* $\varepsilon > 0$ *gerade*

$$\lim_{n\to\infty} P(|A_n - \mu| > \varepsilon) = 0$$

Beweis:

$$\mathbb{E}[A_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = \mu$$

$$\begin{aligned}\operatorname{Var}(A_n) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^n X_i\right)\\ &= \frac{1}{n^2}\left(\sum_{i=1}^n \operatorname{Var}(X_i) + \sum_{\substack{i,j=1\\ i\neq j}}^n \operatorname{Cov}(X_i, X_j)\right)\\ &= \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}(X_i) \leq \frac{1}{n^2}\cdot n\cdot c = \frac{c}{n}\end{aligned}$$

$$\implies P(|A_n - \mathbb{E}[A_n]| > \varepsilon) \leq \frac{\operatorname{Var}(A_n)}{\varepsilon^2} \leq \frac{c}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$$

□

Beispiel V.5 Bernoulli-Experiment mit $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$. Dann gilt

$$P(|A_n - p| > \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$$

Man sagt, „der Mittelwert A_n stabilisiert sich um den Erwartungswert p mit wachsender Zahl von Wiederholungen des Versuchs“. Man beachte hier, dass die Abweichungen für p nahe 0 oder 1 viel kleiner sind als für p nahe $\frac{1}{2}$.

Bemerkung Das schwache Gesetz gilt auch für unabhängige, identisch verteilte Zufallsvariablen in $\mathcal{L}^1$. (Siehe [3])

Korollar V.6 (Weierstraßscher Approximationssatz)
Für $f : [0,1] \to \mathbb{R}$ *stetig definiere das zugehörige* Bernstein-Polynom $f_n : [0,1] \to \mathbb{R}$

$$f_n(p) := \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} p^k (1-p)^{n-k} \qquad p \in [0,1)$$

Dann gilt $\|f - f_n\|_\infty \xrightarrow{n\to\infty} 0$

Beweis: Seien $X_1, \ldots, X_n$ unabhängig, Bernoulli-verteilte (zum Parameter p) Zufallsvariablen.

$$\begin{aligned}\implies \mathbb{E}_p\left[f\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] &= \sum_{k=0}^n f\left(\frac{k}{n}\right) B_{n,p}(\{k\})\\ &= \sum_{k=0}^n f\left(\frac{1}{n}\right)\binom{n}{k} p^k (1-p)^{n-k} = f_n(p)\end{aligned}$$

f ist gleichmäßig stetig auf $[0,1]$. Sei $\varepsilon > 0$ beliebig, dann existiert ein $\delta > 0$, so dass $\forall x, y \in [0,1]$ mit $|x-y| \leq \delta$ gilt: $|f(x) - f(y)| \leq \varepsilon$.

$$\begin{aligned}
|f_n(p) - f(p)| &= \left| \mathbb{E}\left[f\left(\frac{1}{n} \sum_{i=1}^n X_i \right) \right] - f(p) \right| \\
&\leq \mathbb{E}\left[\left| f\left(\frac{1}{n} \sum_{i=1}^n X_i \right) - f(p) \right| \mathbf{1}_{\{|\frac{1}{n}\sum_{i=1}^n X_i - p| \leq \delta\}} \right] \\
&\quad + \mathbb{E}\left[\left| f\left(\frac{1}{n} \sum_{i=1}^n X_i \right) - f(p) \right| \mathbf{1}_{\{|\frac{1}{n}\sum_{i=1}^n X_i - p| > \delta\}} \right] \\
&\leq \varepsilon + 2\|f\|_\infty \underbrace{\mathbb{E}\left[\mathbf{1}_{\{|\frac{1}{n}\sum_{i=1}^n X_i - p| > \delta\}} \right]}_{P(|\frac{1}{n}\sum_{i=1}^n X_i - p| > \delta)} \\
&\leq \varepsilon + 2\|f\|_\infty \frac{p(1-p)}{n\delta^2} \leq \varepsilon + 2\|f\|_\infty \frac{1}{n\delta^2} \\
&\implies \limsup_{n\to\infty} \|f_n - f\|_\infty \leq \varepsilon \qquad \forall \varepsilon > 0 \\
&\implies \lim_{n\to\infty} \|f_n - f\|_\infty = 0
\end{aligned}$$

□

Definition V.7
Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum, Y und $(Y_n)_{n\in\mathbb{N}}$ seien Zufallsvariablen mit Werten in $\mathbb{R}$. $(Y_n)_{n\in\mathbb{N}}$ konvergiert stochastisch gegen Y, *wenn*

$$\forall \varepsilon > 0: \qquad P(|Y_n - Y| \leq \varepsilon) \xrightarrow{n\to\infty} 1 \quad \text{oder auch} \quad \forall \varepsilon > 0 \qquad P(|Y_n - Y| > \varepsilon) \xrightarrow{n\to\infty} 0$$

Man schreibt $Y_n \xrightarrow[n\to\infty]{P} Y$. Man sagt auch „$Y_n$ konvergiert nach Maß“ oder „in Wahrscheinlichkeit“.

Definition V.8
Seien $(\Omega, \mathcal{F}, P)$, Y, Y_n wie oben. $(Y_n)_{n\in\mathbb{N}}$ konvergiert P-fast sicher, *falls*

$$P(\{\omega \in \Omega \mid Y_n(\omega) \xrightarrow{n\to\infty} Y(\omega)\}) = 1 \qquad \text{Man schreibt} \quad Y_n \xrightarrow[P\text{-f.s}]{n\to\infty} Y.$$

Bemerkung
- Im schwachen Gesetz gilt also $A_n \xrightarrow{P} \mu$
- $\{Y_n \to Y\} = \bigcap_{n\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} \bigcap_{l\geq k} \{|Y_l - Y| \leq \frac{1}{n}\} \in \mathcal{F}$ $(= \bigcap_{\varepsilon>0} \bigcup_{k\in\mathbb{N}} \bigcap_{l\geq k} \{|Y_l - Y| \leq \varepsilon\})$.
- Fast sichere Konvergenz impliziert stochastische Konvergenz (siehe Übung).
- Die Umkehrung gilt nicht: Sei $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}_\mathbb{R}$, P gleichmäßige Verteilung auf $[0,1)$,

  $$X_{j,k} = \mathbf{1}_{[k2^{-j},(k+1)2^{-j})}, \qquad j \in \mathbb{N}, \quad k = 0, \ldots, 2^j - 1$$

  Setze $Y_n := X_{j,k}$ für $n = 2^j + k$, $k = 0, \ldots, 2^j - 1$. Dann gilt für $\varepsilon > 0$, dass

  $$P(|Y_n| > \varepsilon) = P(Y_n = 1) = 2^{-j} \xrightarrow{n\to\infty} 0$$

  Andererseits $\forall \omega \in [0,1) \exists$ unendlich viele n mit $Y_n(\omega) = 1 \implies P(\{\omega \mid Y_n(\omega) \to Y(\omega)\}) = 0$

Satz V.9 (Starkes Gesetz der großen Zahlen)
Es seien $(X_i)_{i\geq 1}$ unkorrelierte Zufallsvariablen in $\mathcal{L}^2$ mit dem selben Erwartungswert μ, so dass $\sup_{i\geq 1} \mathrm{Var}(X_i) < \infty$. *Dann gilt für $A_n = \frac{1}{n}\sum_{i=1}^n X_i$:*

$$A_n \xrightarrow{n\to\infty} \mu \quad P\text{-f.s.}$$

Bemerkung Die von Etemadi (1981) angegebene Version des starken Gesetzes der großen Zahlen (GdgZ) benötigt nur die Voraussetzung $X_i \in \mathcal{L}^1$, fordert dafür aber, dass die $(X_i)_{i\geq 1}$ paarweise unabhängig und identisch verteilt (i.i.d.) sind.

Beweis: Setze $v := \sup_i \mathrm{Var}(X_i)$

(a) Zeige $A_{n^2} \to \mu$ P-f.s.: Tschebyscheff-Ungleichung V.3 liefert für $\varepsilon > 0$

$$P(|A_{n^2} - \mu| > \varepsilon) \leq \frac{\mathrm{Var}(A_{n^2})}{\varepsilon^2} = \frac{v}{n^2\varepsilon^2}$$

Damit sind diese Wahrscheinlichkeiten summierbar:

$$\sum_{n\geq 1} P(|A_{n^2} - \mu| > \varepsilon) \leq \frac{v}{\varepsilon^2}\sum_{n\geq 1}\frac{1}{n^2} < \infty$$

1. Teil von Borel-Cantelli II.8 gibt daher $P(|A_{n^2} - \mu| > \varepsilon$ für unendlich viele $n) = 0$.

$$\begin{aligned}&\Longrightarrow P\Bigg(\bigcup_{\substack{\varepsilon\in\mathbb{Q}\\ \varepsilon>0}}\{|A_{n^2}-\mu| > \varepsilon \text{ für } \infty\text{-viele } n\}\Bigg) = 0\\ &\Longrightarrow P(\{A_{n^2}\to\mu\}^C) = 0 \iff A_{n^2}\to\mu \quad P\text{-f.s.}\end{aligned}$$

(b) Betrachte Abweichungen von A_m zu $A_{n^2(m)}$ mit $n(m) \in \mathbb{N}$ so, dass $n^2(m) \leq m < (n(m)+1)^2$. Tschebyscheff:

$$P\Bigg(\Bigg|\sum_{i=1}^m X_i - \sum_{i=1}^{n(m)^2} X_i\Bigg| \geq \varepsilon n(m)^2\Bigg) \leq \frac{\mathrm{Var}\Big(\sum_{i=n(m)^2+1}^m X_i\Big)}{\varepsilon^2 n(m)^4} \leq \frac{(m-n(m)^2)\cdot v}{\varepsilon^2 n(m)^4}$$

$$\begin{aligned}\Longrightarrow \sum_{m\geq 1} P\Bigg(\Bigg|\sum_{i=1}^m X_i - \sum_{i=1}^{n(m)^2} X_i\Bigg| \geq \varepsilon n(m)^2\Bigg) &\leq \frac{v}{\varepsilon^2}\sum_{m\geq 1}\frac{m-n(m)^2}{n(m)^4}\\ = \frac{v}{\varepsilon^2}\sum_{n=1}^\infty \sum_{m=n^2}^{(n+1)^2-1}\frac{m-n^2}{n^4} &\leq \frac{v}{\varepsilon^2}\sum_{n=1}^\infty \frac{2n\cdot(2n+1)}{n^4} < \infty\end{aligned}$$

Borel-Cantelli liefert $P(|\sum_{i=1}^m X_i - \sum_{i=1}^{n(m)^2} X_i| > \varepsilon n(m)^2$ für ∞-viele $m) = 0$. Vereinigung über $\varepsilon \in \mathbb{Q}^+$ liefert wie oben

$$\frac{1}{n(m)^2}\Bigg|\sum_{i=1}^n X_i - \sum_{i=1} n(m)^2 X_i\Bigg| \to 0 \quad P\text{-f.s.}$$

Wegen $\frac{1}{n(m)^2}\sum_{i=1}^{n(m)^2} X_i \xrightarrow{P\text{-f.s.}} \mu$ aus (a) folgt $|\frac{1}{n(m)^2}\sum_{i=1}^{m} X_i - \mu| \xrightarrow{P\text{-f.s.}} 0$ und daher auch

$$\underbrace{\frac{n(m)^2}{m}}_{\to 1}\cdot\left|\frac{1}{n(m)^2}\sum_{i=1}^{m} X_i - \mu\right| = \left|\frac{1}{m}\sum_{i=1}^{m} X_i - \underbrace{\frac{n(m)^2}{m}}_{\to 1}\mu\right| \xrightarrow{P\text{-f.s.}} 0$$

und daher $A_m \to \mu$ P-f.s. □

Definition V.10

Identifiziert man $X, Y \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$, falls $X = Y$ P-f.s. (d.h. $P(\{\omega : X(\omega) = Y(\omega)\}) = 1$) gilt, so erhält man die Menge $L^p(\Omega, \mathcal{F}, P)$ (der Äquivalenzklassen). In der Analysis zeigt man, dass $L^p(\Omega, \mathcal{F}, P)$ mit der Norm $\|X\|_{L^p} = \mathbb{E}[|X|^p]^{\frac{1}{p}} = (\int |X(\omega)|^p P(d\omega))^{\frac{1}{p}}$ für $p \geq 1$ einen Banachraum (d.h. vollständigen normierten Raum) bilden. $L^2(\Omega, \mathcal{F}, P)$ mit dem Skalarprodukt $\langle X, Y\rangle_{L^2} = \mathbb{E}[X \cdot Y]$ ist sogar ein Hilbertraum.

