Einführung in die Stochastik

24. August 2017

Inhaltsverzeichnis

1 Grundlagen

Erinnerung 1.1. *Naive Grundidee der Modellierung des Zufalls:*

Konzept	*mathematisches Objekt*	*Symbol*
„Alle denkbaren Ergebnisse eines zufälligen Geschehens“	*Menge*	Ω
Wahrscheinlichkeit,dass $\omega \in \Omega$ beobachtet wird	*Abbildung $\Omega \to [0,1]$*	
Alle denkbaren Ja-Nein-Fragen, die zum zufälligen Geschehen gestellt werden können	*Potenzmenge von Ω*	$\mathscr{P}(\Omega)$
Wahrscheinlichkeit, dass die zu $a \in \mathscr{P}(\Omega)$ gehörige Frage mit “ja” beantwortet wird.	*Abbildung $\mathscr{P}(\Omega) \to [0,1]$*	$\mathbb{P}$

Beispiel 1.2.1 (6-Seitiger Würfel). $\Omega = \{1,2,3,4,5,6\}$ und $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$.
Ein Beispiel für eine Ja-Nein-Frage: „Ist die gewürfelte Zahl durch 3 teilbar?“ dann ist $A \in \mathscr{P}(\Omega) : A = \{3,6\}$ und $\mathbb{P}(A) = \frac{1}{3}$.

Beispiel 1.2.2. Speziell zufälllige natürliche Zahl: $\Omega = \mathbb{N}$, $p(1) = \frac{1}{2}, p(2) = \frac{1}{4}, ..., p(n) = 2^{-n}$.
Dann gilt $\sum_{\omega=1}^{\infty} p(\omega) = 1$.

a) Ja-Nein-Frage: „Ist die Zahl gerade?“
Zugelassenen Menge: $A = \{2,3,6,...\}$

$$\mathbb{P}(A) = \sum_{j=1}^{\infty} p(2j) = \sum_{j=1}^{\infty} 2^{-2j} = \frac{1}{1-\frac{1}{4}} - 1 = \frac{1}{3}$$

b) Ja-Nein-Frage: „Ist die Zahl Primzahl?“
Zugelassene Menge: $B = \{n \in \mathbb{N} \mid n \text{ ist prim}\}$

$$\mathbb{P}(B) = \sum_{j \in B} p(j) = ???$$

Abschätzung $\mathbb{P}(B) \leq 1 - \mathbb{P}(A) + \mathbb{P}(\{2\}) \leq \frac{2}{3} + \frac{1}{4}$

Definition 1.3. Sei Ω eine abzählbare Menge.
Eine Abbildung $p : \Omega \to [0,1]$ mit $\sum_{\omega \in \Omega} p(\omega) = 1$ heißt **Zähldichte** oder **Wahrscheinlichkeitsdichte** auf Ω.

Definition 1.4. Man nennt dann Ω den „Ergebnisraum“, die „Grundmenge“ oder „Grundgesamtheit“.
Ein spezielles $A \in \mathscr{P}(\Omega)$ nennt man „Ergebnis“ und falls $A = \{\omega\}$ **Elementarereignis**.

Definition 1.5. Sei Ω eine abzählbare Menge und $\mathscr{P}(\Omega)$ die Potenzmenge, p sei eine Zähldichte.
Dann heißt die Abbildung

$$\mathbb{P} : \mathscr{P}(\Omega) \to [0,1], A \mapsto \sum_{\omega \in A} p(\omega)$$

das von p erzeugte **Wahrscheinlichkeitsmaß** (kurz W-Maß).

Bemerkung 1.6. Beachte: p wird in der Notation unterdrückt. Alternativ schreibe $\mathbb{P}_p$.
Außerdem: Statt $\mathbb{P}(\{\omega\})$ wird oft $\mathbb{P}(\omega)$ geschrieben.

Lemma 1.7. *Sei p eine Zähldichte auf Ω. Das von p erzeugte Wahrscheinlichkeitsmaß hat folgende Eigenschaften:*

a) $\mathbb{P}(\Omega) = 1$

b) Falls $(A_n)_{n\in\mathbb{N}}$ eine Folge von paarweise disjunkten Ereignissen ist, dann ist $\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

Beweis. Sei p Zähldichte auf Ω.

a) Nach Definition **??**: $\mathbb{P}(\Omega) = \sum_{\omega\in\Omega} p(\omega) = 1$

b) Es gilt:

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \mathbb{P}\big(\{\omega \in \Omega : \exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}\big) = \sum_{\{\omega\in\Omega:\exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}} p(\omega) \tag{1}$$

Sei nun $N(\omega) = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) =$„Anzahl der A_n die ω enthalten“. Dann ist

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \sum_{\omega\in A_n} p(\omega) = \sum_{\{\omega\in\Omega:\exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}} p(\omega)N(\omega) \tag{2}$$

Da aber die A_n paarweise disjunkt sind ist $N(\omega) = 1$ für alle $\omega \in \{\omega \in \Omega : \exists n \in \mathbb{N} \text{ mit } \omega \in A_n\}$, ist ist (1)=(2). $\square$

Definition 1.8. Für $A \subseteq \Omega$ heißt

$$\mathbb{1}_A : \Omega \to \{0, 1\}, \quad \omega \mapsto \begin{cases} 1 & \omega \in A \\ 0 & \text{sonst} \end{cases}$$

die **Charakteristische Funktion** oder **Indikator** von A. Man schreibt auch $\mathbb{1}_A$.

Beispiel 1.9. Sei $\Omega = \mathbb{N}$, $(a_n)_{n\in N}$ eine Folge mit $a_n \geq 0$, $a := \sum_{n=1}^{\infty} a_n \leq \infty$ und $a > 0$.
Dann ist $p : \Omega \to [0, 1]$, $n \mapsto p(n) := \frac{a_n}{a}$ eine Zähldichte.
Es gilt sogar die Isomorphie:

$$\{\text{Zähldichte auf } \mathbb{N}\} \cong \{\text{Nicht negative Folgen mit } \sum_{n=1}^{\infty} a_n = 1\} \cong \{\text{Nichtnegative summierbare Folgen} \neq 0\}$$

Bemerkung 1.10 (Notation). Sei Ω eine Menge .$|\Omega| \leq \infty$ bezeichne tdie Anzahl der Elemente von Ω „Mächtigkeit der Menge“.

Beispiel 1.11.1 (Einfache Irrfahrt). Dimension d, N Schritte. $\Omega = \{(x_0, x_1, ..., x_N) : X_j \in \mathbb{Z}^d \forall j, x_0 = 0, |x_{j+1} - x_j| = 1 \forall j\}$.
Also ist $|\Omega_N| = (2d)^N$, Setze $p(\omega = \frac{1}{(2d)^N}) \forall \omega \in \Omega$.

Man kann nu folgende Fragestellungen formulieren:

a) $A_N := \{(x_1, ..., x_N)\} \in \Omega_N : \exists j > 0$ mit $x_j = 0\}$. ("Rückkehr zum Startpunkt").
Es ist klar, dass $\mathbb{P}(A_N) \geq \frac{1}{2d} > 0$, falls $N \geq 2$.
Es ist leicht zu zeigen, dass $N \mapsto \mathbb{P}(A_N)$ wächst monoton.

Knifflig: Was ist $\lim\limits_{N\to\infty} ? < 1? = 1?$

Antwort: $= 1$ für $d \leq 2$, < 1 für $d \geq 3$.

b) $B_n, \alpha := \{\omega = (x_0, x_1, ..., x_N) \in \Omega_N : |x_N| \geq N^\alpha\}$ für $0 < \alpha \leq 1$

Frage: $\lim\limits_{N\to\infty} \mathbb{P}(B_{n,\alpha})$?

Antwort: 0, falls $\alpha > \frac{1}{2}$
1, falls $\alpha < \frac{1}{2}$
Für $\alpha = \frac{1}{2}$ gilt

$$\lim_{N\to\infty} \mathbb{P}(B_{n,\alpha}) = \frac{V_k(d)}{(2\pi)^{\frac{d}{2}}} \int_1^\infty r^{d-1} \exp(\frac{1}{2}r^2)dr$$

(dabei ist $V_k(d)$ das Volumen der d-Dimensionalen Einheitskugel).

Beispiel 1.11.2 (Selbstvermeidende Irrfahrt). Dimension d, N Schritte.

a) $\Omega_N^0 = \{(x_0, x_1, ..., x_N) \in \Omega_N : x_i \neq x_j$ falls $i \neq j\}$ Dann gilt für die Anzahl der Pfade:

$$|\Omega_N^0| = \begin{cases} 2, & \text{falls } d = 1 \\ ??, & \text{falls } d > 1 \end{cases}$$

und es ist $p(\omega) = \frac{1}{|\Omega_N^0| \forall \omega \in \Omega_N^0}$.

b) Wie in a)2.

Frage Was ist $\lim\limits_{N\to\infty} \mathbb{P}(B_{N,\alpha}^0)$.

Bekannt $\exists \alpha_c > 0$ mit

$$\lim_{N\to\infty} \mathbb{P}(\mathbb{B}^{\not\vdash}_{\ltimes,\alpha}) = \begin{cases} 0, & \text{falls } \alpha > \alpha_c \\ 1, & \text{falls } \alpha < \alpha_c \end{cases}$$

Bekannte Werte: $d = 1$ $\alpha_c = 1$
$d = 2$ $\alpha_c = \frac{3}{4}$, falls SLE-Conjecture stimmt
$d = 3$ $\alpha_c \approx 0,5876$ (Numerik)
$d \geq 4$ $\alpha_c = \frac{1}{2}$

Beispiel 1.12. Auswählen einer Zufälligen reellen Zahl in $[0, 1]$, alle Zahlen sollen di gleich Wahrscheinlichkeit haben:
$[0, 1]$ ist nicht endlich, also ist Gleiche Wahrscheinlichkeit für alle Zahlen unmöglich.
$[0, 1)$ ist nicht abzählbar, also scheitert der bisherige Ansatz mit der Zähldichte.

Ein möglicher Ausweg Definiere $\mathbb{P}([a,b]) = \mathbb{P}((a,b)) = \mathbb{P}([a,b)) = \mathbb{P}((a,b])$.
Die Erweiterung, sodass $\forall A \in \mathscr{P}([0,1])$ $\mathbb{P}(A)$ definiert ist, ist nicht möglich.

Lösung Definiere $\mathbb{P}$ nicht auf allen Mengen $\mathscr{P}([0,1])$.

Definition 1.13. Sei Ω eine nichtleere Menge.
Ein Mengensystem $\mathscr{F} \subset \mathscr{P}(\Omega)$heißt **$\sigma$-Algebra**, falls

a) $\Omega \in \mathscr{F}$

b) Falls $A \in \mathscr{F}$, dann auch $A^C \in \mathscr{F}$.

c) Falls $A_1, A_2, ... \in \mathscr{F}$, dann auch $\bigcap A_i \in \mathscr{F}$.

$(\Omega, \mathscr{F})$ heißt dann **messbarer Raum** oder **Ereignisraum**.

Bemerkung 1.14. $\mathscr{F}$ ist "die Menge aller Teilmengen von Ω, für die die zugehörige Ja-Nein-Frage beantwortbar ist".
Daher meint

a) "Ist $\omega \in \Omega$" muss beantwortbar sein.

b) Falls "Ist $\omega \in A$?" beantwortbar, so ist auch "Ist $\omega \notin A$?" beantwortbar.

c) Falls "Ist $\omega \in A_i$?" beantwortbar für alle i, dann ist auch "Ist ω in irgendeinem A_i?" beantwortbar.

Beispiel 1.15. Sei $\Omega = [0,1)$, dann ist

a) $\mathscr{F}_0 := \{\emptyset, \Omega\}$

b) $\mathscr{F}_1 := \{\emptyset, [0, \frac{1}{3}), [\frac{1}{3}, 1), \Omega\}$.
Die Frage "Ist $\omega \geq \frac{1}{2}$" ist hier <u>nicht</u> beantwortbar!

c) $A_{j,n} := \left[\frac{j}{n}, \frac{j+1}{n}\right)$, n ist fest, $j \geq n$.
$\mathscr{F}_2 = \{\bigcup_{k=1}^{n} B_{k,n} : B_{k,n} \in \{\emptyset, A_{k,n}\}\}$

d) $\mathscr{F}_3 = \mathscr{P}(\Omega)$ ist ebenfalls eine σ-Algebra.

Satz 1.16. *Sei $\mathscr{G} \subset \mathscr{P}(\Omega)$ ein Mengensystem. Sei $\Sigma := \{\mathscr{A} \subset \mathscr{P}(\Omega) :$ $\mathscr{A}$ ist σ-Algebra und $\mathscr{G} \subset \mathscr{A}$.*
Dann ist auch $\bigcap_{\mathscr{A} \in \Sigma} \mathscr{A}$ eine σ-Algebra.

Definition 1.16. $\sigma(\mathscr{G}) := \bigcap_{\mathscr{A} \in \Sigma} \mathscr{A}$ heißt **die von $\mathscr{G}$ erzeugt σ-Algebra**.

Definition 1.17. Sei $\Omega = \mathbb{R}$, $\mathscr{G} := \{[a,b] : a, b \in \mathbb{R}, a < b\}$.
$\mathscr{B} := \sigma(\mathscr{G})$ heiß **Borel-σ-Algebra**.

Bemerkung 1.18. a) $\mathscr{B}$ enthält alle offenen Mengen, alle abbgeschlossen Mengen und alle halboffenen Intervalle.

b) $\mathscr{B} \subsetneqq \mathscr{P}(\Omega)$.

c) $\mathscr{B}$kann nicht abzählbar konstruiert werden.

d) $\mathscr{B} = \sigma(\{(-\infty, c]\} : c \in \mathbb{R})$.

e) Falls $\Omega_o 0 \subset \mathbb{R}$, $\Omega_0 \neq \emptyset$, dann ist

$$\mathscr{B}_{\Omega_0} := \{A \cap \Omega_0 : A \in \mathscr{B}(\mathbb{R})\}$$

eine σ-Algebra, die **Einschränkung** von $\mathscr{B}$ auf Ω_0.

Definition 1.19. Seien $E_1, E_2, ..., E_N$ Mengen, $N \leq \infty$.
$\mathscr{E}_i$ seien σ-Algebren auf E_i und es sei

$$\Omega = \mathop{\mathrm{X}}_{i=1}^{N} E_i = \{(e_1, ..., e_N) : e_i \in E_i \forall i \leq N\}$$

Eine Menge der Form

$$A_{j,B_j} = \{(e_1, ..., e_N) : e_j \in B_j, \text{ andere } e_k \text{ beliebig}\}$$

mit $B_j \in \mathscr{E}_j, j \leq N$ heißt **Zylindermenge**.