Für eine Folge (X_n) von Zufallsvariablen in $\mathcal{L}^p$, $p \geq 1$ und ein $X \in \mathcal{L}^p$ sagen wir, dass X_n gegen X in $\mathcal{L}^p$ konvergiert, *falls $\|X_n - X\|_{L^p} = \mathbb{E}[|X_n - X|^p]^{\frac{1}{p}} \xrightarrow{n\to\infty} 0$ gilt.*

Lemma V.11

$\mathcal{L}^p$-Konvergenz impliziert stochastische Konvergenz.

Beweis: Die Markov-Ungleichung V.1 liefert

$$\forall \varepsilon > 0 \quad P(|X_n - X| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p} = \varepsilon^{-p}\|X_n - X\|_{L^p}^p \xrightarrow[\text{Vorauss.}]{n\to\infty} 0$$

$$\Longrightarrow X_n \xrightarrow{P} X \text{ stochastisch}$$ □

Beispiel V.12 Betrachte wiederum $S_n := \sum_{k=1}^n \varepsilon_k \cdot \frac{1}{k}$ mit $P(\varepsilon_k = 1) = P(\varepsilon_k = -1) = \frac{1}{2}$ und (ε_k) unabhängig. Wir berechnen

$$\mathbb{E}[S_n] = \sum_{k=1}^n \mathbb{E}[\varepsilon_k]\cdot\tfrac{1}{k} = 0$$

$$\mathrm{Var}(S_m - S_n) \overset{m>n}{=} \mathrm{Var}\left(\sum_{k=n+1}^m \varepsilon_k \cdot \tfrac{1}{k}\right) \overset{\text{Unabh.}}{=} \sum_{k=n+1}^m \frac{1}{k^2}\cdot \mathrm{Var}(\varepsilon_k) = \sum_{k=n+1}^m \frac{1}{k^2}$$

$$\Longrightarrow \|S_m - S_n\|_{L^2}^2 = \mathrm{Var}(S_m - S_n) = \sum_{k=n+1}^m \frac{1}{k^2} < \sum_{k=n+1}^\infty \frac{1}{k^2} \xrightarrow{n\to\infty} 0 \qquad (m > n)$$

$$\Longrightarrow (S_n) \quad \text{ist Cauchy-Folge.}$$

$$\overset{L^2\text{ vollst-}}{\Longrightarrow} \exists S_\infty \in L^2 : S_n \xrightarrow[n\to\infty]{L^2} S_\infty \Longrightarrow S_n \xrightarrow{P} S_\infty \quad \text{stochastisch}$$

Offene Frage: Gilt auch $S_n \to S_\infty$ P-f.s.?

Satz V.13 (Lévys Äquivalenzsatz)

Es sei $(X_k)_{k\geq 1}$ eine Folge unabhängiger Zufallsvariablen und $S_n = \sum_{k=1}^n X_k$. Dann sind äquivalent:

(i) S_n konvergiert P-f.s.

(ii) S_n konvergiert P-stochastisch.

Bemerkung (a) Konvergenz ist im Sinne des Cauchy-Kriteriums gemeint, d.h. in (i) ist $P((S_n \text{ ist Cauchyfolge})) = 1$ und in (ii) gilt

$$\forall \varepsilon > 0 \lim_{N\to\infty} \sup_{m\geq N} P(|S_m - S_N| > \varepsilon) = 0$$

(b) Falls (i) bzw. (ii) nicht gelten, so divergiert (S_n) P-f.s. wegen des 0-1-Gesetzes von Kolmogorov II.23.

Beweis: (i)⇒(ii) folgt wie die Implikation f.s. Konvergenz $\Longrightarrow$ stoch. Konvergenz

(ii)⇒(i) benötigt eine sogenannte *Maximal-Ungleichung*:

Lemma V.14 (Ottaviani-Ungleichung)
Für $\alpha > 0$ gilt $P(\max\limits_{1\leq j\leq n} |S_j| \geq 2\alpha) \leq \dfrac{P(|S_n| \geq \alpha)}{1 - \max_{1\leq j\leq n} P(|S_n - S_j| \geq \alpha)}$ sofern der Nenner nicht 0 ist.

Beweis: Setze $\tau(w) := \inf\big(\{1 \leq j \leq n \mid |S_j(w)| \geq 2\alpha\} \cup \{n+1\}\big)$ („Stoppzeit"). Dann gilt

$$\begin{aligned} P(|S_n| \geq \alpha) &\geq P(|S_n| \geq \alpha, \max_{1\leq j\leq n} |S_j| \geq 2\alpha) = \sum_{j=1}^{n} P(|S_n| \geq \alpha,\ \tau = j) \\ &\geq \sum_{j=1}^{n} P(|S_n - S_j| \leq \alpha, \tau = j) \end{aligned}$$

Da $S_n - S_j = \sum_{k=j+1}^{n} X_k$ gilt und $\{\tau = j\}$ nur von $S_1, \dots, S_j$, d.h. $X_1, \dots, X_j$ abhängt, gilt wegen Unabhängigkeit weiter

$$\begin{aligned} (\dots) &= \sum_{j=1}^{n} P(|S_n - S_j| \leq \alpha) \cdot P(\tau = j) \geq \min_{1\leq j\leq n} P(|S_n - S_j| \leq \alpha) \cdot \sum_{j=1}^{n} P(\tau = j) \\ &= (1 - \max_{1\leq j\leq n} P(|S_n - S_j| > \alpha)) \cdot P(\max_{1\leq j\leq n} |S_j| \geq 2\alpha) \end{aligned}$$ □

Wende Ottaviani auf die Summen $S_n - S_k = \sum_{i=k+1}^{n} X_i$ mit $\alpha = \frac{\varepsilon}{2}$ an und verwende

$$\begin{aligned} P(|S_n - S_j| \geq \alpha) &\leq P\left(|S_n - S_k| \geq \frac{\alpha}{2} \quad \text{oder} \quad |S_j - S_k| \geq \frac{\alpha}{2}\right) \\ &\leq P\left(|S_n - S_k| \geq \frac{\varepsilon}{4}\right) + P\left(|S_j - S_k| \geq \frac{\varepsilon}{4}\right) \leq \frac{\varepsilon}{2} \leq \frac{1}{2} \\ &\Longrightarrow \forall n \geq k : P(\max_{k\leq j\leq n} |S_j - S_k| \geq \varepsilon) \leq \frac{\varepsilon}{2} \end{aligned}$$

σ-Stetigkeit des Wahrscheinlichkeitsmaßes liefert für $n \to \infty$

$$P(\max_{j\geq k} |S_j - S_k| \geq \varepsilon) \leq \frac{\varepsilon}{2}$$

Zu zeigen: $P((S_n)$ Cauchyfolge$) = 1$. Die linke Seite ist gleich

$$P\Bigg(\bigcap_{\substack{\varepsilon>0\\ \varepsilon\in\mathbb{Q}}}\Big\{\exists k \geq 1 \forall m,n \geq k : |S_m - S_n| < \varepsilon\Big\}\Bigg)$$

Zeige also, dass für alle $\varepsilon > 0$ $P(\forall k \geq 1 \exists m,n \geq k : |S_m - S_n| \geq \varepsilon) = 0$. Dies folgt aus

$$\forall \varepsilon > 0 : \quad P(\forall k \geq 1 \sup_{j\geq k} |S_j - S_k| \geq \tfrac{\varepsilon}{2}) = 0$$

Dies folgt wiederum aus $\forall \varepsilon > 0 : \lim_{k\to\infty} P(\sup_{j\geq k} |S_j - S_k| \geq \frac{\varepsilon}{2}) = 0$ (σ-Stetigkeit). Oben haben wir nun gesehen, dass für $k = k(\varepsilon)$ gilt

$$P(\sup_{j\geq k(\varepsilon)} |S_j - S_{k(\varepsilon)}| \geq \varepsilon) \leq \frac{\varepsilon}{2}$$

Halte nun das erste ε fest und lasse $k \to \infty$ gehen, so dass $\lim_{k\to\infty} P(\sup_{j\geq k} |S_j - S_k| \geq \varepsilon) = 0$. □

Anwendung (a) $S_n = \sum_{k=1}^{n} \varepsilon_k \cdot \frac{1}{k}, \quad P(\varepsilon_k = \pm 1) = \frac{1}{2}, \quad (\varepsilon_k)$ unabhängig

$$\overset{\text{s.o.}}{\Longrightarrow} (S_n)_{n\geq 1} \quad \text{konvergiert in } L^2\text{, also stochastisch}$$

$$\overset{\text{Äq-Satz}}{\Longrightarrow} (S_n)_{n\geq 1} \quad \text{konvergiert fast sicher.}$$

(b) Man kann mit dem Äquivalenzsatz auch unsere Variante des starken GdgZ aus dem schwachen GdgZ folgern. Dafür braucht man aus der Analysis ein Reihenkriterium:

$$\frac{1}{n}\sum_{k=1}^{n} \alpha_k \to 0 \iff \sum_{k=1}^{n} \frac{\alpha_k}{k} \to 0 \quad \text{Details siehe [1]}$$

Aus [4] ohne Beweis folgendes einfaches Kriterium für Konvergenz einer Reihe unabhängiger Zufallsvariablen:

Satz V.15 (Drei-Reihen-Satz)
Für unabhängige Zufallsvariablen $(X_k)_{k\geq 1}$ konvergiert $S_n = \sum_{k=1}^{n} X_k$ f.s. (bzw. stochastisch) genau dann, wenn folgende drei Reihen konvergieren:

(i) $\sum_{k=1} P(|X_k| > 1)$

(ii) $\sum_{k=1}^{\infty} \mathbb{E}[X_k \cdot \mathbf{1}_{[-1,1]}(X_k)]$

(iii) $\sum_{k=1}^{\infty} \mathrm{Var}(X_k \cdot \mathbf{1}_{[-1,1]}(X_k))$

Anwendung Beispiel von oben (a), aber mit $P(\varepsilon_k = 1) = p$, $P(\varepsilon_k - 1) = 1 - p$. Dann konvergiert (S_n) genau dann, wenn $p = \frac{1}{2}$ gilt (sonst divergiert die Reihe in *(ii)*).

V.2. Konvergenz in Verteilung

Im schwachen Gesetz der großen Zahlen wurde $A_n \xrightarrow{n\to\infty} \mu$ stochastisch gezeigt. Wie schnell ist diese Konvergenz eigentlich?

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right| > n^{-\alpha}\right) \overset{\text{Tschebyscheff}}{\leq} \frac{\frac{1}{n}\cdot v}{n^{-2\alpha}} \xrightarrow{n\to\infty} 0 \quad \text{für alle } \alpha < \tfrac{1}{2}$$

Für $\alpha > \frac{1}{2}$ gilt $\operatorname{Var}(n^\alpha \cdot A_n) \cong \frac{n^{2\alpha}}{n} \xrightarrow{n\to\infty} \infty$ und es ist keine Konvergenz von (A_n) zu erwarten. Kritischer Wert ist $\alpha = \frac{1}{2}$, wo im Fall $X_i \in \mathcal{L}^2$ identisch verteilt und unkorreliert gilt

$$S_n^* := n^{\frac{1}{2}} \cdot \left(\frac{A_n - \mu}{\sigma(X_1)}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma(X_i)}$$

besitzt die Varianz $\operatorname{Var}(S_n^*) = \frac{1}{n}\sum_{i=1}^n \frac{\operatorname{Var}(X_i)}{\sigma^2(X_i)} = 1$ sowie $\mathbb{E}[S_n^*] = 0$. Man nennt S_n^* *standardisierte Summe* der (X_i). Wir erwarten nicht, dass $S_n^* \to 0$ (stochastisch) konvergiert, weil $\operatorname{Var}(S_n^*) = 1 > 0$ gilt, andererseits könnte die Verteilung von $(S_n^*)_{n\geq 1}$ sich stabilisieren und in geeignetem Sinne konvergieren.

Beispiel V.16 (a) $X_i \sim \mathcal{N}(\mu,\sigma^2) \implies S_n^* \sim \mathcal{N}(0,1)$, d.h. $P^{S_n^*} = \mathcal{N}(0,1)$ für alle $n \geq 1$.

(b) $X_i \sim \mathrm{U}[-1,1] \overset{\text{Übung}}{\implies} f^{S_n^*}$ nähert sich der Gaußschen Glockenkurve an.

(c) „Stabdiagramm“ der Binomialverteilung nähert sich auch der Gaußschen Glockenkurve an (Schule oder später in Vorlesung).