Definition 1.19. Die σ-Algebra in Ω die von allen Zylindermengen Erzeugt wird heißt **Produkt-σ-Algebra**. Man nennt $\mathscr{Z}$ das System der Zylindermenge und $\mathscr{E}_1 \otimes \mathscr{E}_2 \otimes ... \otimes \mathscr{E}_N := \sigma(\mathscr{Z})$.

Definition 1.20. Sei $\Omega, \mathscr{F}$ ein messabere Raum. Eine Abbildung $\mathbb{P} : \mathscr{F} \to [0,1]$ heiß **Wahrscheinlichkeitsmaß** (W-Maß) auf $\mathscr{F}$ (teilweise auch „auf Ω“), falls

a) $\mathbb{P}(\Omega) = 1$

b) Für paarweise disjunkte $A_n \in \mathscr{F}$ gilt $\mathbb{P}(\bigcap_{i=1}^{\infty}) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ („σ-Additivität“).

Dann heißt $\Omega, \mathscr{F}, \mathbb{P}$ **Wahrscheinlichkeitsraum**.

Beispiel 1.21. Sei $(\Omega, \mathscr{F}, \P) = (\mathbb{R}, \mathscr{B}(\mathbb{R}), \delta_x)$. Dabei ist $\mathbb{P}(A) = \begin{cases} 1; & \text{falls } x \in A \\ 0, & \text{sonst.} \end{cases}$
ein Wahrscheinlichkeitsmaß auf $\mathscr{B}(\mathbb{R})$.
Es Modelliert ein „Zufalls“-Experiment, welches sicher x ergibt.

Satz 1.22 (Eigenschaften von Wahrscheinlichkeitsmaßen)**.** *Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B, A_1, A_2, ... \in \mathscr{F}$. Dann ist*

a) $\mathbb{P}(\emptyset) = 0$

b) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
insbesondere $\mathbb{P}(A) + \mathbb{P}(A^C) = 1$

c) Falls $A \subseteq B$, dann ist $\mathbb{P}(A) \leq \mathbb{P}(B)$ („Monotonie“)

d) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ *(σ-Subadditivität)*

e) Falls $A_n \nearrow A_n$ (d.h. $A_1 \subseteq A_2 \subseteq ...$ und $\bigcap_{n=1}^{\infty} A_n = A$), dann ist $\lim\limits_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$ („σ-Stetigkeit“)
Falls $A_n \searrow A_n$ (d.h. $A_1 \supseteq A_2 \supseteq ...$ und $\bigcup_{n=1}^{\infty} A_n = A$), dann ist $\lim\limits_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$

Beweis. a) $\mathbb{P}(\emptyset) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \emptyset\right) = \sum_{n=1}^{\infty} \mathbb{P}(\infty)$. Also $\mathbb{P}(\emptyset) = 0$

b) Falls $A \cap B = \emptyset$, dann ist $\Phi(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(\emptyset) = \mathbb{P}(A) + \mathbb{P}(B)$.
Sei nun $A \cap B \neq \emptyset$.

$\square$

Definition 1.27. Die Abbildung $\lambda : \mathscr{B}(\mathbb{R}^n) \to [0, \infty], A \mapsto \int_{\mathbb{R}^n} \mathbb{1}_A(x)\ dx = \lambda(A)$ heißt **Lebesgue-Maß**

Beispiel 1.28. Sei $\varrho : \mathbb{R}^n \to [0, \infty)$ Borel-messbar und $\int \varrho(x)\ dx = 1$.
Dann ist die Abbildung $\mathbb{P}_p\varrho : \mathscr{B}(\mathbb{R}^n) \to [0, 1], A \mapsto \int_A \varrho(x)\ dx = \int \mathbb{1}_A(y)\varrho(x)\ dx$ ein Wahrscheinlichkeitsmaß.

Definition 1.29. Sei $\varrho : \mathbb{R}^n \to [0, \infty)$ Borel-messbar und $\int \varrho(x)\ dx = 1$, (=1.28) dann heißt ϱ **Dichte** von $\mathbb{P}_p\varrho$.

Beispiel 1.30. Sei ϱ eine Dichte, $x \in \mathbb{R}$. Dann hat $\mathbb{P} := \frac{1}{3}\mathbb{P}_p\varrho + \frac{2}{3}\delta_x$ keine Dichte (siehe **??**)

Bemerkung. Wenn ϱ Dichte ist schreibt man auch $\mathbb{P}_p\varrho \equiv \varrho(x)\ dx$

Definition 1.31. Sei $\Omega \in \mathscr{B}(\mathbb{R}^n)$ mit $\lambda(\Omega) < \infty$. Das Wahrscheinlichkeitsmaß auf Ω mit Dichte $\varrho(y) = \frac{1}{\lambda(\Omega)}$ heißt **Gleichverteilung** auf Ω.
Man fasst dann $\tilde{\varrho}(x) = \frac{1}{\lambda(\Omega)}\mathbb{1}_A\Omega(x)$ als Einbettung in den $\mathbb{R}^n$ auf.

Erinnerung (Zufallsvariable)**.** *Der Begriff „Zufallsvariable" ist historisch gewachsen. (Keine Variable einer Funktion).*

Problemstellung *Von einem komplizierten Zufälligen Geschehen will man nur gewisse Aspekte betrachten.*

Beispiel 1.32.1 (2 mal Würfeln, Würfelsumme). Sei $\Omega : \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 4, 5, 6\}$.
$\omega = (\omega_1, \omega_2)$. Der zu betrachtende Aspekt: $S(\omega_1, \omega_2) = \omega_1 + \omega_2 \in \{2, 3, ..., 12\} \neq \Omega$.

Beispiel 1.32.2. Sei $\Omega = \Omega_N$(siehe **??**, einfache Irrfahrt).
$\omega = (\omega_1, \omega_2, ...\omega_N) \in (\mathbb{Z}^d)^N$.

Aspekt 1 Position nach N Schritten.
Modell: $X_N\big((\omega_1, ..., \omega_N)\big) = \omega_N \in \mathbb{Z}^d \neq \Omega$

Aspekt 2 Maximaler Abstand vom Ursprung bis zum Schritt N.
Modell $M_N\big((\omega_1, ..., \omega_N)\big) = \max\{|\omega_j|, j \leq N\}$.

Beispiel 1.32.3. Sei $\Omega = [0, 1]$, $\mathscr{F} \in \mathscr{B}([0, 1])$, $\omega = x \in [0, 1]$.

Aspekt 1 Erste Ziffer nach dem Komma?
Modell: $y_1(x) = \lfloor 10x \rfloor$

Aspekt 2 Fläche des Quadrates mit Kantenlänge x
Modell: $Q(x) = x^2 \in [0, 1]$.

Fazit: Modellierung durch Abbildungen.

Definition 1.33. Seine $(\Omega, \mathscr{F})$ und $(\Omega', \mathscr{F}')$ Ereignisräume.
eine Abbildung $X : \Omega \to \Omega'$ heißt **Zufallsvariable** (ZV) [oder messbare Abbildung, zufälliges Element von Ω'], falls gilt:
$\forall A' \in \mathscr{F}'$ ist $X^{-1}(A') \in \mathscr{F}$.
Hierbei ist X^{-1} das Urbild von A' unter X.

Bemerkung 1.34. Die Urbild-Abbildung bilte Mengen in $\mathscr{F}'$ (d.h.erlaube Ja-Nein-Fragen) auf Mengen in $\mathscr{P}(\Omega)$ (d.h. Ja-Nein-Fragen) ab.

Beispiel. In 1.32.1 **??** ist $S^{-1}(\{4\}) = \{(1,3),(2,2),(3,1)\}$.
In 1.32.1 **??** ist $Y_1^{-1}(\{3,7\}) = [0,3,0,4) \cup [0,7,0,8)$

Bemerkung. Die Bedingung (*) bedeutet, dass für alle durch $A \in \mathscr{F}'$ erzeugte erlaubten Ja-Nein-Fragen auch die Frage „Liegt $X(\omega)$ in A'?" erlaubt ist.

Bemerkung. Oft nimmar man nur Ω und $(\Omega', \mathscr{F}')$ als gegeben.
Dann ist $X^{-1}(\mathscr{F}') := \{X^{-1}(A') \mid A' \in \mathscr{F}'\}$ die von X erzeugt σ-Algebra.

Bemerkung. Falls $\mathscr{F} = \mathscr{P}(\Omega)$, dann ist jede Abbildung eine Zufallsvariable.

Lemma 1.35. *Seine $(\Omega, \mathscr{F})$ und $(\Omega', \mathscr{F}')$ Ereignisräume, $X : \Omega \to \Omega'$ und sei $\mathscr{G}'$ ein Mengensystem mit $\mathscr{F}' = \sigma(G')$. Dann ist X genau dann Zufallsvariable, wenn $X^{-1}(A') \in \mathscr{F} \forall A' \in \mathscr{G}'$.*

Beweis. **„⇒"** ist klar, da $\mathscr{G}' \subset \mathscr{F}'$

„⇐" Sei $\mathscr{A}' := \{A' \in \Omega' \mid X^{-1}(A') \in \mathscr{F}\}$ ist eine σ-Algebra und $\mathscr{A}' \supset \mathscr{G}'$ nach Annahme.
Daher ist $\mathscr{F}' = \sigma(\mathscr{G}') \subset \mathscr{A}'$, sodass $X^{-1}(A') \in \mathscr{F} \forall A \in \mathscr{F}$.

□

Beispiel 1.36.1. Sei $(\Omega', \mathscr{F}') = (\mathbb{R}, \mathbb{R})$. Nach 1.35 gilt:
$X : \Omega \to \Omega'$ ist genau dann Zufallsvariable, wenn $X^{-1}\big((-\infty, c)\big) \in \mathscr{F} \forall c \in \mathbb{R}$
Für $\Omega' = \mathbb{R}$ heißt X **reelle Zufallsvariable**

Beispiel 1.36.2. Es ist $\overline{\mathbb{R}} := [-\infty, \infty]$ mit σ-Algebra $\mathscr{B}(\overline{\mathbb{R}}) = \sigma\big(\{[-\infty - c] : c \in \overline{\mathbb{R}}\}\big)$.
Die Abbildung $X : \Omega \to \overline{\mathbb{R}}$ ist genau dann Zufallsvariable, wenn $X^{-1}\big([-\infty, c]\big) \in \mathscr{F} \forall c$.
Dann heißt X **numerische Zufallsvariable**

Theorem 1.37. *Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathscr{F}')$ ein Ereignisraum, $X : \Omega \to \Omega'$ eine Zufallsvariable.*
Dann ist die Abbildung

$$\mathbb{P}' : \mathscr{F}' \to [0,1], \quad A' \mapsto \mathbb{P}'(A') := \mathbb{P}(X^{-1}('A))$$

ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathscr{F}')$.

Definition 1.38. $\mathbb{P}'$ heißt **Bildmaß von $\mathbb{P}$ unter X** oder **Verteilung von X unter $\mathbb{P}$**.
Man schreibt $\mathbb{P}' = \mathbb{P} \circ X^{-1}$ oder $\mathbb{P}' = \mathbb{P}_X$

Beweis. Da X eine Zufallsvariable ist, ist $X^{-1}(A') \in \mathscr{F} \forall A' \in \mathscr{F}'$, daher im Definitionsbereich von $\mathbb{P}$.
Also ist $\mathbb{P}'$ wohldefiniert. Prüfe Definition 1.20 **??**.

a) $\mathbb{P}'(\Omega') = \mathbb{P}(X^{-1}(\Omega')) = \mathbb{P}(\Omega) = 1$

b) $A'_1, A'_2, ... \in \mathscr{F}'$ seien paarweise disjunkt. Dann sind $X^{-1}(A'_1), X^{-1}(A'_2), ...$ auch paarweise disjunkt.

$$\begin{aligned}\mathbb{P}'\left(\bigcup_{i=1}^{\infty} A'_i\right) &= \mathbb{P}\left(X^{-1}\left(\bigcup_{i=1}^{\infty} A'_i\right)\right) \\ &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} X^{-1}(A'_i)\right) \\ &= \sum_{i=1} \infty \mathbb{P}(X^{-1}(A'_i)) \\ &= \sum_{i=1} \infty \mathbb{P}'(A'_i)\end{aligned}$$

□

Definition 1.39. Seien $(\Omega_1, \mathscr{F}_1, \mathbb{P}_1)$ und $(\Omega_2, \mathscr{F}_2, \mathbb{P}_2)$ Wahrscheinlichkeitsräume, $X_1 : \Omega_1 \to \Omega'_1$, $X_2 : \Omega_2 \to \Omega'_2$ Zufallsvariablen.
Falls $\mathbb{P}_1(X_1^{-1}(A')) = \mathbb{P}_2(X_2^{-1}(A')) \forall A' \in \mathscr{F}'$, dann heißen X_1 und X_2 identisch verteilt.

Bemerkung 1.40 (Notation). Man schreibt oft:

- $\{X \in A'\}$ statt $X^{-1}(A')$
- $\mathbb{P}(\{X \in A'\})$ oder $\mathbb{P}(X \in A')$ statt $\mathbb{P}(X^{-1}(A'))$
- $\mathbb{P}_X$ statt $\mathbb{P} \circ X^{-1}$.

2. Urnenmodelle mit Reihenfolge, mit Zurücklegen

Beispiel 2.5.1 (Einmal zeihen:). Bsp: Tulpenzwiebeln, N Stück: k_r rote, k_g gelbe, k_o orange.
Dann ist $\mathbb{P}(\text{rot}) = \frac{k_r}{N}$,...
Allgemein: Menge der Merkmale $A = (a_1, ..., a_m)$ mit p_i =Bruchteil der „Kugeln“ mit Merkmal a_i, sodass$\mathbb{P}(\{a_i\}) = p_i$.
Dann ist $\mathbb{P}$ eine Zähldichte mit $\sum_{i=1}^m p_i = 1$.

Beispiel 2.5.2 (N-Mal ziehen). Anzahl der Ziehungen $N \in \mathbb{N}, \Omega = A^N, \mathbb{P}\big((a_{j_1}, ..., a_{j_N})\big) := p_{j_1} \cdot ... \cdot p_{j_N}$, für $j_1, ..., j_N \in \{1, ..., m\}$ mit $p_1, ..., p_m$ wie in 1.
Modelliert N-mal „unabhängig“ (siehe Kapitel 3) ziehen mit Zurücklegen.

Definition 2.6. Spezialfall von 2.: $A = \{0, 1\}$, $p_1 = p, p_0 = 1 - p$.
1$\widehat{=}$„Erfolg“, $p_1\widehat{=}$„Erfolgswahrscheinlichkeit“.
Dann kann für $\omega \in A^N$, $\omega = (\omega_1, ..., \omega_N)$ schreiben:

$$\mathbb{P}(\{\omega\}) = p^{\sum_{i=1}^N \omega_i}(1-p)^{\sum_{i=1}^N (1-\omega_i)}$$

Man nennt dies die **Bernoulli-Verteilung** für N Versuche mit Erfolgswahrscheinlichkeit p.