Definition V.17
Es seien $(X_n)_{n\geq 1}$ und X $\mathbb{R}^d$-wertige Zufallsvariablen (nicht notwendigerweise auf dem selben Wahrscheinlichkeitsraum). Man sagt, dass (X_n) gegen X in Verteilung konvergiert*, Notation $X_n \xrightarrow{d} X$, falls für alle stetigen und beschränkten Funktionen $\varphi : \mathbb{R}^d \to \mathbb{R}$ gilt:*

$$\mathbb{E}[\varphi(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[\varphi(X)]$$

Allgemein definiert man für Wahrscheinlichkeitsmaße (μ_n) und μ auf $\mathcal{B}_{\mathbb{R}^d}$ die schwache Konvergenz *$\mu_n \xrightarrow{w} \mu$, falls*

$$\int_{\mathbb{R}^d}\varphi(x)\mu_n(dx) \to \int_{\mathbb{R}^d}\varphi(x)\mu(dx)$$

gilt für alle stetigen und beschränkten $\varphi : \mathbb{R}^d \to \mathbb{R}$.

Bemerkung (a) Konvergenz in Verteilung $X_n \xrightarrow{d} X$ bedeutet also gerade $P^{X_n} \xrightarrow{w} P^X$, weil

$$\mathbb{E}[\varphi(X_n)] = \int \varphi(x)P^{X_n}(dx) \to \int \varphi(x)P^X(dx) = \mathbb{E}[\varphi(X)]$$

Man kann daher z.B. auch schreiben $X_n \xrightarrow{d} \mathcal{N}(0,1)$ ohne eine Zufallsvariable $X \sim \mathcal{N}(0,1)$ hinzuschreiben.

(b) Schwache Konvergenz $P_n \overset{w}{\to} P$ bedeutet *nicht* $P_n(A) \to P(A)$ für alle Ereignisse A. Einfaches Gegenbeispiel ist für eine Folge reeller Zahlen $(x_n)_{n\geq 1}$ mit $x_n \to x \in \mathbb{R}$ die Folge $(\delta_{x_n})_{n\geq 1}$ der Punktmaße in (x_n) zu betrachten, dann gilt

$$\int \varphi(\xi)\delta_{x_n}(d\xi) = \varphi(x_n) \xrightarrow{\varphi \text{ stetig}} \varphi(x) = \int \varphi(\xi)\delta_x(d\xi)$$

d.h. $\delta_{x_n} \overset{w}{\to} \delta_x$, aber i.A. gilt $\delta_{x_n}(\{x\}) = 0 \neq 1 = \delta_x(\{x\})$, d.h. $\delta_{x_n}(\{x\}) \not\to \delta_x(\{x\})$.

Analog sieht man z.B. für eine Zufallsvariable X und $a_n \to a, b_n \to b$ (reeller Zahlen), dass $a_n X + b_n \overset{d}{\to} aX + b$.

Satz V.18

Konvergiert X_n gegen X stochastisch, so auch in Verteilung $X_n \overset{P}{\to} X \implies X_n \overset{d}{\to} X$.

Beweis: Nach Übungsaufgabe folgt aus $X_n \overset{P}{\to} X$, dass für eine Teilfolge gilt $X_{n_l} \to X$ P-f.s.

$$\implies \mathbb{E}[\varphi(X_{n_k})] = \int \underbrace{\varphi(\overbrace{X_{n_k}(\omega)}^{\overset{\text{f.s.}}{\to} X(\omega)})}_{\substack{\varphi \text{ stetig} \overset{\text{f.s.}}{\to} \varphi(X(\omega)) \\ \text{und } \leq \|\varphi\|_\infty < \infty \in \mathcal{L}^1}} P(d\omega) \xrightarrow[\text{Konvergenz}]{\text{dominierte}} \int \varphi(X(\omega))P(d\omega) = \mathbb{E}[\varphi(X)]$$

$$\implies X_{n_k} \xrightarrow[k\to\infty]{d} X$$

Daraus folgt auch allgemein $X_n \overset{d}{\to} X$, wegen *Widerspruchsarguments*:

Sonst gäbe es eine Teilfolge (X_{n_l}), ein φ und ein $\varepsilon > 0$ mit

$$\forall l \in \mathbb{N} : |\mathbb{E}[\varphi(X_{n_l}) - \mathbb{E}[\varphi(X)]| > \varepsilon \qquad \square$$

Dies wäre dann auch für jede Teilteilfolge $(X_{n_{l_k}})$ wahr. Nach obigem Argument gibt es jedoch eine f.s.-konvergierende Teilfolge $(X_{n_{l_k}})$ mit $\mathbb{E}[\varphi(X_{n_{l_k}})] \to \mathbb{E}[\varphi(X)]$. ↯

Beispiel V.19 (Die Umkehrung gilt nicht) Es sei X $\mathcal{N}(0,1)$-verteilt sowie $X_n = X$, $n \geq 1$ und $Y = -X$. Dann gilt $P^{X_n} = P^X = \mathcal{N}(0,1)$, $P^Y = \mathcal{N}(0,1)$. Offenbar gilt daher $P^{X_n} \overset{w}{\to} P^Y$, d.h. $X_n \overset{d}{\to} Y$. Andererseits ist $P(|X_n - Y| > \varepsilon) = P(|2X| > \varepsilon) \equiv c > 0$, so dass $X_n \overset{P}{\not\to} Y$. Beachte auch, dass für Konvergenz in Verteilung die Zufallsvariablen nicht auf demselben Wahrscheinlichkeitsraum definiert sein muss.

Satz V.20

Für reellwertige Zufallsvariablen $(X_n)_{n\geq 1}$ und X sind äquivalent:

(a) $X_n \xrightarrow[n\to\infty]{d} X$

(b) $F^{X_n}(x) \xrightarrow{n\to\infty} F^X(x)$ *für alle $x \in \mathbb{R}$, wo F^X stetig ist („Stetigkeitspunkte von F^X“).*

Beispiel V.21 $x_n \to x$ reelle Zahlen,

$$X_n \equiv x_n, X \equiv x \implies F^{X_n}(\xi) = P(X_n \le \xi) = \begin{cases} 1, & \xi \ge x_n \\ 0, & \xi < x_n \end{cases} = \mathbf{1}_{[x_n,\infty)}(\xi)$$

Analog ist $F^X(\xi) = \mathbf{1}_{[x,\infty)}(\xi)$.

$$\forall \xi < x : F^{X_n}(\xi) \overset{w}{\to} 0 = F^X(\xi) \qquad \forall \xi > x F^{X_n}(\xi) \to 1 = F^X(\xi)$$

$F^{X_n}(x)$ kann 0 oder 1 sein, d.h. „$F^{X_n}(\xi) \to F^X(\xi)$“ gilt i.A. nicht für $\xi = x$.

Beweis:

(a)⇒(b): Wähle φ stetig und beschränkt mit $\mathbf{1}_{(-\infty,x]} \le \varphi \le 1_{(-\infty,x+\delta]}$ für $x \in \mathbb{R}$, $\delta > 0$. Dann gilt

$$F^{X_n}(x) = \mathbb{E}[\mathbf{1}_{(-\infty,x]}(X_n)] \le \mathbb{E}[\varphi(X_n)] \le \mathbb{E}[\mathbf{1}_{(-\infty,x+\delta]}(X_n)] = F^{X_n}(x+\delta)$$

$$F^X(x) \le \mathbb{E}[\varphi(X)] \le F^X(x+\delta) \qquad \text{mit} \quad \mathbb{E}[\varphi(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[\varphi(X)]$$

Also folgt

$$\left.\begin{aligned} \limsup_{n\to\infty} F^{X_n}(x) \le \limsup_{n\to\infty} \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)] \le F^X(x+\delta) \\ \liminf_{n\to\infty} F^{X_n}(x+\delta) \ge \liminf_{n\to\infty} \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)] \ge F^X(x) \end{aligned}\right. \qquad \forall x \in \mathbb{R}, \delta > 0$$

Ist F^X stetig bei x, so gilt $\lim_{\delta\to 0} F^X(x \pm \delta) = F^X(x)$ und somit

$$\limsup_{n\to\infty} F^{X_n}(x) = \liminf_{n\to\infty} F^{X_n}(x) = F^X(x)$$

(b)⇒(a): Zu φ stetig, beschränkt und $\delta > 0$ wähle Stetigkeitspunkte $x_1 < \ldots < x_K$ von F^X mit $F^X(x_1) < \delta, F^X(x_k) > 1 - \delta$ und

$$\forall x \in [x_{k-1}, x_k] : |\varphi(x) - \varphi(x_k)| < \delta, \quad k = 2, \ldots, K$$

Diese Wahl ist möglich, weil F^X als monotone Funktion nur abzählbar viele Unstetigkeitsstellen hat sowie φ auf einem kompakten Intervall gleichmäßig stetig ist. Es folgt

$$\begin{aligned} \mathbb{E}[\varphi(X_n)] &= \mathbb{E}[\varphi(X_n)\mathbf{1}_{(-\infty,x_1]}(X_n)] + \mathbb{E}[\varphi(X_n)\mathbf{1}_{(x_K,\infty)}(X_n)] \\ &\quad + \sum_{k=2}^{K} \mathbb{E}[\varphi(X_n)\mathbf{1}_{(x_{k-1},x_k)}(X_n)] \\ &\le \|\varphi\|_\infty \cdot (F^{X^n}(x_1) + 1 - F^{X_n}(x_K)) + \sum_{k=2}^{K} (\varphi(x_k) + \delta)(F^{X_n}(x_k) - F^{X_n}(x_{k-1})) \end{aligned}$$

und somit

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{E}[\varphi(X_n)] &\overset{(b)}{\le} \|\varphi\|_\infty (F^X(x_1) + 1 - F^X(x_K)) \\ &\quad + \sum_{k=2}^{K} (\varphi(x_k) + \delta)(F^X(x_k) - F^X(x_{k-1})) \end{aligned}$$

Andererseits gilt

$$\mathbb{E}[\varphi(X)] \geq -\|\varphi\|_\infty \{\underbrace{F^X(x_1)}_{<\delta} + \underbrace{1 - F^X(x_K)}_{<\delta}\} + \sum_{k=2}^{K}(\varphi(x_k) - \delta)(F^X(x_k) - F^X(x_{k-1}))$$

$$\Longrightarrow \limsup_{n\to\infty} \mathbb{E}[\varphi(X_n)] \leq 4\delta\|\varphi\|_\infty + \mathbb{E}[\varphi(X)] + 2\delta \underbrace{\sum_{k=2}^{K}(F^X(x_k) - F^X(x_{k-1}))}_{\leq 1}$$

Mit $\delta \downarrow 0$ folgt also $\limsup_{n\to\infty} \mathbb{E}[\varphi(X_n)] \leq \mathbb{E}[\varphi(X)]$. Vollkommen analog erhält man (ersetze φ durch $-\varphi$) $\liminf_{n\to\infty} \mathbb{E}[\varphi(X_n)] \geq \mathbb{E}[\varphi(X)]$. Daher gilt

$$\lim_{n\to\infty} \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)]$$

□

Bemerkung Im Beweis (a)⇒(b) reicht es aus, $\mathbb{E}[\varphi(X_n)] \to \mathbb{E}[\varphi(X)]$ nur für $\varphi \in \mathcal{C}^\infty(\mathbb{R})$ mit Ableitung $\varphi^{(m)}$ zu wissen, deren Träger kompakt ist. Zum Nachweis der Konvergenz in Verteilung reicht es also, nur solche φ zu betrachten.

Beispiel V.22 (a) $X_n \sim \mathcal{N}(0, \sigma_n^2)$ mit $\sigma_n \to 0$. Dann gilt $X_n \xrightarrow{d} \delta_0$, denn $F^{X_n}(x) \to \mathbf{1}_{[0,\infty)}(x)$ für alle $x \neq 0$ (beachte: $F^{X_n}(0) = \frac{1}{2}$.)