Definition 2.7. Die Zähldichte $\overline{p} : A \times ... \times A \to [0, 1]$, $\overline{p}(a_1, ..., a_n) \mapsto \mathbb{P}(\{a_1\}) \cdot ... \cdot \mathbb{P}(\{a_n\}) = p_1 \cdot ... \cdot p_n$ aus 2.5.2 heißt N-fache **Produktdichte** der Zähldichte $p : A \to [0, 1], a_j \mapsto p(a_j) = p_j$.
Das zugehörige Wahrscheinlichkeitsmaß heißt N-faches **Produktmaß** von ¶.

2.A Urnenmodelle, ohne Reihenfolge, mit Zurücklegen

Beispiel 2.8. In Situation 2. nammt man an dass man die Anzahl der gezogenenen Kugeln pro ??? von Interesse ist. Das ist ein „Aspekt" des Experiments, daher eine Zufallsvariable:

$$X : A^N \mapsto \mathbb{N}_0^m, \quad (a_{j_1}, ..., a_{j_N}) \mapsto \left(\sum_{k=1}^N \mathbb{1}_{\{a_1\}}(a_{j_k}), \sum_{k=1}^N \mathbb{1}_{\{a_2\}}(a_{j_k}), ..., \sum_{k=1}^N \mathbb{1}_{\{a_m\}}(a_{j_k}) \right)$$

(Jede Komponenten entspricht jeweils der Anzahl der gezogenen a_i). Nach der Definition des Produktmaßes ist

$$\begin{aligned}\mathbb{P}\big((n_1, ..., n_m)\big) &= \mathbb{P}\left(X^{-1}\big(\{(n_1, ..., n_m)\}\big)\right) \\ &= p_1^{n_1} p_2^{n_2} ... p_m^{n_m} \cdot \text{„Anzahl der } N\text{-Tupel von Elementen aus } A \text{ mit } n_1 \text{ mal } a_1 \text{ und } n_2 \text{ mal } a_2 \text{ un} \\ &= \prod_{i=1}^m p_i^{n_i} \binom{N}{(n_1, ..., n_m)}\end{aligned}$$

Dabei ist $\binom{N}{(n_1,...,n_m)} = \frac{N!}{n_1! n_2! ... n_m!}$, falls $\sum_{i=1}^m n_i = N$ und $= 0$ sonst.
Man nennt X auch „**Histogramm**".

Definition 2.9. Sei p Zähldichte auf $A = \{a_1, ..., a_m\}$. Dann heißt das Wahrscheinlichkeitsmaß $\mathrm{M}_{n,p}$ auf $(\mathbb{N}^m, \mathscr{P}(\mathbb{N}^m))$ mit

$$\mathrm{M}\left(\{k_1, ..., k_m\}\right) = \begin{cases} 0 & \text{falls } k_1, ..., k_m \neq N \\ \binom{N}{(k_1,...,k_m)} \prod_{i=1}^m p(a_i)^{k_i} & \text{sonst} \end{cases}$$

Multinormalverteilung für N Stichproben mit Ereigniswahrscheinlichkeiten $p(a_1), ..., p(a_n)$ aus der Menge $\{a_1, ..., a_m\}$.

Definition 2.9. Falls in 2.9 gilt, dass $A = \{0, 1\}$ und $p(1) = p$, $p(0) = 1 - p$. In diesem Fall heißt das Wahrscheinlichkeitsmaß $\mathrm{B}_{N,p}$ auf $\{0, 1, ..., N\}$ (oder auf $\mathbb{N}_0$) mit

$$B_{N,p}(\{k\}) = \binom{N}{k} p^k (1-p)^{N-k}$$

Binomialverteilung für N Vrsuche mit Erfolgswahrscheinlichkeit p.

Beispiel 2.10 (Einfache Irrfahrt). Sei $d = 1$, es wird N Schritte gegangen. Nach rechts mit Wahrscheinlichkeit p, nach links mit Wahrscheinlichkeit $1 - p$.
Dann ist X_n der Ort der Irrfahrt nach n Schritten.
Die Sitation lässt sich Modellieren mit $\Omega = \{0, 1\}^N$,

$$\mathbb{P}\big((\omega_1, ..., \omega_n)\big) = p^{\sum_{i=1}^N \omega_i} (1-p)^{\sum_{i=1}^N (1-\omega_i)}$$

Mit Bernoulli-Verteilung:

$$X_n(\omega) = 2 \sum_{i=1}^N \omega_i - N$$

Also ist $\frac{1}{2}(X_n + N)$ eine $B_{N,p}$ verteilte Zufallsvariable. Dann ist

$$\mathbb{P}(X_N = 2k - N) = \mathbb{P}\big(\frac{1}{2}(X_n + N) = k\big) = \binom{N}{k} p^k (1-p)^{N-k}$$

falls $0 \leq l \leq N$.
Im Speziell für $p = 1/2$ gilt

$$\mathbb{P}(X_n = 2k - N) = \binom{N}{k} 2^{-N}$$

Eine Gute Veranschaulichung ist das sogenannte „Galton-Brett“,

2.B Urnenmodelle ohne Zurücklegen

Beispiel 2.11. Hierfür sei auf den Georgii verwiesen.

Definition 2.12. Sei $\Omega_0 = \{1, ..., m\}$ die Menge der „Merkmale“ $k_1, ..., k_m \in \mathbb{N}$ mit $\sum_{j=1}^m k_j = M \in \mathbb{N}$.
Modelliert man nun eine Urne mit M Kugeln, davon k_j mit Merkmal j, dann heißt für $N \in \mathbb{N}_0, N \leq M$ das Wahrscheinlichkeitsmaß auf

$$\overline{\Omega} = \left\{ (n_1, ..., n_m), n_i \leq k_i \forall i, \sum_{i=1}^m n_i = N \right\} \subseteq \mathbb{N}_0^m$$

mit der Zähldichte

$$\mathrm{H}_{N,(k_1,...,k_m)} \left((n_1, ..., n_m) \right) = \frac{1}{\binom{M}{N}} \prod_{i=1}^m \binom{k_i}{n_i}$$

Hypergeometrische Verteilung zu N und $(k_1, ..., k_m)$.
Modelliert man N-mal Ziehen ohne zurück legen mit k_j Kugeln der Farbe j, dann bedeutet das Ereignis $\{n_1, ..., n_m\}$, dass n_i Kugeln der Farbe k_i gezogen wurden.
Falls $\Omega_0 = \{0, 1\}$, dan ist $\overline{\Omega} = \{0, 1, ..., n\}$ und

$$\mathrm{H}_{N,k_1,k_0} \left(\{n\} \right) = \frac{\binom{k_1}{n}\binom{k_0}{N-n}}{\binom{k_1+k_0}{N}}$$

Dabei ist N die Anzahl der Ziehungen, k_1 die Anzahl der „Gewinne“, k_0 die Anzahl der „Nieten“ im Topf und n die Anzahl der gezogenen „Gewinne“.

Beispiel 2.13 (Lotto (6 aus 49)). Sei $k_1 = 6$ die „angekreuzten Zahlen“, $k_0 = 43$. Dann gilt bei 6-maligem Ziehen:

$$\mathbb{P}(\text{„4 richtige“}) = \mathrm{H}_{6,6,49}(4) = \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} \approx 9.686 \cdot 10^{-4}$$

Definition 2.14. Sei $\lambda > 0$. Das Wahrscheinlichkeitsmaß auf $\mathbb{N}_0$ mit Zähldichte $p_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ heißt **Poisson-Verteilung** zum Parameter λ. Schreibe $\mathrm{Poi}(\{k\}) := \mathbb{P}_{p_k}(\{k\})$.

Bemerkung (Bedeutung der Poisson-Verteilung). Modelliert die Anzahl der Erfolge, wenn N =Anzahl der Versuche $\to \infty$, $\mathbb{P}(\text{Erfolg pro Versuch}) \xrightarrow{N\to\infty} 0$, sodass $N\mathbb{P}(\text{Erfolg}) \xrightarrow{N\to\infty} \lambda$.

Definition 2.15. Seien $(a_n), (b_n) \subseteq \mathbb{R}^+$ Folgen, dann heißen $(a_n), (b_n)$ zueinander **asymptotisch äquivalent**, falls $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Lemma 2.16. *Sei $\sim$ die Asymptotische Äquivalenz. Dann gilt*

a) $\sim$ ist Äquivalenzrelation (reflexiv, symmetrisch, transitiv)

b) Falls $\lim_{n\to\infty} a_n = a \leq \infty$ und $(a_n) \sim (b_n)$, dann ist $\lim_{n\to\infty} b_n = a$.

Satz 2.14 (Poisson-Approximation)**.** *Sei $\lambda > 0$ und $(p_n) \subset [0,1]$ eine Folge, mit $\lim_{n\to\infty} p_n = \lambda$. Sei B_{n,p_n} Binomialverteilung, dann gilt $\forall k \in \mathbb{N}$:*

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

Beweis. Es gilt für den Binomialkoeffizienten:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1}{k!} n(n-1)...(n-k+1) \sim \frac{1}{k!} n^k$$

denn für festes $k \in \mathbb{N}$ ist $(n-k) \sim n$ und $\sim$ ist transitiv. Es folgt, dass

$$\begin{aligned} B_{n,p_n}(\{k\}) = \binom{n}{k} p_n^k (a - p_n)^{n-k} &\sim \frac{1}{k!} n^k p_n^k (1-p_n)^{n-k} \\ &\sim \frac{\lambda^k}{k!} \left(1 - n\frac{1}{n} p_n n\right)^n \underbrace{(1-p_n)^{-k}}_{\sim 1} \\ &\sim \frac{\lambda^k}{k!} \big(1 - \frac{1}{n}(p_n n)\big)^n \xrightarrow{n\to\infty} \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

denn falls $\lim_{n\to\infty} a_n = a > 0$, dann $\lim_{n\to\infty} \big(1 - \frac{1}{n} a_n\big)^n e^{-a}$. ☐

Bemerkung 2.17. Sei α die „Rate des Eintreffens“ eiens Ereignisses E, d.h.

$$\lim_{\delta\to 0} \mathbb{P}(E \text{ tritt in } [0,\delta] \text{ ein}) \frac{1}{\delta} = \alpha$$

Sei $[0,t]$ der Beobachtungszeitraum. Teile $[0,t]$ in n Intervalle der Länge t/n. Zusätzlich nehmen wir an

(i) Ob E in einem gewissen Teilintervall auftritt beeinflusst die anderen Teilintervalle nicht.

(ii) Zwei Ereignisse in einem Teilintervall sind verschwindend unwahrscheinlich.

Dann ist

$$\mathbb{P}(E \text{ passiert } k\text{-mal in } [0,t]) \overset{(ii)}{\approx} \mathbb{P}(E \text{ passiert in } k \text{ Intervallen}) \overset{(i)}{\approx} B_{n,p_n}(k) \xrightarrow{n\to\infty} \mathrm{Poi}_{\alpha t}(k)$$

Also beschreibt der Parameter λ der Verteilung Poi_λ beschreibt „Rate mal Zeit“.

Definition 2.18. Für $0 < p \leq 1$ heißt das Wahrscheinlichkeitsmaß $\mathscr{G}_p$ auf $\mathbb{N}_0$ mit Zähldichte $g(k) = p(1-p)^k$ die **geometrische Verteilung** zur Erfolgswahrscheinlichkeit p.
$\mathscr{G}(\{k\}) =$„Wahrscheinlichkeit bei Erfolgswahrscheinlichkeit p pro Versuch genau k Fehlversuche vor dem Ersten Erfolg zu sehen".
Allgemein: Da Wahrscheinlichkeitsmaß $\overline{\mathscr{B}}_{r,p}$ auf $\mathbb{N}_0$ mir $r \in \mathbb{N}$ und $0 < p \leq 1$ mit

$$\overline{\mathscr{B}}(\{k\}) = \frac{r(r+1)...(r+k-1)}{k!} p^r (1-p)^k$$

heißt **negative Binomialverteilung** und modelliert die Wahrscheinlichkeit vor dem $r-$ten Erfolg genau k Misserfolge zu haben.

Definition 2.19. Das Wahrscheinlichkeitsmaß Exp_α auf $(\mathbb{R}_0^+, \mathscr{B}(\mathbb{R}_0^+))$ mit Dichte $\rho_\alpha(x) := \alpha e^{-\alpha x}$ heißst **Exponentialverteilung** zum Parameter $\alpha > 0$

Definition 2.19. Das Wahrscheinlichkeitsmaß Exp_α auf $(\mathbb{R}_0^+, \mathscr{B}(\mathbb{R}_0^+))$ mit $\alpha > 0, r \geq 1$ und Dichte

$$\gamma_{\alpha,r} := \frac{\alpha^r}{(r-1)!} x^{r-1} e^{-\alpha x} = \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x}$$

heißt **Gammaverteilung** zu den Parametern α, r.

Bemerkung 2.20. a) $\int \gamma_{\alpha,r}(x) = 1$

b)

$$\begin{aligned}
\mathrm{Exp}_\alpha([0,t]) = \alpha \int_0^t e^{-\alpha x}\, dx = \alpha \left[-\frac{1}{\alpha} e^{-\alpha x}\right]_0^t = 1 - e^{-\alpha t} = e^{-\alpha t} \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} t^k \\
= \mathrm{Poi}_{\alpha,t}(\{\mathbb{N}\}) = \mathrm{Poi}_{\alpha t}(\text{„Midestens ein Erfolg vor Zeit } t \text{ bei Rate } \alpha\text{"}) \\
= \mathbb{P}(\text{„Wartezeit auf ersten Erfolg} \leq t\text{"})
\end{aligned}$$

c) Es gilt

$$\mathbb{P}_{\gamma_{\alpha,r}}([0,t]) = \frac{\alpha^r}{(r-1)!} \int_0^t x^{r-1} e^{-\alpha x} dx = e^{-\alpha t} \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{k!} \qquad (\star)$$

denn

$$\begin{aligned}
\frac{d}{dt}(\star) &= -\alpha e^{-\alpha t} \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{k!} + e^{-\alpha t} \sum_{k=r}^{\infty} \frac{\alpha^k}{k!} k t^{k-1} \\
&= e^{-\alpha t} \left(-\alpha \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{k!} + \alpha \sum_{k=r}^{\infty} \frac{(\alpha t)^k}{(k-1)!} \right) \\
&= e^{-\alpha t} \alpha \frac{(\alpha t)^{r-1}}{(r-1)!} = \frac{\alpha^r}{(r-1)!} \frac{d}{dt} \int_0^t x^{r-1} e^{-\alpha x}\, dx
\end{aligned}$$

Somit ist also

$$(\star) = \mathrm{Poi}_{\alpha,t}(\{r, r+1, ...,\}) = \mathbb{P}(\text{Wartezeit bis zum } r\text{-ten Erfolg ist} \leq t)$$

Daher heißt $\mathrm{Exp}_\alpha\, \alpha$ und $\mathbb{P}_{\gamma_{\alpha,r}}$ **Wartezeitverteilungen**.