(b) $\mathrm{U}_1, \ldots, U_n \sim \mathrm{U}([0,1])$ unabhängig, $X_n := n \cdot \min(\mathrm{U}_1, \ldots, \mathrm{U}_n)$.

$$x \geq 0: \quad F^{X_n}(x) = P(\min\{\mathrm{U}_1, \ldots, \mathrm{U}_n\} \leq \tfrac{x}{n}) = 1 - (1 - \tfrac{x}{n})_+^n \xrightarrow{n\to\infty} 1 - e^{-x}$$

Verteilungsfunktion von $\mathrm{Exp}(1)$-Verteilung.

$$\Longrightarrow X_n \xrightarrow{d} \mathrm{Exp}(1)$$

Später werden wir folgendes zentrale Resultat der Stochastik beweisen

Satz V.23 (Zentraler Grenzwertsatz)
Es sei $(X_k)_{k\geq1}$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen in $\mathcal{L}^2$ mit $\mathbb{E}[X_k] = \mu$, $\mathrm{Var}(X_k) = \sigma^2 > 0$. Dann gilt für die standardisierte Summe

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{x_k - \mu}{\sigma}: \qquad S_n^* \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$$

also $P(S_n^* \leq x) \to \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy \; \forall x \in \mathbb{R}$.

Satz V.24 (Auswahlsatz von Helly)
Ist $(P_n)_{n\geq1}$ eine Folge von W-Maßen auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ mit Verteilungsfunktionen $(F_n)_{n\geq1}$, so existiert eine Teilfolge $(n_k)_{k\geq1}$ und monoton wachsende, rechtsstetige Funktion $F : \mathbb{R} \to [0,1]$ mit $F_{n_k} \xrightarrow{k\to\infty} F(x)$ für alle Stetigkeitspunkte x von F.

Beweis: Es sei $(q_n)_{n\geq1}$ eine Abzählung von $\mathbb{Q}$, d.h. $\mathbb{Q} = \{q_n \mid n \geq 1\}$. Da $(F_n(q_1))_{n\geq1}$ beschränkt ist, gibt es eine Teilfolge $(n_1(k))_{k\geq1}$ mit $F_{n_1(k)}(q_1) \xrightarrow{k\to\infty} H(q_1)$ für eine reelle

Zahl $H(q_1)$. Ebenso gibt es eine Teilfolge $(n_2(k))_{k\geq 1}$ von $(n_1(k))_{k\geq 1}$ von $F_{n_2(k)}(q_2) \to H(q_2)$ für $H(q_2) \in \mathbb{R}$ geeignet.

Rekursiv erhalten wir Teilfolgen $(n_l(k))_{k\geq 1}$ mit $F_{n_l(k)}(q_l) \xrightarrow{k\to\infty} H(q_l)$. Die *Diagonalfolge* $(n_k(k))_{k\geq 1}$ erfüllt daher $F_{n_k(k)}(q_l) \xrightarrow{k\to\infty} H(q_l) \forall l \in \mathbb{N}$. Mit $n(k) := n_k(k)$ gilt also $F_{n_k}(q) \to H(q)$ für alle $q \in \mathbb{Q}$. Offensichtlich ist $H : \mathbb{Q} \to [0,1]$ monoton wachsend.

Setze $F(x) := \lim_{\substack{q\downarrow x\\ q\in\mathbb{Q}}} H(q) = \inf_{\substack{q>x\\ q\in\mathbb{Q}}} H(q)$, $x \in \mathbb{R}$. Dann ist F rechtsstetig per Konstruktion. Zeige nun: F stetig in $x \implies F_{n_k}(x) \to F(x)$. Wähle dazu $r_1, r_2, s \in \mathbb{Q}$ mit $r_1 < r_2 < x < s$ und $F(x) - \varepsilon \leq F(r_1) \leq F(s) \leq F(x) + \varepsilon$ (möglich, wenn F stetig bei x).

$$\implies \begin{aligned} &\limsup_{k\to\infty} Fn_k(x) \overset{\text{Mon.}}{\leq} \limsup_{k\to\infty} F_{n_k}(s) \underset{\text{von } H}{\overset{\text{Def}}{=}} H(s) \leq F(s) \leq F(x) + \varepsilon \\ &\liminf_{k\to\infty} Fn_k(x) \overset{\text{Mon.}}{\geq} \liminf_{k\to\infty} F_{n_k}(r_2) \underset{\text{von } H}{\overset{\text{Def}}{=}} H(r_2) \geq F(r_1) \geq F(x) - \varepsilon \end{aligned}$$

$$\overset{\varepsilon\downarrow 0}{\implies} \lim_{k\to\infty} F_{n_k}(x) = F(x) \qquad \square$$

Beispiel V.25 Ist $P_n = \mathrm{U}([n, n+1])$, so gilt für die Verteilungsfunktionen $F_n(x) \to 0$ für alle $x \in \mathbb{R}$, d.h. $F(x) = 0$ im obigen Satz. Um sicherzustellen, dass im Grenzwert wieder ein Wahrscheinlichkeitsmaß erscheint, muss man verhindern, dass „Masse“ von den $(P_n)_{n\geq 1}$ nach $\pm\infty$ läuft.

Definition V.26

Eine Folge $(P_n)_{n\geq 1}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ heißt (gleichartig) straff *falls für jedes $\varepsilon > 0$ ein $K_\varepsilon > 0$ existiert mit $\sup_{n\geq 1} P_n([-K_\varepsilon, K_\varepsilon]^C) < \varepsilon$.*

Korollar V.27

Ist $(P_n)_{n\geq 1}$ eine straffe Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$, so gibt es eine Teilfolge (n_k) und ein Wahrscheinlichkeitsmaß P auf $(\mathbb{R}, \mathcal{B}_\mathbb{R})$ mit $P_{n_k} \xrightarrow{k\to\infty} P$. (Kompaktheitskriterium)

Beweis: Nach Konstruktion des Lebesgue-Stieltjes-Maßes gibt es ein Maß P mit

$$P((a,b]) = F(b) - F(a) \quad \text{für} \quad a < b$$

und F aus dem Satz. Es bleibt zu zeigen, dass P ein Wahrscheinlichkeitsmaß ist, d.h. $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$. Sei dazu $K_\varepsilon > 0$ mit $\sup_{n\geq 1} P_n([-K_\varepsilon, K_\varepsilon]^C) < \varepsilon$ und $x > K_\varepsilon$ so, dass $\pm x$ Stetigkeitspunkt von F ist. Dann folgt

$$1 - F(x) + F(-x) = \lim_{k\to\infty} (1 - F_{n_k}(x) + F_{n_k}(-x)) \leq \sup_{n\geq 1} (\underbrace{1 - F_n(x)}_{P_n((x,\infty))} + \underbrace{F_n(-x)}_{P_n((-\infty,-x]}) \leq \varepsilon$$

und $\varepsilon \downarrow 0$ liefert die Behauptung. $\square$

V.3. Charakteristische Funktionen

Definition V.28

Für eine reellwertige Zufallsvariable X bezeichnet

$$\varphi^X(u) := \mathbb{E}[e^{iu\cdot X}] = \mathbb{E}[\cos(u\cdot X)] + i\mathbb{E}[\sin(u\cdot X)], \qquad u \in \mathbb{R}$$

die charakteristische Funktion *von X. Analog ist für ein Wahrscheinlichkeitsmaß P auf* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$

$$\varphi^P(u) := \int_{\mathbb{R}} e^{iux} P(dx) = \int_{\mathbb{R}} \cos(ux) P(dx) + i \int_{\mathbb{R}} \sin(ux) P(dx), \qquad u \in \mathbb{R}$$

die char. Funktion von P.

Lemma V.29

Die charakteristische Funktion erfüllt $\varphi(0) = 1$, $|\varphi(u)| \leq 1 \forall u \in \mathbb{R}$ *und* φ *ist gleichmäßig stetig.*

Beweis:

$$\varphi(0) = \int e^{i0x} P(dx) = \int 1 P(dx) = 1$$

$$|\varphi(u)| = \left| \int e^{iux} P(dx) \right| \overset{\Delta\text{-Ungl.}}{\leq} \int \underbrace{|e^{iux}|}_{=1} P(dx) = 1$$

$$|\varphi(u) - \varphi(v)| = \left| \int (e^{iux} - e^{ivx}) P(dx) \right| \leq \int |e^{iux} - e^{ivx}| P(dx)$$

$$= \int \underbrace{|e^{i(u-v)x} - 1|}_{\substack{|u-v| \to 0 \\ \to 0}} P(dx) \xrightarrow[|u-v| \to 0]{\text{DCT}} 0$$

und damit ist φ gleichmäßig stetig. □

Beispiel V.30 (a) $P = \operatorname{Bin}(n, p) : \varphi^P(u) = \int_{\mathbb{R}} e^{iux} P(dx) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot e^{iku}$

$$= \sum_{k=0}^{n} \binom{n}{k} (p \cdot e^{iu})^k (1-p)^{n-k} = (p \cdot e^{iu} + 1 - p)^n$$

(b) $P = \operatorname{Exp}(\lambda) : \varphi^p(u) = \int_{\mathbb{R}} e^{iux} P(dx) = \int_0^{\infty} e^{iux} \cdot \lambda e^{-\lambda x} \, dx$

$$= \lambda \int_0^{\infty} e^{(iu-\lambda)x} \, dx = \lambda \frac{e^{(iu-\lambda)x}}{iu - \lambda} \Bigg|_{x=0}^{+\infty} = -\lambda \frac{1}{iu - \lambda} = \left(1 - \frac{iu}{\lambda}\right)^{-1}$$

(c) $P = \mathcal{N}(0,1) : \varphi^P(u) = \int_{\mathbb{R}} e^{iux} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx$

$$= e^{-\frac{u^2}{2}} \underbrace{\int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-iu)^2}{2}\right) dx}_{G(u)} \overset{(*)}{=} e^{-\frac{u^2}{2}}$$

Nun ist $G : \mathbb{C} \to \mathbb{C}$ holomorph mit $G(ir) = 1$ für alle $r \in \mathbb{R}$. Daraus folgt $G(u) = 1$ für alle $u \in \mathbb{C}$ und $x \to \infty$ (Eindeutigkeitssatz). Damit folgt $(*)$.

Lemma V.31

(a) Sind X, Y *unabhängige Zufallsvariablen, so gilt* $\varphi^{X+Y}(u) = \varphi^X(u) \cdot \varphi^Y(u)$, $u \in \mathbb{R}$.

(b) Für $X \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$ *mit* $p \in \mathbb{N}$ *ist* $\varphi^X \in \mathcal{C}^p(\mathbb{R})$ *und es gilt*

$$(\varphi^X)^{(k)}(0) = i^k \mathbb{E}[X^k], \quad k = 0, 1, \ldots, p$$

Beweis: Übungsaufgabe.

Satz V.32 (Eindeutigkeit der char. Funktion)

Zwei Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *mit derselben charakteristischen Funktion sind identisch.*

Beweis: Zeige die Inversionsformel für $a < b$

$$P((a,b)) + \tfrac{1}{2}P(\{a,b\}) = \lim_{u\to\infty} \frac{1}{2\pi} \int_{-u}^{u} \frac{e^{-iua} - e^{-iub}}{iu} \varphi^P(u)\, du$$

wobei der Integrand bei $u = 0$ stetig ergänzt sei. Wir verwenden $\lim_{u\to\infty} \int_{-u}^{u} \frac{\sin x}{x}\, dx = \pi$ aus der Analysis. Setze

$$\begin{aligned}
I(u) &= \int_{-u}^{u} \frac{e^{-i\tilde{u}a} - e^{-i\tilde{u}b}}{i\tilde{u}} \varphi^P(\tilde{u})\, d\tilde{u} \\
&\overset{\substack{\text{Def}\\ \text{Char.Fkt.}}}{=} \int_{-u}^{u} \left(\int_{-\infty}^{\infty} \frac{e^{-i\tilde{u}a} - e^{-i\tilde{u}b}}{i\tilde{u}} e^{i\tilde{u}x} P(\,dx) \right) d\tilde{u} \\
&\overset{\substack{\text{Fubini}\\ \text{beschr.}\\ \text{Integrand}}}{=} \int_{-\infty}^{\infty} \int_{-u}^{u} \frac{e^{-i\tilde{u}(a-x)} - e^{-i\tilde{u}(b-x)}}{i\tilde{u}}\, d\tilde{u} P(\,dx) \\
&\overset{\substack{\text{Symmetrie}\\ \text{in } \tilde{u}}}{=} \int_{-\infty}^{\infty} \int_{-u}^{u} \frac{\sin(\tilde{u}(x-a)) - \sin(\tilde{u}(x-b))}{\tilde{u}}\, d\tilde{u} P(\,dx) \\
&\overset{\substack{\text{subst.}\\ v=|\tilde{u}(x-a)|}}{=} \int_{-\infty}^{\infty} \left(\operatorname{sgn}(x-a) \underbrace{\int_{-u|x-a|}^{u|x-a|} \frac{\sin v}{v}\, dv}_{\xrightarrow{u\to\infty} \begin{cases} \pi, & x \neq a \\ 0, & x = a \end{cases}} - \operatorname{sgn}(x-b) \underbrace{\int_{-u|x-b|}^{u|x-b|} \frac{\sin v}{v}\, dv}_{\begin{cases} \pi, & x \neq b \\ 0, & x = b \end{cases}} \right) P(\,dx) \\
&\xrightarrow[u\to\infty]{\substack{\text{dominierte}\\ \text{Konvergenz}}} \int_{-\infty}^{\infty} \begin{Bmatrix} \pi - \pi, & x < a \text{ oder } x > b \\ \pi \pm 0, & x = a \text{ oder } x = b \\ \pi - (-\pi), & a < x < b \end{Bmatrix} P(\,dx) \\
&= 2\pi P((a,b)) + \pi P(\{a,b\})
\end{aligned}$$