Definition 2.21. Das Wahrscheinlichkeitsmaß $\mathscr{N}$ auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit Dichte $\phi_{0,1(x)} = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ heißst **Standard-Normalverteilung**. Das Wahrscheinlichkeitsmaß $\mathscr{N}_{m,v}$ auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit Dichte $\phi_{m,v}(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v}\right)$ heißst **Normalverteilung** (oder Gaußverteilung/Gaußmaß) mit Mittelwert m und Varianz v.

Proposition 2.22. $\phi_{m,v}(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2v}\right)$ ist eine Dichte auf $\mathbb{R}$.

Beweis. Es gilt

$$\begin{aligned}
\frac{1}{\sqrt{2\pi v}} \int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2v}} &\overset{y=x-m}{=} \frac{1}{\sqrt{2\pi v}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2v}}\ dy \\
&\overset{z=\frac{y}{\sqrt{v}}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{z}}\ dz
\end{aligned}$$

Berechne dann stattdessen

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\ dx\right) = \left(\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\ dx\right)\left(\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\ dy\right) &= \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}}\ dx\ dy \\
&= \int_0^{\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r\ d\theta\ dr \\
&= 2\pi \int_0^{\infty} e^{-\frac{r^2}{2}} r\ dr
\end{aligned}$$

mit $s = -\frac{r^2}{2}$

$$\begin{aligned}
&= 2\pi \int_{-\infty}^{0} e^s\ ds \\
&= 2\pi(e^0 - e^{-\infty}) = 2\pi
\end{aligned}$$

Es folgt, dass $\int_{-\infty}^{\infty} e^{-\frac{(x-m)^2}{2v}} = \sqrt{2\pi}$. □

Bemerkung 2.23.

a) Bedeutung der Normalverteilung:
„universeller" Grenzwert unabhängiger Summen in der „einzig sinnvollen" Skalierung: $\frac{1}{n}\sum_{i=1}^{\infty} X_i \to m$.
Es folgt, dass $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - m) \sim \mathscr{N}(0, v)$.

b) Geometrische Bedeutung: $\mathscr{N}_{0,v}$ ist die erste Koordinate der sogenannte „Gleichverteilung auf $\mathbb{R}^{\infty}$".

Satz 2.24. *Sei $B_N(r) := \left\{x \in \mathbb{R}^N : |x| \leq r\right\}$ Kugel mit Radius r im $\mathbb{R}^N$, sei $\mathbb{P}_N$ die Gelichverteilung auf $B_N(r)$ und sei $X_1 : B_N(r) \to \mathbb{R}, x = (x_1, ..., x_N) \mapsto x_1$ die Zufallsvariable „Projektion auf die erste Koordinate". Dann gilt*

$$\lim_{N\to\infty} \mathbb{P}_{N,r_N}(a \leq X_1 \leq b) = \begin{cases} 0, & \textit{falls } \lim\limits_{n\to\infty} \frac{r_N}{\sqrt{N-1}} = \infty \\ \delta_0([a,b]), & \textit{falls } \lim\limits_{N\to\infty} \frac{r_N}{\sqrt{N-1}} = 0 \textit{ und } a, b \neq 0 \\ \frac{1}{\sqrt{2\pi v}} \int_a^b e^{-\frac{x^2}{2v}}\ dx, & \textit{falls } \lim\limits_{N\to\infty} \frac{r_N}{\sqrt{N_1}} = \sqrt{v} > 0 \end{cases}$$

Beweis. (nur Fall $a, b \neq 0$ im Fall 2) Integration einer Kugel durch Zerlegen in „Kreisscheiben“.

$$\begin{aligned}
h_r(a,b) :&= \int_{\mathbb{R}^N} \mathbb{1}_{x_1^2+\ldots+x_N^2 \leq r^2}(x) \mathbb{1}_{\{a \leq x_1 \leq b\}} \ dx \\
&= \int_a^b dz \int_{\mathbb{R}^{N-1}} \mathbb{1}_{x_1^2+\ldots+x_N^2 \leq r^2 - z^2}(x) \ dx \\
&= \int_a^b \int_{B_{N-1}(\sqrt{r^2-z^2})} \ dx = \int_a^b (r^2-z^2)^{N-1/2} \ dz \underbrace{\int_{B_{N-1}} (1) \ dx}_{=V_1(N-1)} \\
&= r^{N-1} V_1(N_1) \int_a^b \left(1 - \frac{z^2}{r^2}\right)^{\frac{N-1}{2}} \ dz
\end{aligned}$$

Durch Subtitution $y = \frac{z}{r}\sqrt{N-1}$

$$= r^{N-1} V_1(N-1) \frac{r}{\sqrt{N-1}} \int_{\frac{\sqrt{N-1}}{r}a}^{\frac{\sqrt{N-1}}{r}b} \overbrace{\left(1 - \frac{1}{N-1} y^2\right)}^{=: f_N(y)} \ dy$$

Da $\mathbb{P}_{N,r_N}(a \leq x_1 \leq b) = \frac{h_{r_N}(a,b)}{h_{r_N}(-r_n, r_N)}$ ist kürzen sich die Faktoren vor den Integralen.
Außerdem ist $\lim_{N\to\infty} f_N(y) = \exp(-y^2)^{1/2} = e^{-\frac{1}{2}y^2}$ gleichmäßig auf Kompakta.
(d.h. $\lim_{N\to\infty} \sup \left\{ \left| f_N(y) - e^{-\frac{y^2}{2}} \right| : |y| \leq C \right\} = 0$)
Mit dieser Information kann man relativ leicht zeigen, dass für $a_\infty = \lim_{N\to\infty} \frac{\sqrt{N-1}}{r_N} a$, $b_\infty = \lim_{N\to\infty} \frac{\sqrt{N-1}}{r_N} b$ gilt

$$\lim_{N\to\infty} \int_{\frac{\sqrt{N-1}}{r_N}a}^{\frac{\sqrt{N-1}}{r_N}b} f_N(x) dy = \begin{cases} \int_{a_\infty}^{b_\infty} e^{-\frac{y^2}{2}} \ dy, & \text{falls } -\infty \leq a_\infty b_\infty \leq \infty \\ 0, & \text{falls } a_\infty = b_\infty \in \{-\infty, 0, \infty\} \end{cases}$$

Aus **??** für $h_{r_N}(-r_n, r_N)$ folgt die Behauptung in allen Fällen. $\square$

3 Bedingte Wahrscheinlichkeit, Unabhängigkeit

Definition 3.1. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $A \in \mathscr{F}$ mit $\mathbb{P}(A) > 0$. Dann heißt die Abbildung

$$\mathbb{P}_A : \mathscr{F} \to [0,1], \quad B \mapsto \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

bedingtes Wahrscheinlichkeitsmaß unter der Bedingung A.
Für festes $B \in \mathscr{F}$ heißt die Zahl $\mathbb{P}(B|A) := \mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$ die **bedingte Wahrscheinlichkeit** von B unter der Bedingung A.

Proposition 3.2. $\mathbb{P}_A$ ist das eindeutige Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$ mit den Eigenschaften

(i) $\mathbb{P}_A(A) = 1$

(ii) $\exists c > 0$ mit $\mathbb{P}_A(B) = c\mathbb{P}(B) \forall B \in \mathscr{F}$ mit $B \subseteq A$.

Beweis. a) $\mathbb{P}_A$ ist ein Wahrscheinlichkeitsmaß und erfüllt (i),(ii).

b) $\mathbb{P}$ ist eindeutig:
Sei $\tilde{\mathbb{P}}_A$ ein Wahrscheinlichkeitsmaß mit den Eigenschaften (i),(ii).
Dann gilt $\tilde{\mathbb{P}}_A(B) = \tilde{\mathbb{P}}_A(A \cap B) + \tilde{\mathbb{P}}_A(B \setminus A = c\mathbb{P}(A \cap B))$.
Mit $B = A$ folgt $1 = \tilde{\mathbb{P}}_A(A) = c\mathbb{P}(A)$, also $\frac{1}{\mathbb{P}(A)}$. Also ist

$$\tilde{\mathbb{P}}_A(B) = \frac{1}{\mathbb{P}(A)} \mathbb{P}(A \cap B) = \mathbb{P}_A(B)$$

□

Bemerkung 3.3. $\mathbb{P}_A$ modelliert die Situation, dass wir wissen, dass A sicher eintritt. $\mathbb{P}_A$ beschreibt das Modell welches diese Information berücksichtigt (**??**(i)), aber sonst möglichst wenig ändert (**??**(ii)).

Beispiel. 2-Mal Würfeln, Information ‚Summe wird 10 sein'.
(In der Praxis: alle Würfe mit $X_1 + X_2 \neq 10$ werden ungültig gemacht.)

$$\mathbb{P}_A(\underbrace{X_1 = 5}_{B}) = \mathbb{P}\big(\underbrace{\{(5,x) : 1 \leq x \leq 6\}}_{B} \cap \underbrace{\{(x,y) : x + y = 10\}}_{A}\big) / \mathbb{P}\big(\underbrace{\{(x,y) : x + y = 10\}}_{A}\big)$$
$$= \frac{\mathbb{P}(\{(5,5)\})}{\mathbb{P}(\{(4,6),(5,5),(6,4)\})} = \frac{1}{3}$$

Satz 3.4. *Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $N \leq \infty$, $(B_i)_{i=1,\dots,N}$ mit paarweise disjunkten $B_i \in \mathscr{F}$ und $\bigcap_{i=1}^N B_i = \Omega$ (eine abzählbare Partition von Ω)*

a) Fallunterscheidungsformel:

$$\forall A \in \mathscr{F} \textit{ ist } \mathbb{P}(A) = \sum_{i=1}^{N} \mathbb{P}(B_a)\mathbb{P}(A|B_i)$$

(Konvention: $\mathbb{P}(A|B_i) = 0$, falls $\mathbb{P}(B_i) = 0$)

b) Formal von Bayes: $\forall k \leq N$ ist

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k)\mathbb{P}(A|B_k)}{\sum_{i=1}^N \mathbb{P}(B_i)\mathbb{P}(A|B_i)} \qquad (\star)$$

Beweis. a) $\sum_{i=1}^N \mathbb{P}(B_i)\mathbb{P}(A|B_i) = \sum_{i=1}^N \mathbb{P}(B_i \cap A) = \mathbb{P}\left(\bigcup_{i=1}^N B_i \cap A\right) = \mathbb{P}(A)$

b) $\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B_k)\mathbb{P}(B_k|A)}{\mathbb{P}(A)} = (\star)$

□

Beispiel 3.5 (Bayes Formel in der Medizin, False-Positive beim HIV-Test). Sei $B_1 = \{\text{Menschn mit HIV}\}$, $B_2 = \Omega \setminus B_1 = \{gesundeMenschen\}$. Empirisch Bekannt $\mathbb{P}(B_1) = 0.02$ (2% infizierte), $\mathbb{P}(A|B_1) = 0.95$ (Sensitivität 95%), $\mathbb{P}(A|B_2) = 0.1$ (Spezifität 10%).
Angenommen „Test ist Positiv“:

$$\mathbb{P}(B_1|A) = \frac{\mathbb{P}(B_1)\mathbb{P}(A|B_1)}{\mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2)} = \frac{0.02 \cdot 0.95}{0.02 \cdot 0.95 + 0.98 \cdot 0.1} \approx \frac{1}{6}$$

Viel kleiner als die naiv vermuteten 0.9.

3.A Galton-Watson-Prozesse und Wahrscheinlichkeitsbäume

Beispiel 3.6 (Mehrstufiges Modell).

a) 1 Lebewesen bekommt $X_{1,1} \in \mathbb{N}_0$ Nachkommen und stirbt danach.
$X_{1,1}$ ist Zufallsvariable.

b) Die $X_{1,1}$ Nachkommen bekommen jeweils $X_{2,1}, X_{2,2}, ..., X_{2,X_{1,1}}$ Nachkommen und stirbt. Nun leben

$$Y_2 = \sum_{i=1}^{X_{1,1}} X_{2,i}$$

Lebewesen.

c) Die Y_2 Lebewesen bekommen jeweils $X_{3,1}, X_{3,2}, ..., X_{3,Y_2}$ Nachkommen.

Proposition 3.7. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ein Wahrscheinlichkeitsraum und $A_1, ..., A_n \in \mathscr{F}$. Dann gilt

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

Beweis. Falls $\mathbb{P}(A_1 \cap ... \cap A_n) = 0$, dann (Konvention) ist auch $\mathbb{P}(A_n|A_1 \cap ... \Cap A_{n-1}) = 0$. Sonst

$$\begin{aligned}
&\mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \cdot ... \cdot \mathbb{P}(A_1|A_1 \cap ... \cap A_{n-1}) \\
&= \mathbb{P}(A_1) \frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)} \frac{\mathbb{P}(A-1 \cap A_2 \cap A-3)}{\mathbb{P}(A_1 \cap A_2)} \cdots \frac{\mathbb{P}(A_1 \cap .... \cap A_n)}{\mathbb{P}(A-1 \cap ... \cap A_{n-1})} \\
&= \mathbb{P}(A-1 \cap ... \cap A_n).
\end{aligned}$$

□

Satz 3.8. *Sei $N \leq \infty$ und seien $(\Omega_i, \mathscr{F}_i)$ Messräume und die Ω_i abzählbar. Sei ρ_1 Zähldichte auf Ω_1 und*
$\forall k < N, \omega_i \in \Omega_i, i \leq k$ sei $\rho_{k+1|\omega_1,...,\omega_k}$ Zähldichte auf Ω_{k+1}.
Sei $\Omega = \bigtimes_{i=1}^{N} \Omega_i$ und $X_o : \Omega \to \Omega_i$, $\omega = (\omega_1, ...,) \mapsto \omega_i$ die i-te Projektion.