Da es nur abzählbar viele $a, b \in \mathbb{R}$ gibt mit $P(\{a,b\}) > 0$ erhalten wir

$$F(b) - F(a) = P((a,b]) = \lim_{u\to\infty} \frac{1}{2\pi} \int_{-u}^{u} \frac{e^{-iua} - e^{-iub}}{iu} \varphi^P(u)\, du$$

für alle bis auf abzählbar viele $a < b$. Wegen Rechtsstetigkeit ist die Verteilungsfunktion F daher eindeutig festgelegt. □

Beispiel V.33 (a) S_n sei Bin(n,p)-verteilt (d.h. $S_n = \sum_{i=1}^n X_i$, $X_i \sim \text{Bin}(1,p)$ unabh.).

$$S_n^* = \frac{S_n - np}{\sqrt{npq}}, \quad q := 1 - p \implies \mathbb{E}[S_n^*] = 0, \quad \operatorname{Var}(S_n^*) = 1$$

$$\begin{aligned}
\varphi^{S_n^*} &= \mathbb{E}\left[e^{iuS_n^*}\right] = \mathbb{E}\left[e^{iu\frac{S_n - np}{\sqrt{npq}}}\right] = \varphi^{S_n}\left(\frac{u}{\sqrt{npq}}\right) e^{\frac{-iunp}{\sqrt{npq}}} \overset{\substack{\text{Üb.}\\ \text{Lemma}}}{=} \left(\varphi^{X_1}\left(\frac{u}{\sqrt{npq}}\right)\right)^n e^{-\frac{iunp}{\sqrt{npq}}} \\
&= \left(p \cdot e^{i\frac{u}{\sqrt{npq}}} + q \cdot 1\right)^n e^{-iu\frac{up}{\sqrt{npq}}}
\end{aligned}$$

$$\overset{\text{Taylor}}{=} \left(p\left(1+\cancel{\frac{iuq}{\sqrt{npq}}}+\frac{1}{2}\left(\frac{iuq}{\sqrt{npq}}\right)^2+o(\tfrac{1}{n})\right)\right)+q\left(1-\cancel{\frac{iup}{\sqrt{npq}}}+\frac{1}{2}\left(\frac{iup}{\sqrt{npq}}\right)^2+o(\tfrac{1}{n})\right)^n$$

$$=\left(1-\frac{1}{2}\left(\frac{u^2\cancel{p}q^{\cancel{2}}}{n\cancel{p}\cancel{q}}+\frac{u^2p^{\cancel{2}}\cancel{q}}{n\cancel{p}\cancel{q}}\right)+o(\tfrac{1}{n})\right)^n=\left(1-\frac{u^2}{2n}+o(\tfrac{1}{n})\right)^n \xrightarrow{n\to\infty} e^{-\frac{u^2}{2}} \quad \forall u\in\mathbb{R}$$

Beachte dazu $\left(1+\frac{z}{n}+o\left(\frac{1}{n}\right)\right)^n \to e^z$ weil

$$n\log\left(1+\tfrac{z}{n}+o(\tfrac{1}{n})\right) \overset{\text{Taylor}}{=} n\left(\tfrac{z}{n}+\underbrace{o(\tfrac{1}{n})+\mathcal{O}\left(\left(\tfrac{z}{n}+o(\tfrac{1}{n})\right)^2\right)}_{o(\frac{1}{n})}\right) \xrightarrow{n\to\infty} z$$

Die char. Funktionen von S_n^* konvergieren also für $n\to\infty$ gegen die char. Funktion $\varphi^{\mathcal{N}(0,1)}(u)=e^{-\frac{u^2}{2}}$. Für ein Bild siehe [3], S. 132.

(b) (vgl. Übung) Es seien $X_j \sim \mathrm{U}([-1,1])$, $j\geq 1$ unabhängig

$$S_n^*=\frac{\sqrt{3}}{\sqrt{n}}\sum_{j=1}^n X_j \implies \mathbb{E}[S_n^*]=0,\quad \mathrm{Var}(S_n^*)=1$$

$$\varphi^{S_n^*}(u)=\mathbb{E}\left[e^{iu\frac{\sqrt{3}}{\sqrt{n}}\sum_{j=1}^n X_j}\right]=\left(\mathbb{E}\left[e^{iu\frac{\sqrt{3}}{\sqrt{n}}X_1}\right]\right)^n=\left(\frac{1}{2}\int_{-1}^1 e^{iu\frac{\sqrt{3}}{\sqrt{n}}\cdot x}\,dx\right)^n=\left(\frac{\sin\left(\sqrt{\frac{3}{n}}u\right)}{\sqrt{\frac{3}{n}}\cdot u}\right)^n$$

$$\overset{\text{Taylor}}{=}\left(1-\frac{\left(\sqrt{\frac{3}{n}}u\right)^2}{3!}+o(\tfrac{1}{n})\right)^n=\left(1-\frac{u^2}{2n}+o(\tfrac{1}{n})\right)^n\to e^{-\frac{u^2}{2}}=\varphi^{\mathcal{N}(0,1)}(u)$$

Gilt $X_n \overset{d}{\to} X$, so folgt aus der Definition sofort $\varphi^{X_n}(u)\to\varphi^X(u)$ punktweise für alle $u\in\mathbb{R}$ (beachte $\sin(ux),\cos(ux)$ sind stetig, beschränkt). Überraschenderweise gilt auch die Umkehrung unter schwachen Voraussetzungen:

Satz V.34 (Stetigkeitssatz von Lévy)
Sind $(P_n)_{n\geq 1}$ Wahrscheinlichkeitsmaße auf $(\mathbb{R},\mathcal{B}_\mathbb{R})$ mit char. Funktionen φ_n und gilt $\varphi_n(u)\xrightarrow{n\to\infty}\psi(u)$, $u\in\mathbb{R}$, mit einer bei $u=0$ stetigen Funktion $\psi:\mathbb{R}\to\mathbb{C}$, so ist $\psi=\varphi^P$, die char. Funktion eines Wahrscheinlichkeitsmaßes P und es gilt $P_n\overset{w}{\to}P$.

Beweis: (a) Zeige die Ungleichung $P([-\frac{2}{n},\frac{2}{n}]^C)\leq\frac{1}{4}\int_{-u}^u(1-\varphi^P(v))\,dv$, $u>0$. Es gilt

$$\frac{1}{u}\int_{-u}^u(1-e^{ivx})\,dv=\frac{1}{u}\left(2u-\int_{-u}^u\cos(vx)\,dv\right)=2-2\frac{\sin(ux)}{ux} \qquad (*)$$

$$\overset{\text{Fubini}}{\implies}\frac{1}{u}\int_{-u}^u(1-\varphi^P(v))\,dv\overset{(*)}{=}2\int_\mathbb{R}\left(1-\frac{\sin(ux)}{ux}\right)P(\,dx)$$

$$\geq 2 \cdot \int_{|x| \geq \frac{2}{u}} \underbrace{\left(1 - \overbrace{\frac{1}{|ux|}}^{\leq \frac{1}{2}}\right)}_{\geq \frac{1}{2}} P(dx) \geq P([-\tfrac{2}{u}, \tfrac{2}{u}]^C)$$

(b) Zeige: $(P_n)_{n\geq 1}$ ist straff. Zu $\varepsilon > 0$ wähle aufgrund der Stetigkeit von ψ bei Null ein $u > 0$ mit $\frac{1}{u}\int_{-u}^{u}(1-\psi(v))\,dv < \varepsilon$. Dominierte Konvergenz impliziert

$$\exists N \geq 1 \forall n : \frac{1}{u}\int_{-u}^{u}(1-\varphi_n)(v))\,dv \leq 2\varepsilon \overset{\text{(a)}}{\Longrightarrow} \forall n \geq N : P_n([-\tfrac{2}{n}, \tfrac{2}{n}]^C) \leq 2\varepsilon$$

Für ein $R \geq \frac{2}{n}$ gilt daher $\forall n \geq 1 : P_n([-R,R]^C) \leq 2\varepsilon \implies (P_n)_{n\geq 1}$ ist straff.

(c) Schluss: Nach dem Satz von Helly (Satz 5.24) existiert eine Teilfolge n_k und ein Wahrscheinlichkeitsmaß P mit $P_{n_k} \xrightarrow[k\to\infty]{w} P$. Dann gilt auch (s.o.) $\varphi_{n_k}(u) \to \varphi^P(u)$ für alle $u \in \mathbb{R}$, d.h. $\psi(u) = \varphi^P(u) \forall u \in \mathbb{R}$. Also besitzt jede Teilfolge von (P_n) eine gegen P schwach konvergierende Teilteilfolge. Damit folgt $P_n \overset{w}{\to} P$. □

V.4. Zentraler Grenzwertsatz

Wir haben in obigen Beispielen für Binomial- und gleichmäßiger Verteilung gesehen, dass $\varphi^{S_n^*}(u) \xrightarrow{n\to\infty} \varphi^{\mathcal{N}(0,1)}(u)$. Nach dem Stetigkeitssatz von Lévy folgt also $S_n^* \overset{d}{\to} \mathcal{N}(0,1)$. Dies gilt in größter Allgemeinheit für standardisierte Summen unabhängiger Zufallsvariablen:

Satz V.35 (Zentraler Grenzwertsatz)
Es sei $(X_j)_{j\geq 1}$ eine Folge unabhängiger, identisch verteilter Zufallsvariablen in $\mathcal{L}^2$ und $\mu = \mathbb{E}[X_j]$ sowie $\sigma^2 = \mathrm{Var}(X_j) > 0$. Dann gilt

$$S_n^* := \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{X_j - \mu}{\sigma} \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$$

Beweis: Es reicht nach Lévy zu zeigen: $\varphi^{S_n^*}(u) \to \varphi^{\mathcal{N}(0,1)}(u) = e^{-\frac{u^2}{2}}$. Es gilt mit $\tilde{X}_j = \frac{X_j-\mu}{\sigma}$:

$$\begin{aligned}
\varphi^{S_n^*}(u) &= \mathbb{E}\left[e^{\frac{iu}{\sqrt{n}}\sum_{i=1}^{n}\tilde{X}_j}\right] = \left(\mathbb{E}[e^{i\frac{u}{\sqrt{n}}\tilde{X}_1}]\right)^n \\
&\underset{\text{Taylor}}{\overset{\text{Übung}}{=}} \left(1 + \frac{iu}{\sqrt{n}}\underbrace{\mathbb{E}[\tilde{X}_1]}_{=0} + \tfrac{1}{2}\left(\frac{iu}{\sqrt{n}}\right)^2 \mathbb{E}[\tilde{X}_1^2] + o\left(\left(\frac{u}{\sqrt{n}}\right)^2\right)\right)^n \\
&= \left(1 - \frac{1}{2}\frac{u^2}{n} + o(\tfrac{1}{n})\right)^n \xrightarrow[\text{s.o. Beispiel}]{n\to\infty} e^{-\frac{u^2}{2}} = \varphi^{\mathcal{N}(0,1)}(u)
\end{aligned}$$

□

Interpretation des ZGWS

Der ZGWS besagt, dass bei vielen unabhängigen und identisch verteilten stochastischen Fehlern (Störgrößen) die standardisierte Summe gerade approximativ $\mathcal{N}(0,1)$-verteilt ist. Bei physikalischen Messungen z.B. wird daher gerne von normalverteilten Fehlern ausgegangen. Die allgemeine ZGWS von Lindeberg – Fehler gibt eine hinreichende und notwendige Bedingung für $S_n^* \to \mathcal{N}(0,1)$, wenn die X_i unabhängig, aber nicht unbedingt identisch verteilt sind. Ebenso gibt es Verallgemeinerungen für abhängige Zufallsvariablen („Mischungskoeffizienten")

Wann gilt für die Binomialverteilung die Normal- und wann die Poisson-Approximation? (ZGWS vs. Poissonscher GWS (I.12). Sei dazu $S_n \sim \text{Bin}(n,p)$, $0 < p < 1$, dann gilt einerseits

$$\begin{aligned} &\text{(ZGWS)} \quad S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0,1) \\ \Longrightarrow \; &P(a < S_n^* \le b) = P\left(np + a\sqrt{np(1-p)} < S_n \le np + b\sqrt{np(1-p)}\right) \to \varphi(b) - \varphi(a) \end{aligned}$$

und andererseits

$$\text{(Poiss. GWS)} \quad S_n \xrightarrow{d} \text{Poiss}(\lambda) \quad \text{falls } np_n \to \lambda \text{ (insbesondere } p_n \to 0)$$

Grundsätzlicher Unterschied: Beim Poiss. GWS müssen die Einzelwahrscheinlichkeiten p_n klein sein und S_n nimmt Werte asymptotisch in $\mathbb{N}_0$ an. Beim ZGWS ist $p_n = p$ =const. (beliebig), die Werte von S_n^* sind zwar diskret, aber verteilen sich asymptotisch auf der reellen Achse.

Genauer muss man den Approximationsfehler in beiden GWS genauer untersuchen. Im Poissonschen GWS gilt:

Satz V.36
Für $p \in (0,1)$ und $n \in \mathbb{N}$ gilt

$$\sum_{k \ge 0} |\,\text{Bin}_{n,p}(k) - \text{Poiss}_{n \cdot p}(k)| \le 2np^2$$

Beweis: („coupling"-Argument) Seien die Zufallsvariablen $X_i \sim \text{Bin}(1,p)$ unabhängig („Bernoulli-Folge"), $1 \le i \le n$ und $S_n = \sum_{i=1}^n X_i \sim \text{Bin}(n,p)$ sowie $Y_i \sim \text{Poiss}(p)$ unabhängig, $1 \le i \le n$ und $T_n = \sum_{i=1}^n Y_i \sim \text{Poiss}(np)$. Wir werden nun die Familien $(X_i)_{1 \le i \le n}$ *nicht* unabhängig voneinander wählen, sondern geschickt auf denselben Wahrscheinlichkeitsraum miteinander koppeln. Konstruiere dazu unabhängige Zufallsvariablen $Z_1, \ldots, Z_n$ mit Werten in $\mathbb{N}_0 \cup \{-1\}$, so dass

$$P(Z_i = k) = \begin{cases} \text{Poiss}_p(k) & \text{für } k \ge 1 \\ 1 - p & \text{für } k = 0 \\ e^{-p} - (1-p) & \text{für } k = -1 \end{cases}$$

gilt (beachte $e^{-p} \ge 1 - p$). Setze $X_i = \mathbf{1}_{\{Z_i \ne 0\}}$, $Y_i := \max(Z_i, 0)$. Dann gilt $X_i \sim \text{Bin}(1,p)$ sowie $Y_i \sim \text{Poiss}(p)$, und $(X_i)_{1 \le i \le n}$ und $(Y_i)_{1 \le i \le n}$ sind jeweils unabhängige Familien von Zufallsvariablen, da $(Z_i)_{1\le i\le n}$ unabhängig ist. Damit erhalten wir

$$\begin{aligned}
&\sum_{k\ge 0} |\operatorname{Bin}_{n,p}(k) - \operatorname{Poiss}_{np}(k)| \\
&= \sum_{k\ge 0} |P(S_n = k) - P(T_n = k)| \\
&= \sum_{k\ge 0} \Bigg| P\Big(\{S_n = k\} \cap \bigcup_{1\le i\le n} \{X_i \ne Y_i\}\Big) - P\Big(\{T_n = k\} \cap \bigcup_{1\le i\le n} \{X_i \ne Y_i\}\Big) \\
&\qquad + \cancel{P\Big(\{S_n = k\} \cap \bigcap_{1\le i\le n} \{X_i = Y_i\}\Big)} - \cancel{P\Big(\{T_n = k\} \cap \bigcap_{1\le i\le n} \{X_i = Y_i\}\Big)} \Bigg| \\
&\le 2P\Big(\bigcup_{1\le i\le n} \{X_i \ne Y_i\}\Big) = 2P\Big(\bigcup_{1\le i\le n} \{Z_i \notin \{0,1\}\}\Big) \\
&\le 2n(1 - e^{-p} - e^{-p}p + e^{-p} - (1-p)) = 2n(1-e^{-p})p \le 2np^2
\end{aligned}$$

□

Bemerkung Im ZGWS gilt unter der Bedingung $X_i \in L^3$ folgende Fehlerabschätzung (*Satz von Berry-Esséen*, siehe z.B. [2])

$$\|F^{S_n^*} - \varphi\|_\infty \le 0.8 \cdot \frac{\mathbb{E}[|X_i - \mu|^3]}{\sigma^{\frac{3}{2}}} \cdot \frac{1}{\sqrt{n}} = \mathcal{O}(n^{-\frac{1}{2}})$$

Man kann sogar eine asymptotische Entwicklung (*Edgeworth-Entwicklung*) der Form

$$F^{S_n^*}(x) = \varphi(x) + \frac{1}{\sqrt{n}} \cdot H_1(x) + \frac{1}{n} H_2(x) + \cdots + \frac{1}{n^{\frac{k}{2}}} H_k(x) + o(n^{-\frac{k}{2}})$$

herleiten mit Koeffizienten $H_k(x)$, die auf *Hermite-Polynomen* beruhen.

VI. Einführung in die Schätztheorie

VI.1. Grundbegriffe

Definition VI.1
Es sei $(\mathcal{X} = \Omega, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell sowie $g : \Theta \to \mathbb{R}^d$ eine Funktion. Für jeden Parameter ϑ heißt $g(\vartheta)$ abgeleiteter Parameter *. Jede messbare Funktion $\hat{g} : \mathcal{X} \to (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ heißt* Schätzer *von $g(\vartheta)$. Für eine* Realisierung *(Daten) $x \in \mathcal{X}$ heißt $\hat{g}(x)$* Schätzung *oder* Schätzwert.

In der Statistik wird das Verhalten des Schätzers $\hat{g}$ als Zufallsvariable bezüglich der wahren, aber unbekannten Wahrscheinlichkeitsverteilung P_ϑ untersucht. Ein häufig angewandtes Gütekriterium ist der MSE (*mean squared error*):

Definition VI.2 (Mittlerer quadratischer Fehler, MSE)
Der mittlere quadratische Fehler *eines Schätzers $\hat{g}$ von $g(\vartheta)$ ist gegeben durch*

$$R(\hat{g}, \vartheta) := \mathbb{E}_\vartheta\left[|\hat{g} - g(\vartheta)|^2\right] \quad \text{(Risikofunktion)}$$

wobei $|\cdot|$ *die Euklidische Norm bezeichne. Die Differenz*

$$B(\hat{g},\vartheta) := \mathbb{E}_\vartheta[\hat{g}] - g(\vartheta)$$

heißt Verzerrung *oder* Bias *von* $\hat{g}$ *bei* ϑ*. Gilt* $B(\hat{g},\vartheta) = 0$ *für alle* $\vartheta \in \Theta$*, so heißt* $\hat{g}$ erwartungstreu *oder* unbiased.

Beispiel VI.3 $\mathcal{X} = \mathbb{R}$, $\mathcal{F} = \mathcal{B}_\mathbb{R}$, $\Theta = \{\text{W-Maße } P \text{ auf } \mathcal{B}_\mathbb{R} \mid \int_{-\infty}^{\infty} x^2 P(\,dx) < \infty\}$, $P_\vartheta := \vartheta$ bzw. stat. Modell $(\mathbb{R}, \mathcal{B}_\mathbb{R}, \{\text{W-Maße } P^{\otimes n} \text{ auf } \mathcal{B}_\mathbb{R} \mid \int x^2 P(\,dx) < \infty\})$. $g : \Theta \to \mathbb{R}$ Erwartungswert, also $g(P) = \int_{-\infty}^{\infty} x P(\,dx)$. Stichprobenraum $\mathcal{X} = \mathbb{R}^n$, d.h. Beobachtungen $x_1, \ldots, x_n \in \mathbb{R}$, jedes $x_i \sim P$ und $x_1, \ldots, x_n$ unabhängig.

$$\hat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{Arithmetisches Mittel}$$

Es gilt $\mathbb{E}_p[\hat{g}] = \cancel{\tfrac{1}{n}\textstyle\sum_{i=1}^{n}} \int_\mathbb{R} x\,dP = g(P)$, d.h. $\hat{g}$ ist ist erwartungstreuer Schätzer von $g(P)$. MSE:

$$\begin{aligned} R(P,\hat{g}) = \mathbb{E}_P[(\hat{g} - g(P))^2] = \mathrm{Var}_P(\hat{g}) &\overset{\text{Unabh.}}{=} \frac{1}{n^{\cancel{2}}} \cancel{\sum_{i=1}^{n}} \int_\mathbb{R} (x - g(P))^2 P(\,dx) \\ &= \frac{1}{n} \underbrace{\int_\mathbb{R} (x - g(P))^2 P(\,dx)}_{\text{„Varianz von } P\text{“}} \end{aligned}$$

Außerdem gilt (starkes Gesetz der großen Zahlen) mit der Notation $\hat{g}_n = \hat{g} : \hat{g}_n \xrightarrow{P\text{-f.s.}} g(P) \forall P \in \Theta$ (in der Statistik sagt man, dass $\hat{g}_n$ ein (start) *konsistenter Schätzer* ist für $n \to \infty$).

Lemma VI.4 (Bias-Varianz-Zerlegung)
Für jeden Schätzer $\hat{g} \in \mathcal{L}^2(\mathcal{X}, \mathcal{F}, P_\vartheta)$ *gilt*

$$R(\hat{g},\vartheta) = \underbrace{B(\hat{g},\vartheta)^2}_{\mathit{Bias}^2} + \underbrace{\mathrm{Var}_\vartheta(\hat{g})}_{\mathit{Varianz}}$$

Beweis:

$$\begin{aligned} R(\hat{g},\vartheta) &= \mathbb{E}_\vartheta[(\hat{g} - \mathbb{E}_\vartheta[\hat{g}] + \mathbb{E}_\vartheta[\hat{g}] - g(\vartheta))^2] \\ &= \mathbb{E}_\vartheta[(\hat{g} - \mathbb{E}_\vartheta[\hat{g}])^2] + \overbrace{2\mathbb{E}[\hat{g} - \mathbb{E}_\vartheta[\hat{g}]]}^{=0}(\mathbb{E}_\vartheta[\hat{g}] - g(\vartheta)) + (\mathbb{E}_\vartheta[\hat{g}] - g(\vartheta))^2 \\ &= \mathrm{Var}_\vartheta(\hat{g}) + B(\hat{g},\vartheta)^2 \end{aligned}$$

□

Beispiel VI.5 (a) Bernoulli-Experiment $\mathcal{X} = \{0,1\}^n$, $\mathcal{F} = \mathcal{P}(\mathcal{X})$, P_ϑ habe Zähldichte

$$p^\vartheta(x_1, \ldots, x_n) = \vartheta^{\sum_{i=1}^{n} x_i} \cdot (1-\vartheta)^{\sum_{i=1}^{n}(1-x_i)}$$

$\vartheta \in \Theta = [0,1]$. Betrachte Identität $g(\vartheta) = \vartheta$ und den Schätzer $\hat{\vartheta}(x) := \frac{1}{n}\sum_{i=1}^{n} x_i$ (rel. Häufigkeit). Wie oben im Beispiel ist $\hat{\vartheta}$ erwartungstreu, konsistent für $n \to \infty$ und es gilt $R(\hat{\vartheta},\vartheta) = \frac{\vartheta(1-\vartheta)}{n}$. Stets gilt $R(\hat{\vartheta},\vartheta) \leq \frac{1}{4n}$, aber z.B. $R(\hat{\vartheta},0) = 0$.

Ein anderer Schätzer ist $\tilde{\vartheta}(x) = \frac{1}{\pi} : R(\vartheta,\tilde{\vartheta}) = (\frac{1}{\pi} - \vartheta)^2$. Man wird bei der Suche nach einem besten Schätzer im Allgemeinen die Klasse der Schätzer einschränken (z.B. Erwartungstreue) und/oder den in ϑ maximalen MSE bzw. gemittelten MSE betrachten.