Dann esxistiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \bigotimes_{i=1}^{N} \mathscr{F}_i)$ mit den Eigenschaften

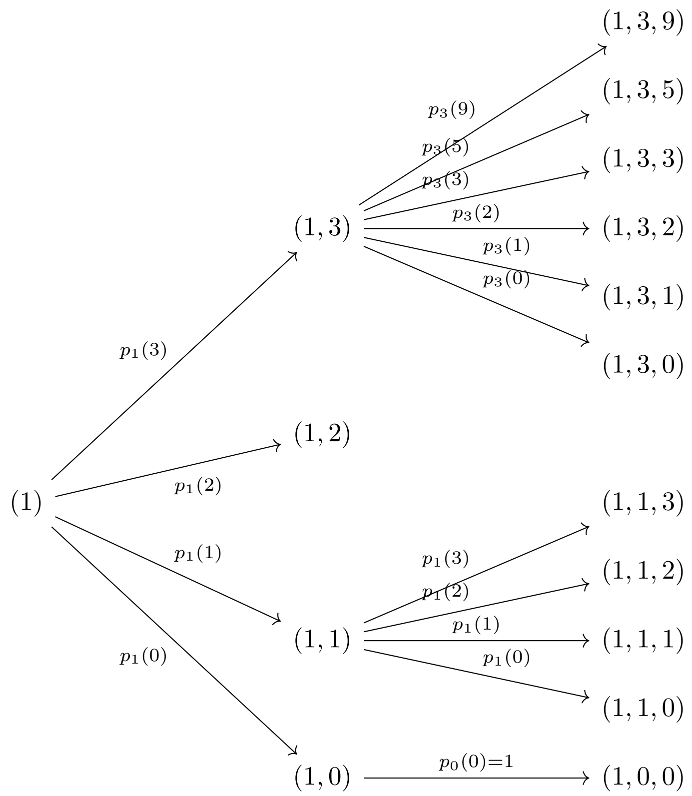

Abbildung 1: ...

a) $\mathbb{P}(X_1 = \omega_1) = \rho_1(\omega_1)$ *für alle* $\omega_1 \in \Omega_1$.

b) $\forall k < N, \omega \in \Omega$ *und falls* $\mathbb{P}(X_1 = \omega_1, ..., X_k = \omega_k) \neq 0$*, dann ist*

$$\mathbb{P}(X_{k+1} = \omega_{k+1} \mid X_j = \omega_j \forall j \leq k) = \rho_{\omega_{k+1}|\omega_1,...,\omega_k}(\omega_{k+1})$$

Bemerkung 3.9. a) Insbesondere osz $\mathbb{P}(A_1 = \omega_1, ..., X_{k+1} = \omega_{k+1}) = \rho_1(\omega_1) \cdot ... \cdot \rho_{k+1|\omega_1,...,\omega_k}(\omega_{k+1})$. („Produkt entlang der Äste“)

b) Falls $N < \infty$ dann hat $\mathbb{P}$ die Zähldichte $\rho : \omega = (\omega_1, ..., \omega_k) \mapsto \rho_1(\omega_1) \cdot ... \cdot \rho_{N|\omega_1,...,\omega_{N+1}}(\omega_N)$.

c) Falls $N = \infty$, dann hat $\mathbb{P}$ im Allgemeinen keine Zähldichte

Beweis. a) Falls $N < \infty$: Nachrechnen

b) Falls $N = \infty$ Bilde $[0, 1] \to \Omega$ mittels $x \mapsto (\omega_1(x), ...)$ mit $\omega_i(x)$= dasjenige $\omega_i \in \Omega_i$, sodass x im zu ω_i gehörigen Intervall liegt, in Stufe i.
Zeige dann $X : x \mapsto \omega(x)$ ist Zufallsvariable von $([0, 1], \mathscr{B}([0, 1]))$ nach $(\Omega, \mathscr{F})$. $\mathbb{P}$ ist dann das Bildmaß .

□

Beispiel 3.10.1 (unendlich oft wiederholter Münzwurf). Sei $\Omega = \{-1,1\}^N$ und $\mathscr{F} = \bigotimes_{i=1}^{\infty} \mathscr{P}(-1,1)$.
($\mathscr{F}$ wird also von den Mengen

$$\left\{\{\omega \in \Omega : \omega_1 = k_1, ..., \omega_n = k_n\} : K_i \in \{-1,1\} \forall i \leq nnn \in \mathbb{N}\right\}$$

erzeugt. und $\mathbb{P}(\omega_1 = k_1, ..., \omega_n = k_n) = 2^{-n}$.
Im Fall von Satz 3.8 bedeutet das $\rho_1(\omega_1) = \frac{1}{2}, \rho_{2,\omega_1}(\omega_2) = \frac{1}{2}$.

Beispiel 3.10.2 (Unendlich oft wiederholtes würfeln, 10-Seitiger Würfel). Sei $\Omega = \{0,1,...,9\}$. Die Abbildung X aus 3.8 ist hier:

$$X : [0,1) \mapsto \Omega = \mathop{\times}_{i=1}^{\infty} \Omega_i, x \mapsto \omega(x)$$

Definition 3.11. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $A, B \in \mathscr{F}$.
A, B heißen **unabhängig** (oder **unabhängige Ereignisse**), falls $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Man schreibt $A \perp\!\!\!\perp B$.

Bemerkung.

a) Falls $A \perp\!\!\!\perp B$, dann ist

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

b) Beachte:
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ („additiv“), falls A, B
$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ („multiplikativ“), falls A, B unabhängig (per Definition).

c) Aussage b) gilt zwar für $A_1, A_1, A_3 \mathscr{F}$ falls $\mathbb{P}(A_i \cup A_j) = \mathbb{P}(A_i) + \mathbb{P}(A_j)$ für $i, j \in \{1,2,3\}$und $i \neq j$. Dann gilt auch

$$\mathbb{P}(A_1 \cup A_2 \cup A_3) = \mathbb{P}(A_1) + \mathbb{P}(A_2)\mathbb{P}(A_3)$$

Aber: gilt nicht für Unabhängigkeit.
Unabhängigkeit ist eine algebraische Eigenschafte von Mengen und Wahrscheinlichkeitsmaßen.
Die Interpretation „A beeinflusst B nicht“ ist nicht immer richtig.

d) Unabhängigkeit trotz Kausalität: $\Omega = \{1,2,3,4,5,6\}^2, \mathbb{P}(A \cap B)\mathbb{P}(\{6,1\}) = \frac{1}{36} = \mathbb{P}(A)\mathbb{P}(B)$. Geht nicht wenn man 7 durch 8 ersetzt!

e) $\mathbb{P}(A) \in \{0,1\}$, es folgt $A \perp\!\!\!\perp A$

f) Im wichtigen Fall der Produktmaße sind jedoch c) bis e) nicht relevant

Definition 3.12. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $I \neq 0$ eine Indexmenge. $(A_i)_{i \in I}$ heißt **unabhängige** Familie von Mengen, wenn $\forall J \subseteq$, mit $|J| < \infty$

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) \prod_{j \in J} \mathbb{P}(A_j)$$

(d.h. Multiplikativ für alle Kombinationen von endlich vielen Mengen) gilt.

Definition 3.13. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $I \neq 0$ sei Indexmenge und $\forall i \in I$ seien $(\Omega_i, \mathscr{F}_i)$ Ereignisräume. Seien $Y_i : \Omega \to \Omega_i$ Zufallsvariablen. Die Familie $(Y_i)_{i\in I}$ heißt **unabhängig** (oder **unabhängige Familie von Zufallsvariablen**), wenn $\forall J \subseteq$, mit $|J| < \infty$ und für alle $(B_j)_{j\in J}$ mit $B_j \in \mathscr{F}_j$

$$\mathbb{P}\left(\bigcap_{j\in J} Y_j^{-1}(B_j)\right) = \prod_{j\in J} \mathbb{P}\left(Y_j^{-1}(B_j)\right)$$

gilt.

Definition 3.13. Eien Familie $(\mathscr{A}_i)_{i\in I}$ von **Megensystemen** heißst **unabhängig** (oder auch unabhänguge Famile von Mengensystemen), wenn $\forall J \subseteq$, mit $|J| < \infty$, für alle $(A_j)_{j\in J}$ mit $A_j \in \mathscr{A}_j$ sind die Mengen $(A_j)_{j\in J}$ unabhängig.
(Dabei darf aus jedem $\mathscr{A}_j$ höchstens ein A_j gewählt werden.)
Daher $X \perp\!\!\!\perp Y \Leftrightarrow \sigma(A) \perp\!\!\!\perp \sigma(Y)$.

Satz 3.14. *In 3.13 sei $\mathscr{G}_i$ ein Schnitt-stabiler Erzeuger von $\mathscr{F}_i$. Dann*

$$\textbf{??}\ \textit{gilt}\ \forall B_i \in \mathscr{G}_i \Leftrightarrow \textbf{??}\ \textit{gilt}\ \forall B_i \in \mathscr{F}_i$$

Beweis. **„$\Leftarrow$“** klar, da $\mathscr{G}_i \subseteq \mathscr{F}_i$

„$\Rightarrow$“ Durch Induktion nach $n := |\{i \in J : B_i \in \mathscr{F} - i \setminus \mathscr{G}_i\}|$:

$n = 0$ bedeutet $B_i \in \mathscr{G}_i \forall i \in J$.

$n \to n+1$ Seien $(B_i)_{i\in J}$ mit $B_i \in \mathscr{F}_i \forall i \in J$ und $B_i \in \mathscr{F}_i \setminus \mathscr{G}_i$ (n+1)-mal. Wähle $j \in J$ und setze $J' = J \setminus \{j\}$, dann folgt aus der Induktionsannahme, dass für

$$A : \bigcap_{i\in J'} Y_i^{-1}(B_i)$$

$$\mathbb{P}(A) = \prod_{i\in J'} \mathbb{P}(Y_i^{-1}(B_i)) \qquad (\star)$$

gilt.
Falls $\mathbb{P}(A) = 0$, dann ist auch $\mathbb{P}(A \cap Y^{-1}(B_j)) = 0$.
Falls $\mathbb{P}(A) > 0$, dann definiere die Wahrscheinlichkeitsmaße

$$\begin{aligned} \mathbb{P}_i : \mathscr{F}_j &\to [0,1], \\ \tilde{B}_i &\mapsto \mathbb{P}(Y_j^{-1}(\tilde{B}_j)|A) = \frac{\mathbb{P}(Y_j^{-1}(\tilde{B}_j) \cap A)}{\mathbb{P}(A)} \qquad (\star\star) \\ \mathbb{P}_2 : \mathscr{F} : i &\to [0,1], \\ \tilde{B}_j &\mapsto \mathbb{P}(Y_j^{-1}(\tilde{B}_j)) \end{aligned}$$

Da **??** für n gilt ist $\forall \tilde{B}_j \in \tilde{\mathscr{G}}_j$, also

$$\mathbb{P}_1(\tilde{B}_j) = \frac{\mathbb{P}\big(Y_j^{-1}(\tilde{B}_j)\big) \cdot \prod_{i\in J'} \mathbb{P}\big(Y_i^{-1}(B_i)\big)}{\mathbb{P}(A)} \overset{(\star)}{=} \mathbb{P}\big(Y_j^{-1}(\tilde{B}_j)\big) = \mathbb{P}_2(\tilde{B}_j)$$

Es folgt, dass $\mathbb{P}_1(\tilde{B}) = \mathbb{P}_2(\tilde{B}) \forall \tilde{B} \in \mathscr{G}_j$, sodass aus **??** folgt, dass $\mathbb{P}_1 = \mathbb{P}_2$. Dann

$$\prod_{i \in J} \mathbb{P}\big(Y_i^{-1}(B_i)\big) \overset{(\star)}{=} \mathbb{P}\big(Y_j^{-1}(b_i)\big)\mathbb{P}(A) = \mathbb{P}_2(B_j)\mathbb{P}(A)$$
$$= \mathbb{P}_1(B_j)\mathbb{P}(A) = \mathbb{P}\big(Y_j^{-1}(B_j) \cap A\big) = \mathbb{P}\left(\bigcap_{i \in J} Y_i^{-1}(B_i)\right)$$

für alle $B_i \in \mathscr{F}_i$. Es folgt die Behauptung.

$\square$

Korollar 3.15. *Seien $(Y_i)_{i=1,\dots,n}$ Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$, dann gilt*

a) Falls $Y_i : \Omega \to E_i$ mit E_i abzählbar, dann

$$(Y_i) \text{ ist unabhängig} \Leftrightarrow \mathbb{P}(Y_1 = e_1, ..., Y_n = e_n) = \prod_{i=1}^{n} \mathbb{P}(Y_i = e_i)$$

für alle $e_1 \in E_1, ..., e_n \in E_n$.

b) Falls $Y_i : \Omega \to \mathbb{R}$, dann ist

$$(Y_i) \text{ ist unabhängig} \Leftrightarrow \mathbb{P}(Y_1 \leq c_1, ..., Y_n \leq c_n) = \prod_{i=1}^{n} \mathbb{P}(Y_i \leq c_i)$$

für alle $c_1, ..., c_n \in \mathbb{R}$.

Korollar 3.16. *Sei $(\Omega, \mathscr{F}\mathbb{P})$ ein Wahrscheinlichkeitsraum, I eine Menge und für alle $i \in I$ sei $A_i \in \mathscr{F}$. Dann gilt*

a) $(A_i)_{i \in I}$ sind unabhängige Mengen $\Leftrightarrow$ $(\mathbb{1}_{A_i})_{i \in I}$ unabhängige Zufallsvariablen sind.

b) Insbesondere: Falls $(A_i)_{i \in I}$ unabhängig und $C_i \in \{\emptyset, \Omega, A_i, A_i^C\}$ für alle $i \in I$, dann sind auch $(C_i)_{i \in I}$ unabhängig.

Beweis. Für $a \in \mathscr{F}$ ist $\mathbb{1}_A : \Omega \to \{0,1\}, \omega \mapsto \mathbb{1}_A(\omega)$ eine Zufallsvariable und das einelementige Mengensystem $\{\}$ ist ein Schnitt-stabiler Erzeuger von $\mathscr{P}(\{0,1\})$ und $\mathbb{1}_a^{-1}(\{1\}) = A$.
Daher gilt, dass (A_i) unabhängig $\Leftrightarrow$ **??** gilt falls $B_i = \{1\} \overset{3.14}{\Leftrightarrow} \mathbb{1}_{A_i}$ ist unabhängig. Also gilt 1.
Für 2. Setze

$$B_i = \begin{cases} \{1\} & \text{falls } c_i = A \\ \{0\} & \text{falls } c_i = A^c \\ \emptyset & \text{falls } c_i = \emptyset \\ \{0,1\} & \text{falls } c_i = \Omega \end{cases}$$

Dann folgt 2 aus der Unabhängigkeit der $\mathbb{1}_{A_i}$ und aus **??** mit den so gewählten B_i. $\square$

Beispiel 3.17. Punkt auf Kreisscheibe: Winkel und Radius unabhängig

$$
\begin{tikzcd}
\Omega \arrow[r, "Y_1"] \arrow[dr, "Y_2"] \arrow[rr, bend left, "Z_1"] \arrow[drr, bend right, "Z_2"'] & \Omega_1 \arrow[r, "\phi_1"] & E_1 \\
& \Omega_2 \arrow[r, "\phi_2"'] & E_2
\end{tikzcd}
$$

Abbildung 2: ...