(b) Taxiproblem: In einer Stadt werden alle N Taxis mit einer gut sichtbaren Lizenznummer zwischen 1 und N verstehen. Ein Tourist sieht Taxis mit den Nummern $X_1, \dots, X_n$ (ohne Doppelte). Welche sinnvolle Schätzung von N könnte ihm einfallen.

a) Er nimmt den Mittelwert und multipliziert ihn mit 2: $\hat{N}_1 = \frac{2}{n} \sum_{i=1}^n X_i$.

b) Er nimmt das Maximum $\hat{N}_2 = \max(X_1, \dots, X_n)$

c) Er addiert den mittleren Abstand zwischen Beobachtungen zu Maximum hinzu $\hat{N}_3 = \max(X_1, \dots, X_n) + \frac{1}{n-1} \underbrace{\sum_{i=1}^{n-1}(X_{(i)} - X_{(i+1)})}_{X_{(1)}-X_{(n)}}$ mit $(X_{(i)})_{i=1}^n$ die geordneten Zufallsvariablen.

d) betrachte beim letzten Beispiel auch den Abstand $X_{(n)}$ zu 0:

$$\hat{N}_4 = X_{(1)} + \frac{1}{n} X_{(1)} = \frac{n+1}{n} \max\{X_1, \dots, X_n\}.$$

Alle Vorschläge haben gewisse Meriten, aber wir wollen den bzgl. MSE optimalen Schätzer finden. Zur leichten Analyse (sonst siehe [6]) nehmen wir hier ein stetiges Modell an, nämlich $X_1, \dots, X_n \sim \mathrm{U}([0, \vartheta])$ unabhängig mit $\vartheta \in \Theta = (0, \infty)$ unbekannt (Approx. für N groß). Zufallszahlen glm. in $[0, \vartheta]$, ϑ zu schätzen).

a) $\hat{\vartheta}_1 = \frac{2}{n} \sum_{i=1}^n X_i \overset{\text{s.o.}}{\Longrightarrow} R(\hat{\vartheta}_1, \vartheta) = \frac{4}{n} \operatorname{Var}_\vartheta(X_1) = \frac{\vartheta^2}{3n}$

b) $\hat{\vartheta}_2 = \max(X_1, \dots, X_n) \implies R(\hat{\vartheta}_2, \vartheta) = B(\hat{\vartheta}_2, \vartheta)^2$,

$$\operatorname{Var}_\vartheta(\hat{\vartheta}_2) = \frac{1}{(n+1)^2}\vartheta^2 + \left(\frac{\vartheta^2 n}{n+2} - \frac{n^2}{(n+1)^2}\vartheta^2\right) = \vartheta^2 \cdot \frac{2}{(n+1)(n+2)}$$

c) $\hat{\vartheta}_3 = \frac{n+1}{n} \cdot \hat{\vartheta}_2 \implies R(\hat{\vartheta}_3, \vartheta) = \underbrace{B(\hat{\vartheta}_3, \vartheta)^2}_{=0} + \operatorname{Var}_\vartheta(\hat{\vartheta}_3) = (\frac{n+1}{n})^2 \operatorname{Var}_\vartheta(\hat{\vartheta}_2) = \vartheta^2 \cdot \frac{1}{n(n+2)}$

Wir stellen fest, dass $\hat{\vartheta}_3$ für alle $\vartheta > 0$ und (zumindest) große n den kleinsten MSE besitzt. Darüber hinaus ist $\hat{\vartheta}_3$ erwartungstreu. Man kann zeigen, dass $\hat{\vartheta}_3$ in der Tat bzgl. MSE optimaler Schätzer ist.

VI.2. Cramér-Rao-Ungleichung und Maximum-Likelihood-Prinzip

Satz VI.6 (Cramér-Rao-Schranke, 1943)
Es sei $(\mathcal{X}, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell mit $\Theta \subseteq \mathbb{R}$ offen und $\hat{g}$ ein erwartungstreuer Schätzer von $g(\vartheta)$ mit $g : \Theta \to \mathbb{R}$ differenzierbar. Weiterhin besitze P_ϑ eine Dichte $f_\vartheta : \mathbb{R} \to [0, \infty)$, so dass $\frac{\partial}{\partial\vartheta} f_\vartheta$ existiere für λ-f.a. $x \in \mathcal{X}$ und Vertauschungen seien erlaubt:

$$\frac{\partial}{\partial\vartheta} \int h(x) f_\vartheta(x)\, dx = \int h(x) \frac{\partial}{\partial\vartheta} f_\vartheta(x)\, dx$$

für $h(x) = 1$ und $h(x) = \hat{g}(x)$. Dann gilt

$$\forall \vartheta \in \Theta: \quad \mathbb{E}_\vartheta[|\hat{g} - g(\vartheta)|^2] \geq \frac{g'(\vartheta)^2}{I(\vartheta)}$$

sofern die Fisher-Information $I(\vartheta) = \mathbb{E}_\vartheta\left[\left(\frac{\frac{\partial}{\partial\vartheta} f_\vartheta}{f_\vartheta}\right)^2\right]$ *endlich ist.*

Beweis: Da $\hat{g}$ erwartungstreu ist, gilt die Bias-Varianzzerlegung (VI.4)

$$\mathbb{E}[(\hat{g} - g(\vartheta))^2] = \mathrm{Var}(\hat{g})$$

Für $Y = \frac{\frac{\partial}{\partial\vartheta} f_\vartheta(x)}{f_\vartheta(x)}$ gilt nun

$$\mathbb{E}_\vartheta[Y] = \int \frac{\partial}{\partial\vartheta} f_\vartheta(x)\, dx = \frac{\partial}{\partial\vartheta} \underbrace{\int f_\vartheta(x)\, dx}_{1} = 0$$

$$\mathbb{E}_\vartheta[Y^2] = I(\vartheta) \qquad (\Longrightarrow Y \in \mathcal{L}^2)$$

Wir erhalten

$$\begin{aligned} \mathrm{Cov}_\vartheta(\hat{g}, Y)^2 &\leq \mathrm{Var}_\vartheta(\hat{g})\,\mathrm{Var}(Y) = \mathbb{E}[(\hat{g} - g(\vartheta))^2] I(\vartheta) \\ \text{und} \quad \mathrm{Cov}_\vartheta(\hat{g}, Y) &= \mathbb{E}\big[\big(\hat{g} - \mathbb{E}[\hat{g}]\big)(Y - \mathbb{E}[Y])\big] \\ &= \mathbb{E}[\hat{g}Y] - \cancel{\mathbb{E}[\hat{g}]\mathbb{E}[Y]} \\ &= \int \hat{g}(x) \frac{\partial}{\partial\vartheta} f_\vartheta(x)\, dx = \frac{\partial}{\partial\vartheta} \underbrace{\int \hat{g}(x) f_\vartheta(x)\, dx}_{=g(\vartheta)} = g'(\vartheta) \end{aligned}$$

□

Bemerkung (a) Im Fall von Zähldichten p_ϑ anstatt von Dichten f_ϑ gilt exakt die selbe Aussage (allg. für Radon-Nikodym-Ableitung von P_ϑ bzgl. eines dominierenden Maßes).

(b) Im Fall $g'(\vartheta) = 0$ wird die rechte Seite als 0 interpretiert. Ist $I(\vartheta) = 0$ und $g'(\vartheta) = 0$ folgt $\mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))^2] = \infty$.

(c) Die Bedingungen an f_ϑ sind die, die an ein reguläres stat. Modell gestellt werden. Sie sind für bzgl. ϑ stetig differenzierbares f_ϑ mit $\int \sup_{\vartheta\in\Theta} \left|\frac{\partial}{\partial\vartheta} f_\vartheta(x)\right| dx < \infty$ stets erfüllt.

(d) Gleichheit in der Cramér-Rao-Schranke gilt genau dann, wenn $Y = \frac{\frac{\partial}{\partial\vartheta} f_\vartheta}{f_\vartheta}$ und $\hat{g}$ linear unabhängig sind. In diesem Fall nennt man $\hat{g}$ *effizient* . Die lineare Unabhängigkeit führt auf die Exponentialfamilien als stat. Modell.

Lemma VI.7
Im Produktmodell $(\mathcal{X}^n, \mathcal{F}^{\otimes n}, (P_\vartheta^{\otimes n})_{\vartheta\in\Theta})$ mit P_ϑ wie oben gilt für die Fisher-Information

$$I_n(\vartheta) = n I_1(\vartheta)$$

(n unabhängige und identisch verteilte Beobachtungen $\Longrightarrow$ n-fache Information.)

Beweis: $P_\vartheta^{\otimes n}$ hat die Dichte $\prod_{i=1}^n f_\vartheta(x_i), x \in \mathcal{X}^n$ und daher

$$\begin{aligned}
I_n(\vartheta) &= \mathbb{E}_\vartheta\left[\left(\sum_{i=1}^n \frac{\frac{\partial}{\partial\vartheta} f_\vartheta(X_i)}{f_\vartheta(X_i)}\right)^2\right] \\
&= \operatorname{Var}_\vartheta\left(\sum_{i=1}^n \frac{\frac{\partial}{\partial\vartheta} f_\vartheta(X_i)}{f_\vartheta(X_i)}\right) \overset{\text{unabh.}}{=} \sum_{i=1}^n \operatorname{Var}_\vartheta\left(\frac{\frac{\partial}{\partial\vartheta} f_\vartheta(X_i)}{f_\vartheta(X_i)}\right) \\
&= n I_1(\vartheta)
\end{aligned}$$

□

Insbesondere ist die Rate $\frac{1}{n}$ für den MSE optimal.

Beispiel VI.8 (a) Bernoulli-Experimet $X_1, \dots, X_n \overset{\text{Unabh.}}{\sim} \text{Bin}(1, \vartheta)$, $\vartheta \in (0,1)$.

$$\begin{aligned}
P_\vartheta(x_1, \dots, x_n) &= \prod_{i=1}^n \vartheta^{x_i}(1-\vartheta)^{1-x_i} \qquad x \in \{0,1\}^n \\
\implies I_1(\vartheta) &= \mathbb{E}\left[\left(\frac{X_1\vartheta^{X_1-1}(1-\vartheta)^{1-X_1} - (1-X_1)\vartheta^{X_1}(1-\vartheta)^{-X_1}}{\vartheta^{X_1}(1-\vartheta)^{1-X_1}}\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{X_1}{\vartheta} - \frac{1-X_1}{1-\vartheta}\right)^2\right] \\
&= \underbrace{P(X_1 = 1)}_{=\vartheta}\frac{1}{\vartheta^2} + \frac{1}{(1-\vartheta)^2}\underbrace{p(X_1 = 0)}_{=(1-\vartheta)} \cancel{-2\frac{1}{\vartheta(1-\vartheta)}\mathbb{E}[X_1 - X_1^2]} \\
&= \frac{1}{\vartheta} + \frac{1}{1-\vartheta} = \frac{1}{\vartheta(1-\vartheta)} \implies I_n(\vartheta) = \frac{n}{\vartheta(1-\vartheta)}
\end{aligned}$$

$g(\vartheta) = \vartheta$ und Modell ist regulär. Daher ist die Cramér-Rao-Schranke: $\frac{1}{I(\vartheta)} = \frac{\vartheta(1-\vartheta)}{n}$ also ist $\hat{\vartheta} = \frac{1}{n}\sum_{i=1}^n X_i$ effizient.

(b) Taxiproblem. Vereinfachtes Modell $P_\vartheta = \text{U}_{[0,\vartheta]}^{\otimes n}$ auf $\mathbb{R}^n$. Daher ist

$$f_\vartheta(x) = \prod_{i=1}^n \frac{1}{\vartheta}\mathbf{1}_{[0,\vartheta]}(x_i)$$

ist *nicht* differenzierbar bzgl. ϑ (nicht stetig bei x_1). Also ist das statistische Modell nicht regulär. Wir erhalten die schnellere Konvergenzrate $\frac{1}{n^2}$. Man kann zeigen, dass diese optimal ist.

Idee des Maximum-Likelihood-Prinzips: Ist $(\mathcal{X}, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein diskretes stat. Modell, so wollen wir bei Vorliegen der Stichprobe $x \in \mathcal{X}$ einen Schätzwert $\hat{\vartheta}(x)$ für ϑ bestimmen. Das Maximum-Likelihood-Prinzip ϑ besagt, dass man als $\hat{\vartheta}(x)$ denjenigen Parameter ϑ nehmen soll, für den $p_\vartheta(x)$ maximal ist, d.h. so dass der Versuchsausgang x bei Vorliegen der Verteilung P_ϑ am wahrscheinlichsten ist. Im stetigen Fall benutzt man analog Dichten f_ϑ.

Definition VI.9

(a) Ist $(\mathcal{X}, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ ein diskretes stat. Modell mit Zähldichten $p_\vartheta(\cdot)$, so heißt

$$\begin{aligned} L(\vartheta, x) &:= p_\vartheta(x) \qquad \text{Likelihood-Funktion} \\ \textit{sowie} \quad l(\vartheta, x) &:= \log L(\vartheta, x) \qquad \text{Log-Likelihood} \end{aligned}$$

Entsprechend in einem stat. Modell $(\mathcal{X}, \mathcal{F}, (P_\vartheta)_{\vartheta\in\Theta})$ mit Dichten $L(\vartheta, x) = f_\vartheta(x)$ und $l(\vartheta, x) = \log L(\vartheta, x)$.

(b) Gilt für einen Schätzer $\hat{\vartheta}$

$$L(\hat{\vartheta}(x), x) = \sup_{\vartheta\in\Theta} L(\vartheta, x) \quad \textit{bzw. äquivalent} \quad l(\hat{\vartheta}(x), x) = \sup_{\vartheta\in\Theta} l(\vartheta, x)$$

so heißt $\hat{\vartheta}$ Maximum-Likelihood-Schätzer *(MLE)* .

Lemma VI.10 (Plugin-Prinzip)
Ist $g : \Theta \to \Theta'$ bijektiv, so ist $g(\hat{\vartheta}^{ML})$ der Maximum-Likelihood-Schätzer von $\vartheta' = g(\vartheta)$, wobei $\hat{\vartheta}^{ML}$ der ML-Schätzer von $\vartheta \in \Theta$.

Beweis: $\sup_{\vartheta'\in\Theta'} L'(\vartheta', x) = \sup_{\vartheta'\in\Theta'} f_{g^{-1}(\vartheta')}(x) = \sup_{\vartheta\in\Theta} f_\vartheta(x) = \sup_{\vartheta\in\Theta} L(\vartheta, x)$ □

Beispiel VI.11 (a) Binomialverteilung mit unbekannter Erfolgswahrscheinlichkeit $\vartheta \in (0,1)$.

$$\begin{aligned} \text{Likelihood-Funktion} \quad L(\vartheta, x) &= \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} \qquad x \in \{0, \dots, n\} \\ \text{Log-Likelihood} \quad l(\vartheta, x) &= \log \binom{n}{x} + x \log \vartheta + (n-x) \log(1-\vartheta) \\ \text{Maximum bestimmen:} \quad 0 &\overset{!}{=} \frac{\partial}{\partial \vartheta} l(\hat{\vartheta}, x) = \frac{x}{\hat{\vartheta}} - \frac{n-x}{1-\hat{\vartheta}} \\ \iff 0 &= (1-\hat{\vartheta})x - \hat{\vartheta}(n-x) = x - \hat{\vartheta}n \iff \hat{\vartheta}(x) = \frac{x}{n} \end{aligned}$$

ist ML-Schätzer und nach obigen Beispiel optimal.

(b) Schätzung eines Fischbestandes. $\mathcal{X} = \{0, \dots, n\}$ (n =„Anzahl der Fische beim 2. Fangen“), $\Theta = \{w, w+1, \dots\}$ (w = „Anzahl der markierten Fische“), $x \in \mathcal{X}$ = “Anzahl markierter Fische beim 2. Fang“.

$$\begin{aligned} P_\vartheta &= \mathrm{Hyp}_{n,w,\vartheta-w} \\ L(\vartheta, x) &= \frac{\binom{w}{x}\binom{\vartheta-w}{n-x}}{\binom{\vartheta}{n}} \to \max_\vartheta \end{aligned}$$

Betrachte $\frac{L(\vartheta,x)}{L(\vartheta-1,x)} \implies L(\cdot, x)$ ist monoton wachsend auf $\vartheta \in \{w, \dots, \lfloor \frac{nw}{x} \rfloor\}$ und fallend auf $\vartheta \in \{\lfloor \frac{nw}{x} \rfloor + 1, \dots\} \implies \hat{\vartheta}^{\mathrm{ML}}(x) = \lfloor \frac{nw}{x} \rfloor$.

Bemerkung (a) Das ML-Prinzip liefert in komplexen Modellen gute Ideen, wie Schätzer konstruiert werden können, aber jenseits von asymptotischen Untersuchungen existieren kaum allgemeine Sätze über deren Eigenschaften. Ob der ML-Schätzer eine gute Wahl ist, muss im Einzelfall geprüft werden.

(b) Unter Regularitätsannahmen an das Modell kann man zeigen

$$\sqrt{n}(\hat{\vartheta}^{\mathrm{ML}} - \vartheta) \xrightarrow{d} \mathcal{N}(0, I^{-1}(\vartheta_0))$$

VI.3. Likelihood-Quotienten-Tests

Motivation: Mendels Erbse (1865). Beobachtung bei Erbsen: Ausprägung „rund" (A) oder „kantig" (a) sowie „gelb" (B) oder „grün" (b). Das Merkmal „rund" ist dominant, d.h. die Genotypen AA, Aa, aA führen zum Phänotypen „rund", bei aa ergibt sich „kantig". Ebenso ist „gelb" dominant.

Betrachtet man nun Nachkommen des heterozygoten Genotypen $AaBb$, dann sollten die vier Phänotypen („rund, gelb", „rund, grün", „kantig, gelb", „kantig, grün") im Verhältnis 9:3:3:1 auftauchen.

Mendel erhob folgende Daten ($n = 556$ Erbsen):

	gelb	grün
rund	315	108
kantig	101	32

Dies führt auf die Multinomialverteilung mit $r = 4$ Klassen und (theor.) Wahrscheinlichkeiten $p_1 = \frac{9}{16}, p_2 = \frac{3}{16} = p_3, p_4 = \frac{1}{16}$ für die verschiedenen Klassen. Statistisches Modell

$$\mathcal{X} = \{k \in \{0, 1, \dots, n\}^r \mid k_1 + \dots + k_r = n\}, \quad \mathcal{F} = \mathcal{P}(\mathcal{X})$$

Parametermenge $\Theta = \{\vartheta \in (0,1)^r \mid \vartheta_1 + \vartheta_2 + \dots + \vartheta_r = 1\}$ (entsprechen den $(p_i)_{1 \le i \le r}$).

$$p_\vartheta(k) = P_\vartheta(\{k\}) = \frac{n!}{k_1! \cdots k_r!} \vartheta_1^{k_1} \cdot \vartheta_2^{k_2} \cdots \vartheta_r^{k_r}$$

Teste $H_0 : \vartheta = \vartheta_0$ gegen $H_1 : \vartheta \neq \vartheta_0$ mit $\vartheta_0 = \begin{pmatrix} \frac{9}{16} & \frac{3}{16} & \frac{3}{16} & \frac{1}{16} \end{pmatrix}^\top$.

Weitere Anwendung: Zufallszahlen auf $\{1, \dots, n\}$ und Hypothese $H_0 : \vartheta = \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix}$. Inspiriert von Neyman-Pearson-Test verwendet man für diese Probleme folgende Testkonstruktion:

Definition VI.12
In einem statistischen Modell mit Likelihood-Funktion $L(\vartheta, x)$ betrachte das Testproblem $H_0 : \vartheta \in \Theta_0$ gegen $H_1 : \vartheta \in \Theta_1$ mit $\Theta = \Theta_0 \cup \Theta_1$. Dann heißt ein Test der Form

$$\varphi(x) = \mathbf{1}\left(\frac{\sup_{\vartheta \in \Theta_1} L(\vartheta, x)}{\sup_{\vartheta \in \Theta_0} L(\vartheta, x)} > c_\alpha \right)$$

mit $c_\alpha > 0$ geeignet heißt Likelihood-Quotienten-Test. *Mit $\hat{\vartheta}_0, \hat{\vartheta}_1$, Maximum-Likelihood-Schätzern in der Parametermenge Θ_0 bzw. Θ_1, gilt entsprechend*

$$\varphi(x) = \mathbf{1}\left(\frac{L(\hat{\vartheta}_1(x), x)}{L(\hat{\vartheta}_0, x)} > c_\alpha \right)$$

In unserem Multinomialmodell ergibt sich

$$\frac{\sup_{\vartheta\in\Theta_1} L(\vartheta,x)}{\sup_{\vartheta\in\Theta_0} L(\vartheta,x)} \underset{\text{in } \vartheta}{\overset{\text{Stetigkeit}}{=}} \frac{\sup_{\vartheta\in\Theta} L(\vartheta,x)}{L(\vartheta_0,x)} = \frac{\sup_{\vartheta\in\Theta} p_\vartheta(x)}{p_{\vartheta_0}(x)} = \frac{\sup_{\vartheta\in\Theta} \vartheta_1^{x_1}\cdots\vartheta_r^{x_r}}{\vartheta_{0,1}^{x_1}\cdots\vartheta_{0,r}^{x_r}}$$

$$\underset{\text{Multiplikatoren}}{\overset{\text{Lagrange-}}{=}} \frac{(\frac{x_1}{n})^{x_1}\cdots(\frac{x_r}{n})^{x_r}}{\vartheta_{0,1}^{x_1}\cdots\vartheta_{0,1}^{x_r}}$$

Wir erhalten weiter

$$\log\left(\frac{\sup_{\vartheta\in\Theta_1} L(\vartheta,x)}{\sup_{\vartheta\in\Theta_0} L(\vartheta x)}\right) = \sum_{i=1}^r x_i \log\left(\frac{x_i}{n\vartheta_{0,i}}\right)$$

und der Likelihood-Quotienten-Test hat die Form

$$\varphi(x) = \mathbf{1}\left(\sum_{i=1}^r x_i \log\left(\frac{x_i}{n\vartheta_{0,i}}\right) > \log c_\alpha\right)$$

Taylor-Entwicklung liefert $x\log(\frac{x}{x_0}) = (x-x_0) + \frac{(x-x_0)^2}{2x_0} + o((x-x_0)^2)$ und somit

$$\varphi(x) \approx \mathbf{1}(V^2(x) > v_\alpha) \quad \text{mit} \quad V^2(x) = 2n\sum_{i=1}^r \left(\left(\frac{x_i}{n} - \vartheta_{0,i}\right) + \frac{\left(\frac{x_i}{n}-\vartheta_{0,i}\right)^2}{2\vartheta_{0,i}}\right)$$

$$\underset{\sum \vartheta_{0,i}=1}{\overset{\sum \frac{x_i}{n}=1}{=}} \sum_{i=1}^r \frac{(x_i - n\vartheta_{0,i})^2}{n\vartheta_{0,i}} = \sum_i \frac{(B_i-E_i)^2}{E_i}$$

symbolisch mit B_i: beobachtete Anzahl in Klasse i und E_i erwartete Anzahl in Klasse i.

$$\tilde{\varphi}(x) = \mathbf{1}(V^2(x) > v_\alpha) \quad \text{heißt } \chi^2\text{-Test}$$

Zur Bestimmung des kritischen Wertes v_α für ein vorgegebenes Niveau α bedient man sich der asymptotischen Näherung

$$\forall v > 0: \quad \lim_{n\to\infty} P_{\vartheta_0}(V^2 \le v) = \int_0^v f_{\chi^2(r-1)}(x)\,dx \quad \text{mit} \quad f_{\chi^2(r-1)}(x) = \frac{1}{\Gamma(\frac{r+1}{2})2^{\frac{r-1}{2}}} x^{\frac{r-1}{2}-1} e^{-\frac{x}{2}}$$

(der Dichte der $\chi^2(r-1)$-Verteilung). Das heißt, man wählt v_α als das $(1-\alpha)$-Quantil der $\chi^2(r-1)$-Verteilung.

Bei Mendels Daten ergibt sich eine klare Annahme der Hypothese zum Niveau $\alpha = 0.05$ (oder sogar größer). In der Tat ist die Wahrscheinlichkeit, dass V^2 den Wert entsprechend diesen Daten oder kleiner annimmt, ca. 10%. (Manipulationsverdacht)

Eine weitere Anwendung des Likelihood-Quotienten-Tests ist der sogenannte t-Test für $X_1,\dots,X_n \sim \mathcal{N}(\mu,\sigma^2)$ unabhängig und $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$ ($\sigma > 0$ beliebig, unbekannt).

Index

Literatur

[1] Heinz Bauer. *Wahrscheinlichkeitstheorie*. De Gruyter, 1991.

[2] Kai Lai Chung. *A Course in Probability Theory*. Academic Press, 1994.

[3] Hans-Otto Georgii. *Stochastik*. De Gruyter, 2007.

[4] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2006.

[5] Max Koecher. *Lineare Algebra und analytische Geometrie*. Springer, 1997.

[6] Ulrich Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg, 2005.