Satz 3.18. *Seien $(Y_i)_{i=1}^{\infty}$ unabhängige Zufallsvariable, mit $Y_i : (\Omega, \mathscr{F}) \to (\Omega_i \mathscr{F}_i)$. Dann gilt*

a) Falls $\phi_i : (\Omega_i \mathscr{F}_i) \to (E_i \mathscr{E}_i)$ messbar und $Z_i : \Omega \to E_i, \omega \mapsto \phi_i(Y_i(\omega))$, dann sind auch $(Z_i)_{i\in\mathbb{N}}$ unabhängig. („Stabilität unter einsetzen in Funktionen“)

b) Seien $J_1, J_2, ...,$ paarweise disjunkte Teilmengen von $\mathbb{N}$. Für jedes $k \in \mathbb{N}$ definiere

$$W_k(\omega) = (X_i(\omega))_{i\in J_k} \in \bigtimes_{i\in J_k} \Omega_i$$

Dann sind die Zufallsvariablen $(W_k)_{k\in\mathbb{N}}$ unabhängig:
(„Stabilität gegenüber Zusammenfassen in disjunkte Blöcke“)

Beweis. a) Sei $I \subset \mathbb{N}$ endlich. Dann ist (für $A_i \in \mathscr{E}_i$)

$$\mathbb{P}\left(\bigcap_{i\in I} Z_i^{-1}(A_i)\right) = \mathbb{P}\left(\bigcap_{i\in I} Y_i^{-1}(A_i)\right) \overset{Y_i \text{ unahängig}}{=} \prod \mathbb{P}(Y_i^{-1}(\phi^{-1}(A_i)))$$

b) Es gilt $\times_{i=1}^{n} A_i) = \{(\omega_1, ..., \omega_n) : \omega_1 \in A_1, ..., \omega_n \in A_n\} \subset \times_{i=1}^{n} \Omega_i$.
Für $k \in \mathbb{N}$ setze

$$\mathscr{G}_k := \left\{ \bigtimes_{i\in J_k} A_i^{(k)} : A^{(k)} \in \mathscr{F}_i \forall i \in J_k \; A_i^{(k)} \neq \Omega_i \text{ nur endlich oft} \right\}$$

$\mathscr{G}_k$ ist Schnitt-stabiler Erzeuger von $\bigotimes_{i\in J_k} \mathscr{F}_i$, also miss **??** nur auf $(G_k)_{k\in\mathbb{N}}$ geprüft werden.
Seien also $B_{k_1} \in \mathscr{G}_{k_1}, ..., B_{k_n} \in \mathscr{G}_{k_n}$. Da $W_{k_j} \in B_{k_j}$ genau dann wenn $X_i \in A_i^{kj}$ ist

$$\begin{aligned}
\mathbb{P}(W_{k_1} \in B_{k_1}, ..., W_{k_n \in P_{k_n}}) &= \mathbb{P}(X_i \in A_i^{(k)} \forall i \in J_{k_1,...,} X_i \in A_i^{(k)} \forall i \in J_k) \\
&= \prod_{l=1}^{n} \prod_{i\in J_{k_l}} \mathbb{P}(X_i \in A_i^{(k_l)}) = \prod_{l=1}^{n} \mathbb{P}(W_{k_l} \in B_{k_l})
\end{aligned}$$

□

Bemerkung. Unabhängigkeit ist eng verwandt mit dem Produktmaß:

Definition 3.21. Sei $(\Omega, \mathscr{F}_i, \mathbb{P}_i)$ ein Wahrscheinlichkeitsraum für $i \in \mathbb{N}$.
Jedes Maß $\mathbb{P}$ auf $(\Omega, \mathscr{F}) := (\times_{i=1}^{\infty} \Omega_i, \bigotimes_{i=1}^{\infty} \mathscr{F}_i)$

($\bigotimes_{i=1}^{\infty} \mathscr{F}_i$ erzeugt durch Zylindermengen)
mit der Eigenschaft

$$\mathbb{P}\left(\{\omega = (\omega_i)_{i\in\mathbb{N}} \in \Omega : \omega_{j_i} \in A_{j_i}, ..., \omega_{j_m} \in A_{j_m}\}\right) = \prod_{k=1}^{m} \mathbb{P}_{jk}(A_{jk}) \tag{3.2}$$

für alle $j_1 < j_2 < ... < j_m$, mit $j_i \in \mathbb{N}$ und $A_{ji} \in \mathscr{F}_{ji}$ heißt **Produktmaß** der Maße $\mathbb{P}_i$. Schreibweise $\mathbb{P} = \bigotimes_{i=1}^{\infty} \mathbb{P}_i$

Bemerkung 3.22. a) In 3.21existiert immer genau ein Produktmaß. (Eindeutigkeit nach Satz **??**, Existenz aus dem Maßfortsetzungssatz)

b) Wichtiger Spezialfall:

(i) Sei $\Omega_i = \{0, 1\}$, $\mathbb{P}_i(0) = \mathbb{P}_i(1) = \frac{1}{2}$ für alle i (unendlich oft wiederholter Münzwurf)

(ii) Sei $\Omega_i = \mathbb{R}$, $\mathscr{F}_i\mathscr{B}(\mathbb{R})$, $\mathbb{P}_i = \rho(x)\ dx$ für alle i.
(Zufällige Folge mit Gliedern die gemäß $\rho(x)\ dx$ verteilt sind.)

Proposition 3.23. a) Seien $(X_i)_{i\in\mathbb{N}}$Zufallsvariablen, $X_i : (\Omega, \mathscr{F}) \to (\Omega_i\mathscr{F}_i)$, dann gilt

(X_i) ist unabhängig $\Leftrightarrow$ Das Bildmaß der Zufallsvariablen $X : \Omega \to \bigtimes_{i=1}^{\infty} \Omega_i, \omega \mapsto (X_i(\omega))_{i\in\mathbb{N}}$ ist ein Pro

b) Seien $(\Omega_i\mathscr{F}_i)_{i\in\mathbb{N}}$ seien Maßräume, $\Omega = \times_{i=1}^{\infty}, \mathscr{F} = \bigotimes_{i=1}^{\infty} \mathscr{F}_i$.
$\mathbb{P}$ sei ein beliebiges Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$. Dann gilt:

$\mathbb{P}$ ist Produktmaß $\Leftrightarrow$ Die Projektionenn $X_i : \Omega \to \Omega_i$, $\omega = (\omega_1, \omega_2, ...) \mapsto \omega_i$ sin unabhängige Zufallsvar

Bemerkung 3.24. In 3.23a) heißt der Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}}) := (\times_{i=1}^{\infty} \Omega_i, \bigotimes_{i=1}^{\infty}, \mathbb{P}_X)$ mit $X = (X_i)_{i\in N}$ der kanonische Wahrscheinlichkeitsraum der Zufallsvariablen X_i. Die i-te Koordinatenabbildung $Y_i : \tilde{\Omega} \to \Omega_i$ entspricht dann jeweis X_i, d.h. $(Y_{i_1}, ..., Y_{i_n})$ hat die gleiche Verteilung wie $(X_{i_1}, ..., X_{i_n})$.

Definition 3.25. Seien $\Pr_1, ..., \mathbb{P}_n$ Wahrscheinlichkeitsmaße auf $\mathbb{R}$ mit der Dichte $\rho_1, ..., \rho_n$. Dann heißt

$$\rho : \mathbb{R}^n \to \mathbb{R}_0^+, (x_1, ..., x_n) \mapsto \rho(x_1, ..., x_n) = \rho_1(x_1)...\rho_n(x_n)$$

die **Produktdichte**.

Satz 3.25. *Die Produktdichte ist die Dichte des Produktmaßes.*

Beweis. Es gilt

$$\begin{aligned}
\mathbb{P}\big((-\infty, c_1] \times (-\infty, x_2] \times ... \times (-\infty, c_1]\big) &= \mathbb{P}(X_1 \le c_1, ..., X_n \le c_n) \\
&\overset{\textbf{??}b)}{=} \mathbb{P}(X_1 \le c)...\mathbb{P}(X_n \le n) \\
&= \int_{-\infty}^{c_1} \rho_1(x_1)\ dx_1 \cdot ... \cdot \int_{-\infty}^{c_n} \rho_n(x_n)\ dx_n \\
&\overset{\substack{\text{Satz von} \\ \text{Fubini}}}{=} \int_{(-\infty,c_1]\times...\times(-\infty,c_n]} \rho_1(x_1)...\rho_n(x_n)\ dx_1...\ dx_n \\
&= \int_{(-\infty,c_1]\times(-\infty,x_2]\times...\times(-\infty,c_1]} \rho(x)\ dx
\end{aligned}$$

Dann folgt mit **??**, dass $\mathbb{P}$ die Dichte ρ hat. $\square$

Bemerkung. Für unabhängige Zufallsvariablen X, Y kann man die Dichte $X+Y$ mittels Faltungsprodukt berechnen.

Satz 3.26. *Seien $X, Y : \Omega \to \Omega'$ unabhängige Zufallsvariablen*

a) Falls $\Omega' = \mathbb{Z}$ und falls $\mathbb{P}_X$ (bzw $\mathbb{P}_Y$) Zähldichte ρ_1 (bzw. ρ_2), dann hat $\mathbb{P}_{X+Y}$ die Zähldichte

$$\rho_1 * \rho_2 : \mathbb{Z} \to [0,1], z \mapsto \rho_1 * \rho_2(z) = \sum_{k \in Z} \rho_1(k)\rho_2(z-k)$$

b) Falls $\Omega' = \mathbb{R}$ und falls $\mathbb{P}_X$ (bzw $\mathbb{P}_Y$) die Dichte ρ_1 (bzw. ρ_2), dann hat $\mathbb{P}_{X+Y}$ die Dichte

$$\rho_1 * \rho_2 : \mathbb{R} \to \mathbb{R}_0^+, x \mapsto \rho_1 * \rho_2(x) = \int_{\mathbb{R}} \rho_1(y)\rho_2(x-y)\ dy$$

Beweis. a) Da nur $\mathbb{P}_X, \mathbb{P}_Y, \mathbb{P}_{X+Y}$ untersucht werden kann man$\Omega = \mathbb{Z} \times \mathbb{Z}$, $X\big((z_1, z_2)\big) = z_1$ und $Y\big((z_1, z_2)\big) = z_2$ wählen und $\mathbb{P}_{(X,Y)}$ hat die Produkt-Zähldichte $(z_1, z_2) \mapsto \rho_1(z_1)\rho_2(z_2)$.
Dann gilt $X + Y = w \in \mathbb{Z}$,genau dann wenn $z_1 + z_2 = w$, bzw. $z = 1 = w - z_2$.
Also $(X + y)^{-1}(\{w\}) = \{(z_1, z_2) : z_1 = w - z_2\}$, es folgt

$$\mathbb{P}_{X+Y}(\{w\}) = \mathbb{P}_{(X,Y)}\big((X+Y)^{-1}(\{w\})\big) = \sum_{\substack{z_1 \in \mathbb{Z} \\ z_2 \in \mathbb{Z} \\ z_1 + z_2 = w}} \rho(z_1)\rho(z_2) = \sum_{z \in Z} \rho_1(z)\rho_2(w-z)$$

b) Setze $\Omega = \mathbb{R}^2$, $X\big((x_1, x_2)\big) = x_1$, $Y\big((x_1, x_2)\big) = x_2$, $\mathbb{P}\big([a,b] \times [c,d]\big) = \int_{[a,b]\times[b,c]} \rho_1(x_1)\rho_2(x_2)\ dx_1\ dx_2$. Dann ist

$$\begin{aligned}
\mathbb{P}(X + Y \leq c) &= \mathbb{P}\big(\{(x,y) \in \Omega\} : x + y \leq c\big) \\
&= \int_{-\infty}^{\infty} \rho_1 \int_{-\infty}^{c-x} \rho_2(y)\ dy\ dx \\
&= \int_{-\infty}^{\infty} \rho_1 \int_{-\infty}^{c} \rho_2(\tilde{y} - x)\ d\tilde{y}\ dx \\
&= \int_{-\infty}^{c} \underbrace{\int_{-\infty}^{\infty} \rho_1(x)\rho_2(\tilde{y} - x)\ dx}_{=\rho_1 * \rho_2(\tilde{y})}\ d\tilde{y}
\end{aligned}$$

Dann folgt mit **??**, dass $\mathbb{P}_{X+Y}$ die Dichte $\rho_1 * \rho_2$ hat.

□

Beispiel 3.27.1. Sei $X \sim \mathscr{N}_{\mu,\nu}$ und $Y \sim \mathscr{N}_{\mu',\nu'}$, dann ist $X + Y \sim \mathscr{N}_{\mu+\mu',\nu+\nu'}$

Beispiel 3.28. Sei $\Omega = \mathbb{R}^n$, $\mathscr{F} = \otimes_{x=1}^{\infty} \mathscr{B}(\mathbb{R})$, $X_i(\omega) = \omega_i$ Projektionen. Sei

$$\begin{aligned}
A_1 &= \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} \omega_n \text{ exitsiert}\right\} \\
A_2 &= \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_n(\omega) = 0\right\} \\
A_3 &= \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_n(\omega) = \lim_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_n \text{ exitsiert}\right\}
\end{aligned}$$

Für A_1, A_2, A_3 gilt: Sei $\omega \in A_j$. Wenn $\tilde{\omega}_i = \omega_i$ für alle außer endlich viele $i \in \mathbb{N}$ gilt, dann ist $\tilde{\omega} \in A_j$. Ebenso wenn $\omega \in A_j^C$, dann $\omega \in A_j^C$.

Definition 3.29. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und seien $X_i : (\Omega, \mathscr{F}) \to (\Omega_i, \mathscr{F}_i)$ seien Zufallsvariablen. Setze

$$\mathscr{F}_{\{i\}} = \sigma(X_i) = \{X_i^{-1}(B) : B \in \mathscr{F}_i\} \qquad \text{(„von } X_i \text{ erzeugte } \sigma\text{-Algebra“)}$$

Setze für $K \subseteq \mathbb{N}$

$$\mathscr{F}_K := \sigma(\{X_i : i \in K\}) = \sigma(\mathscr{F}_{\{i\}} : i \in K) \qquad \text{(„von } (X_i)_{i\in K} \text{ erzeugte } \sigma\text{-Algebra“)}$$

Schreibe für $n \in \mathbb{N}$

$$\begin{aligned} \mathscr{F}_n &:= \mathscr{F}_{\{1,\dots,n\}} && \text{(„}\sigma\text{-Algebra bis zur ‚Zeit‘ } n\text{“)} \\ \tau_n &:= \mathscr{F}_{n+1,n+2,\dots} && \text{(„}\sigma\text{-Algebra nach der Zeit } n\text{“)} \\ \tau &:= \bigcap_{n\in\mathbb{N}} \tau_n && \text{(„terminale/asymptotische } \sigma\text{-Algebra“)} \end{aligned}$$

Beispiel 3.30. Für $\Omega = \mathbb{R}^N$, $\mathscr{F} = \mathscr{B}(\mathbb{R}^N)$, X_i ist i-te Projektion. Dann ist

$$\mathscr{F}_{\{1,\dots,n\}} = \Big\{ \big\{ x \in \mathbb{R}^N : (x_1, \dots, x_n) \in B, x_{n+1} \text{ beliebig}, x_{n+1} \text{ beliebig}, \dots \big\} : B \in \mathscr{B}(\mathbb{R}^n) \Big\}$$

sind Zylindermengen.
Für $|K| = \infty$ (insbesondere für τ_n) kann man $\mathscr{F}_K$ nicht explizit angeben. Stattdessen: τ_n =kleinste σ-Algebra, die $\mathscr{F}_{n+1,n+2,\dots,m}$ enthält für alle $m > n$.

Definition 3.31. Sei $(\Omega, \mathscr{F})$ Messraum, $A_i \in \mathscr{F}$ für alle $\in \mathbb{N}$. Definiere

$$\limsup_{n\to\infty} A_n := \bigcap_{m=1}^{\infty} \bigcup_{k\geq m} A_k = \{\omega, \in \Omega : \omega \in A_i \text{ für unendlich viele } i\}$$

$$\liminf_{n\to\infty} A_n := \bigcup_{m=1}^{\infty} \bigcap_{k\geq m} A_k = \{\omega, \in \Omega : \omega \in A_i \text{ für alle außer endlich viele viele } i\}$$

Proposition 3.32. a) Die Gleichheiten in 3.31 gelten

b) $\bigcap_{n=1}^{\infty} A_n \subseteq \liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A - n \subseteq \bigcap_{n=1}^{\infty} A_n$

c) Für alle $\omega \in \Omega$ gilt

$$\liminf_{n\to\infty} \mathbb{1}_{A_i}(\omega) = \mathbb{1}_{\liminf_{n\to\infty} A_i}(\omega)$$
$$\limsup_{n\to\infty} \mathbb{1}_{A_i}(\omega) = \mathbb{1}_{\limsup_{n\to\infty} A_i}(\omega)$$

d) Falls $X_i : (\Omega, \mathscr{F}) \to (\Omega_i \mathscr{G}_i)$ Zufallsvariablen sind uns $A_i = X^{-1}(B_i)$ mit $B_i \mathscr{G}_i$, dann sind $\limsup_{n\to\infty} A_n \in \tau$ und $\liminf_{n\to\infty} A_n \in \tau$.
(Nicht aber $\bigcap_{n=1}^{\infty} A_n$ oder $\bigcup_{n=1}^{\infty}$!)

Beispiel 3.33. In der Situation von 3.29 mit $\Omega_i = \mathbb{R}$ ist

$$A = \left\{ \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i \text{ exitsiert und liegt in } [a, b] \right\} \in \tau$$

für alle $a \leq b$.

Satz 3.34.

Bemerkung. In der Sitation von **??** können also nur Wahrscheinlichkeiten $\{0, 1\}$ auftreten, wenn die X_i unabhängig sind. Welche der beiden Möglichkeiten trifft zu?
Diese frage is i.A. schwierig zu beantworten, aber im Fall $\limsup_{n\to\infty} A_n$ gibt es folgenden Satz:

Satz 3.35 (Lemma von Borel-Cantelli)**.** *Sei* $(\Omega, \mathscr{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $A_j \in \mathscr{F}$*, für alle* $j \in \mathbb{N}$*. Dann*

a) Falls $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$*, dann ist* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$

b) Falls $(A_k)_{k\in\mathbb{N}}$ *unabhängig und* $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$*, dann ist* $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$.

Beweis. a) Für alle m ist $A \subseteq \bigcap_{k=1}^{\infty} A_k$. Also ist $\mathbb{P}(A) \leq \sum_{k=m}^{\infty} \mathbb{P}(A_k)$.
Da $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$ gilt, muss $\lim\limits_{n\to\infty} \sum_{k=m}^{\infty} \mathbb{P}(A_k) = 0$ gelten. Da die Mengenfolge $(\bigcup_{k=m}^{\infty})_{m\in\mathbb{N}}$ absteigend ist gilt

$$\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k=m}^{\infty} A_k\right) \leq \lim_{m\to\infty} \sum_{k=m}^{\infty} \mathbb{P}(A_k) = 0$$

b) Es gilt für das Komplement:

$$A^C = \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty}$$

Also ist

$$\begin{aligned}
\mathbb{P}(A^C) &\leq \sum_{m=1}^{\infty} \mathbb{P}\left(\bigcap_{k=m}^{\infty} A_K^C\right) \overset{\substack{\text{Satz ?? c)} \\ \text{absteigende} \\ \text{Mengefolge}}}{=} \sum_{m=1}^{\infty} \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=m}^{\infty} A_K^C\right) \\
&\overset{\text{unabh.}}{=} \sum_{m=1}^{\infty} \lim_{n\to\infty} \prod_{k=m}^{n} (1 - \mathbb{P}(A_k)) \\
&= \sum_{m=1}^{\infty} \lim_{n\to\infty} e^{-\sum_{k=m}^{\mathbb{P}(A_k)}} \\
&= \sum_{m=1}^{\infty} 0 = 0
\end{aligned}$$

□

4 Erwartungswert und Varianz

Bemerkung 4.1 (Motivation). Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X_i)_{i\in\mathbb{N}}$ mit $X_i : \Omega \to \mathbb{Z}$, $\mathbb{P}(X_i = k) = p_k$ seien unabhängige Zufallsvariablen.
(d.h. (p_k) ist zähldichte für jedes X_i)
Wenn man nun N-mal mit (X_i) „würfelt", erhält man N zufällige ganze Zahlen $X_1(\omega), ..., X_n(\omega)$.
(„Realisierung von $(X_1, ..., X_n)$")

Frage Wie groß ist der (zufällige) Mittelwert für $\frac{1}{N}\sum_{i=1}^{n} X_i$ typischerweise wenn n groß (unendlich groß) wird?

Beispiel $\mathbb{P}(X_i = k) = \frac{1}{6}$, falls $k \in \{1, ..., 6\}$. Dann ist der Mittelwert ≈ 3.5. Es ist aber auch $(1, 1, 1, ..., 1)$ möglich.

Überlegung p_k modelliert die „relative Häufigkeit“ mit der der Wert k auftritt. D.h. für große n sollte man cs $p_k \cdot n$-mal den Wert k erhalten.
Wähle z.B. $(-4, 2, 1, 2, -4, 0, 1, 3, 2, -3, 1, 0)$:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) &= \frac{1}{n}\left(0|\{i \le n : X_i(\omega = 0)\}| + 1|\{i \le n : X_i(\omega) = 1\}| + (-1)|\{i \le n : X_i(\omega = -1)\}| + ...\right) \\
&= \sum_{k\in\mathbb{Z}} \frac{k}{n}|\{i \le n : X_i(\omega) = k\}| \\
&\overset{?}{\approx} \sum_{k\in\mathbb{Z}} kp_k \qquad (\star)
\end{aligned}$$

Beobachtung Das letzte $\approx$ stimmt sicher nicht immer, sollte aber „meistens“ richtig sein.
Siehe dazu **??** (Gesetze der großen Zahlen)

Betrachte nun die rechte Seite von $(\star)$:

Definition 4.2. Eine reelle Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt **diskret** wenn die Menge $X(\Omega)$ höchstens abzählbar ist

Definition 4.3. Sei X eine diskrete Zufallsvariable. X heißt **integrierbar** (bzw „**Erwartungswert existiert**“), falls

$$\sum_{y\in X(\Omega)} |y|\mathbb{P}(X = y) < \infty$$

Man schreibt $X \in \mathscr{L}^1$ bzw. $X \in \mathscr{L}^1(\mathbb{P})$.

Definition 4.3. Für $X \in \mathscr{L}^1$ heißt

$$\mathbb{E}(X) = \mathbb{E}_{\mathbb{P}}(X) := \sum_{y\in X(\Omega)} \mathbb{P}(X = y) \in \mathbb{R}$$

der **Erwartungswert** von X.

Beispiel 4.4.1. Sei $X : \Omega \to \{a, b\}$, $a, b \in \mathbb{R}$ und $\mathbb{P}(X = a) = p = 1 - \mathbb{P}(X = b)$. Dann ist $\mathbb{E}(X = p \cdot 0 + (1-p)b = b + p(a - b)$.

$a = 1$**,** $b = 0$ Dann ist $\mathbb{E}(X = p$ (Münzwurf)

$a = 1$**,** $b = -1$ Dann ist $\mathbb{E}(X = 1 - 2p$ (Schrittweise Irrfahrt)

Beispiel 4.4.2. Sei T geometrisch verteilt, also $\mathbb{P}(T = k) = p(1-p)^{k-1}$, dann ist

$$\mathbb{E}(X = \sum_{k=1}^{\infty} p\overbrace{(1-p)}^{=:q}{}^{k-1} = p\underbrace{\sum_{k=1}^{\infty} kq^{k-1}}_{=:f(q)} = (\star)$$

Für die Funktion $f(q)$ gilt

$$f(q) = \frac{d}{dq}\sum_{k=0}^{\infty} q^k = \frac{d}{dy}\frac{1}{1-q}$$
$$= \frac{1}{(1-q)^2} = \frac{1}{p^2}$$

Also ist $(\star) = \frac{p}{p^2} = p$.

Satz 4.5. *Seien $X, Y \in \mathscr{L}^1$ diskrete Zufallsvariablen.*

a) Falls $X \geq Y$, dann ist $\mathbb{E}(X) \geq \mathbb{E}(Y)$

b) (i) Sei $c \in \mathbb{R}$, dann ist auch $cX \in \mathscr{L}^1$ und $\mathbb{E}(cX) = c\mathbb{E}(X)$.

(ii) Es ist auch $X + Y \in \mathscr{L}^1$ und $\mathbb{E}(X+Y) = \mathbb{E}(X) + E(Y)$

Beweis. Aus der Maßtheoreie, da $\mathbb{E}(X)$ das (elementare) Integral ist. □

Definition 4.6. Sei X reelle Zufallsvariable. Dann heißt $X_{(n)}$ für alle $n \in \mathbb{N}$

$$X_{(n)}(\omega := \frac{1}{n})\lfloor nX(\omega)\rfloor$$

die $\frac{1}{n}$-Approximation.
(Dabei ist $\lfloor\cdot\rfloor$ die Abrunden-Funktion.)

Satz 4.7. *Sei X reelle Zufallsvariable, $X_{(n)}$ aus 4.6.*

a) Für alle $n \in \mathbb{N}$ ist $X_{(n)} \leq X \leq X_{(n)} + \frac{1}{n}$

b) Falls $X_(N_0) \in \mathscr{L}^1$ für ein $n_0 \in \mathbb{N}$, dann ist $X_{(n)} \in \mathscr{L}^1$ für alle n.

c) Falls b) gilt, so ist $\left(\mathbb{E}(X_{(n)})\right)_{n\in\mathbb{N}}$ eine Cauchyfolge.

Beweis.

a) ist klar.

b) Aus a) folgt, dass $|X_{(n)}(\omega)| \leq |X_{(m)}| + \max\{\frac{1}{n}, \frac{1}{m}\}$.
Da insbesondere $1 \in \mathscr{L}^1(\mathbb{P})$ folgt die Behauptung.

c) Aus b) folgt, dass $|\mathbb{E}(X_{(n)}) - \mathbb{E}(X_{(m)})| \leq \max\{\frac{1}{n}, \frac{1}{m}\} \xrightarrow{n,m\to\infty} 0$. Ist Cauchyfolge.

□

Satz 4.8. *Die Aussagen von Satz 4.5 (Linearität und Monotonie) gelten auch für Reelle Zufallsvariablen.*

Beweis. Durch Grenzwertsätze der Maßtheorie: □

Satz 4.10 (Monotonen Konvergenz)**.** *Seien $(X_n)_{n\in\mathbb{N}}, X$ relle Zufallsvariablen mit punktweise $X_n(\omega) \leq X_{n+1}(\omega)$ für alle $\omega \in \Omega$ und $\lim\limits_{n\to\infty} X_n(\omega) = \sup_{n\in\mathbb{N}} X_n(\omega) = X(\omega)$.*
Dann gilt

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

Satz 4.11 (Majorisierte Konvergenz). *Seien $(X_n)_{n\in\mathbb{N}}, X$ relle Zufallsvariablen und es existiert $Y \in \mathscr{L}^1$, mit $|X_n(\omega)| \leq |Y(\omega)|$ für alle $\omega \in \Omega$ und $\lim\limits_{n\to\infty} X_n(\omega) = X(n)$ für alle $\omega \in \Omega \setminus \Omega_0$, mit Ω_0 Nullmenge. Dann ist*

$$\lim \liminf_{n\to\infty} \mathbb{E}(X_n) = E(X)$$

Definition 4.12. Sei $A := \{\omega \in \Omega \mid \omega \text{ hat Eigenschaft } E\} \in \mathscr{F}$.
Falls $\mathbb{P}(A^c) = 0$, dann sagen wir „die Eiegnschaft E gilt $\mathbb{P}$-**fast sicher**“.

Beispiel 4.13.1. Sei $\mathbb{P} = U[0,1] = \lambda_{[0,1]}$, $X(\omega) = \omega$.. Dann ist $X \notin \mathbb{Q}$ $\mathbb{P}$ fast sicher, denn $\lambda(\mathbb{Q} \cap [0,1]) = 0$

Beispiel 4.13.2. Sei $\mathbb{P} = U[-1,1]$, $X_n(y) = \sin\left(\frac{1}{|y|+\frac{1}{n}}\right)$, dann konvergiert $X_n(y)$ $\mathbb{P}$ fast sicher.

Bemerkung 4.14. Falls $X \geq 0$, dann erlauben wir auch $\mathbb{E}(X) = \infty$. Wir definieren dann $\mathbb{E}(X) = \lim\limits_{n\to\infty} \mathbb{E}(\min\{X, n\})$.
Satz 4.8a),b) und 4.10 gilt weiterhin, falls $X, Y \geq 0$ und $c \geq 0$.

Satz 4.15. *Seien (X_n) reele Zufallsvariable und $X_n \geq 0$, dann ist*

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$$

Beispiel 4.16. Sei $(\Omega, \mathscr{F}, \mathbb{P}) = \big((0,1), \mathscr{B}\big((0,1)\big), \lambda_{(0,1)}\big)$ und $X_n = n\mathbb{1}_{0,\frac{1}{n}}$.
Dann ist $\mathbb{E}(X_n) = 1 \forall n$, aber $\liminf_{n\to\infty} X_n(\omega) = 0$.
Also kann die Ungleichung in Satz 4.15 kann also Strikt sein.

Satz 4.17 (Integration bezüglich des Bildmaßes). *Sei $X : \Omega \to \Omega'$ Zufallsvariable, $f : \Omega' \to \mathbb{R}$ messbar.*
Es gelte entweder $f(\omega') \geq 0 \forall \omega' \in \Omega'$ oder

$$\mathbb{E}\left(|f(X)|\right) = \int_\Omega |f\big(X(\omega)\big)| \mathbb{P}(d\omega) < \infty$$

Dann gilt:

$$\mathbb{E}(f(X)) = \int_{\Omega'} f(\omega')\mathbb{P}_X(d\omega') = \mathbb{E}_{\mathbb{P}_X}(f) = \int_{\Omega'} f(\omega')\mathbb{P} \circ X^{-1}(d\omega')$$

Beweis. (i) Sei $f(\omega') = \mathbb{1}_B(\omega')$ mit $B \in \mathscr{F}'$.
Dann ist $\mathbb{E}(f(X)) = \mathbb{E}(\mathbb{1}_B(X)) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1(B)}) = \mathbb{P}_X(B)$ nach Definition von $\mathbb{P}_X$. Dann gilt die Gleichung aus 4.17.

(ii) Sei nun $f(\omega') = \sum_{n=1}^m c_n \mathbb{1}_{B_n}(\omega')$, mit $C_n \in \mathbb{R}$, $B_n \in \mathscr{F}'$. Dann ist

$$\mathbb{E}(f(X)) \overset{4.8}{=} \sum_{n=1}^m c_n \mathbb{E}(\mathbb{1}_{B_n}(X)) \overset{(i)}{=} \sum_{n=1}^m c_n \mathbb{E}_{\mathbb{P}_X}(\mathbb{1}_{B_n}) \overset{4.8b)}{=} \mathbb{E}_{\mathbb{P}_X}\left(\sum_{n=1}^m c_n \mathbb{1}_{B_n}\right) = \mathbb{E}_{\mathbb{P}_X}(f)$$

(iii) Sei nun $f(\omega') \geq 0 \forall \omega \in \Omega$. Setze

$$f_n(\omega') := \sup\left\{\frac{1}{n}\lfloor f(\omega')\rfloor, n\right\}$$

(Abgeschnittene $\frac{1}{n}$ Approximation.)
Dann folgt aus (ii), dass $\mathbb{E}(f_n(X)) = \mathbb{E}_{\mathbb{P}_X}(f_n)$. Dann folgt aus der monotonen Konvergenz, dass im Grenzwert $n \to \infty$ gilt: $\mathbb{E}(f(X)) = \mathbb{E}_{\mathbb{P}_X}(f)$.
Dann gilt die Gleichung aus 4.17.

(iv) Sei $\mathbb{E}(|f(X)|) < \infty$. Da $|f(X(\omega))| = f_+(X(\omega)) + f_-(\omega')$ ist auch $\mathbb{E}(f_-(X)) < \infty$ und $\mathbb{E}(f_+(X)) < \infty$.
Da $f(\omega') = f_+(\omega') - f_-(\omega')$ gilt ist

$$\mathbb{E}(f(X)) = \mathbb{E}(f_+(X)) - \mathbb{E}(f_-(X)) \overset{(iii)}{=} \mathbb{E}_{\mathbb{P}_X}(f_+) - \mathbb{E}_{\mathbb{P}_X}(f_-) = \mathbb{E}_{\mathbb{P}_X}(f)$$

$\square$

Bemerkung 4.18. Sei $H : [a,b] \to \mathbb{R}$ streng monoton wachsend und $h \in C^1$. Sei $\Omega = [a,b], \mathscr{F}\mathscr{B}([a,b]), \mu(dx) = h'(x)\ dx$.
Dann ist $\mu([a,b]) = \int_a^b h'(x)\ dx = h(b) - h(a)$ das Bildmaß von μ unter h.

$$\begin{aligned}\mu \circ h^{-1}\big((-\infty, x]\big) &= \mu\big(\{y \in [a,b] : -\infty \leq h(y) \leq x\}\big) = \mu\big([a,b] \cap (-\infty, h^{-1}(x))\big) \\ &= \int_a^{h^{-1}(x)} h'(y)\ dy = [h(y)]_a^{h^{-1}(x)} = h(h^{-1}(x)) - h(a) = x - h(a)\end{aligned}$$

falls $h(a) \leq x \leq h(b)$. Dann ist

$$\mu \circ h^{-1} = \lambda_{[h(a), h[b]]}$$

Dann folgt aus 4.17

$$\int_a^b f(h(x))h'(x)\ dx = \mathbb{E}_\mu(f(h)) = \mathbb{E}_{\mathbb{P}_n}(f) = \int_{h(a)}^{h(b)} f(y)\ dy$$

Korollar 4.19. *Seien $X : \Omega_1 \to \Omega'$, $Y : \Omega_2 \to \Omega'$ Zufallsvariablen mit $\mathbb{P}_X = \mathbb{P}_Y$. Dann gilt $\mathbb{E}(f(x)) = \mathbb{E}(f(Y))$ für alle $f : \Omega' \to \mathbb{R}$ mit $\mathbb{E}(|f(x)|) < \infty$ oder $f \geq 0$.*
Umgekehrt: sei $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$ für alle f messbar, beschränkt, dann gilt $\mathbb{P}_X = \mathbb{P}_Y$.

Beweis. **„$\Rightarrow$“** Folgt aus 4.17.

„$\Leftarrow$“ $\mathbb{P}_X(A) = \mathbb{E}(\mathbb{1}_A(X)) = \mathbb{E}(\mathbb{1}_A(X)) = \mathbb{P}_Y(A)$.

$\square$

Beispiel 4.20.1. Sei $X \sim \operatorname{Exp}(\alpha)$, d.h. $\mathbb{P}_X(dx) = \alpha e^{-\alpha x}\ dx$. Dann ist

$$\mathbb{E}(X^k) = \alpha \int_0^\infty x^k e^{-\alpha x}\ dx = \frac{k!}{\alpha^k}$$

Beispiel 4.20.2. Sei $X \sim \mathscr{N}(m, v)$, dann ist $\mathbb{E}(X) = m$ und $\mathbb{E}\big((X-m)^2\big) = v$

Korollar 4.21. *a) Sei X reelle Zufallsvariable. $\mathbb{P}_X$ habe dichte ρ bezüglich des Lebesgue-Maß. Dann ist*

$$\mathbb{E}(f(X)) = \int_{-\infty}^\infty \rho(x) f(x)\ dx$$

(falls $\int \rho(x)|f(x)|\ dx < \infty$). Inesbesondere

$$\mathbb{E}(X) = \int_{-\infty}^\infty x\rho(x)\ dx$$

b) Sei $\Omega = \mathbb{R}$, $\mathbb{P}(dy) = \rho(y)\ dy$ (ρ ist Dichte von $\mathbb{P}$).
Sei X reelle Zufallsvariable . Dann gilt:

$$\mathbb{E}(f(X)) = \int_{-\infty}^{\infty} f(X(y))\rho(y)\ dy$$

Insbesondere $\mathbb{E}(X = \int_{-\infty}^{\infty} X(y)\rho(y)\ dy$

Definition 4.22. Sei X reelle Zufallsvariable

a) Falls $|X^n| \in \mathscr{L}^1$ für $n \in \mathbb{N}$, so heißst $\mathbb{E}(X^n)$ das **n-te Moment von X**.

b) Falls $|X^p| \in \mathscr{L}^1$, für $p > 0$, so schreibt man $X \in \mathscr{L}^p$.

c) Die Zahl $\mathbb{V}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$ heißt die **Varianz** von X (bezüglich $\mathbb{P}$).
$\sqrt{\mathbb{V}(X)}$ heißt **Standardabweichung** von X.

Bemerkung 4.23. a) Für die Varianz gilt

$$\mathbb{V}(X) = \mathbb{E}(X^2 - E\mathbb{E}(X)X + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

b) Da $|X|^p \leq 1 + |X|^q$, falls $p \leq q$ ist

$$\mathscr{L}^q \leq \mathscr{L}^p$$

falls $p \leq q$.

Definition 4.24. Sei $X, Y \in \mathscr{L}^2$, dann heißt

$$\mathrm{Cov}(X, Y) := \mathbb{E}\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

die **Kovarianz** von X und Y.
Man nennt X und Y

positiv korreliert falls $\mathrm{Cov}(X, Y) > 0$

unkorreliert falls $\mathrm{Cov}(X, Y) = 0$

negtaiv korreliert falls $\mathrm{Cov}(X, Y) < 0$

Beispiel 4.25. Sei $\Omega = [0, 1]$ mit Lebesgue-Maß, sei $X(\omega) = \omega$,$Y(\omega) = \omega^2$ und $Z(\omega) = \frac{1}{2}\sin(\pi\omega)$. Dann ist

$$\mathbb{E}(X) = \int_0^1 d\ dx = \frac{1}{2}$$
$$\mathbb{E}(Y) = \int_0^1 x^2\ dx = \frac{1}{3}$$
$$\mathbb{E}(Z) = \int_0^1 \frac{\sin(\pi\omega)}{2}\ dx = \frac{1}{2\pi}$$

Dann folgt für die Kovarianz:

$$\mathrm{Cov}(X, Y) = \mathbb{E}(\left(X - \frac{1}{2}\right)\left(Y - \frac{1}{3}\right)) = \frac{1}{12}$$

$\mathrm{Cov}(X, Z) = \int_0^1 \left(x - \frac{1}{2}\right)\left(\sin(\pi x) - \frac{1}{2\pi}\right)\ dx = 0$Zum Vergleich: $\mathrm{Cov}(X, X) \quad = \mathbb{V}(X) = \frac{1}{12}$

Daher eigenen sich die Korrelationskoeffizienten besser al Maß

Definition 4.26. Für $X, Y \in \mathscr{L}^2$ heißt

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

Korrelationskoeffizient von X und Y.

Beispiel 4.27. In Beispiel4.25 gilt

$$\rho(X,Y) = \frac{\sqrt{15}}{4} \approx 0.968$$
$$\rho(X,X) = 1$$

Satz 4.28 (Ungleichungen für Integrale)**.** *a)* ***Jensen-Ungleichung*** *Sei X reelle Zufallsvariable, $\phi(\mathbb{R} \to \mathbb{R})$ sei konvex.*
Sei zusätzlich noch $\mathbb{E}(|X|) < \infty$ und $\mathbb{E}(|\phi(X)|) < \infty$.
Dann gilt

$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}(X))$$

d) ***Chebyshev-Markov-Ungleichung*** *Sei X reelle Zufallsvariable, $f : [0,\infty) \to [0,\infty)$ sei monoton wachsend. Dann gilt $\forall c > 0$ mit $f(c) > 0$, dass*

$$\mathbb{P}(|X| \geq c) \leq \frac{\mathbb{E}(f(|X|))}{f(c)}$$

Insbesondere gilt für $f(x) = x^2$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) \leq \frac{\mathbb{V}(X)}{c^2}$$

b) ***Hölder Ungleichung*** *Seine X, Y reelle Zufallsvariablen. Dann gilt für alle $p, q \in [1, \infty]$ mit $\frac{1}{p} + \frac{1}{q} = 1$, dass*

$$\mathbb{E}(|XY|) \leq \|X\|_p \|Y\|_q$$

dabei ist

$$\|A\|_p := \begin{cases} \mathbb{E}(|X|^p)^{\frac{1}{p}} & p \leq \infty \\ \sup\{|X(\omega)| : \omega \in \Omega\} & p = \infty \end{cases}$$

c) ***Cauchy-Schwarz-Ungleichung*** *Seien $X, Y \in \mathscr{L}^2$. Dann ist $XY \in \mathscr{L}^2$ und*

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

Beweis. a) Da ϕ konvex ist, ist per Definition $A := \{(x,y) \in \mathbb{R}^2 : y \geq \phi(x)\}$ eine konvexe Menge.
$\Leftrightarrow \forall x \in \mathbb{R}$ gibt es eine Gerade durch $(x, \phi(X))$, die das Innere A^0 von A nicht schneidet.
$\Leftrightarrow \forall x \in \mathbb{R} \exists a \in \mathbb{R} : \ell(y) := \phi(x) + a(y - x) \leq \phi(y)$.
Wähle nun $x = \mathbb{E}(X)$. Dann ist

$$\phi\big(X(\omega)\big) \geq \ell\big(X(\omega)\big) = \phi\big(\mathbb{E}(X)\big) + a\big(X(\omega) - \mathbb{E}(X)\big)$$

Da der Erwartungswert monoton ist erhält anwenden des Erwartungswerts auf beide Seiten die Ungleichung:

$$\mathbb{E}\big(\phi(X)\big) \geq \mathbb{E}\big(\phi(\mathbb{E}(X)) + a(X - \mathbb{E}(X))\big) = \mathbb{E}\big(\phi(\mathbb{E}(X))\big) + a\mathbb{E}\big(X - \mathbb{E}(X)\big) = \phi\big(\mathbb{E}(X)\big)$$

d) Es gilt

$$\begin{aligned}
\mathbb{P}(|X| \geq c) &= \mathbb{P}(f(|X|) \geq f(c)) \\
&= \mathbb{P}\left(\frac{f(|X|)}{f(c)} \geq 1\right) = \mathbb{E}\left(\mathbb{1}_{\left\{\frac{f(|X|)}{f(c)} \geq 1\right\}}\right) \\
&\leq \mathbb{E}\left(\mathbb{1}_{\left\{\frac{f(|X|)}{f(c)} \geq 1\right\}} \frac{f(|X|)}{f(c)}\right) \\
&\leq \mathbb{E}\left(\frac{f(|X|)}{f(c)}\right) = \frac{1}{f(c)} \mathbb{E}(f(|X|))
\end{aligned}$$

b) Für $p = \infty$ gilt $\mathbb{E}(|XY|) \leq \mathbb{E}(\|X\|_\infty |Y|) = \|X\|_\infty \|Y\|_1$.
Für $p < \infty$: Falls $\|X\|_p = 0$, dann folgt für $c > 0$ aus d), dass $\mathbb{P}(|X| > c) \leq \|X\|_p^p (c^p) = 0$. Somit its

$$\mathbb{P}(|X| > 0) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} \{\omega : X(\omega) > \frac{1}{n}\}\right) \overset{??}{=} \lim_{n \to \infty} \mathbb{P}(X \geq 1/n) = 0$$

Es folgt, dass $\mathbb{E}(|XY|) = \mathbb{E}(|XY|\mathbb{1}_{\{|X|>0\}}) = 0$, sodass die Ungleichung gilt (als Gleichung).
Sei $\|X\|_p > 0$. Da in der Behauptung nur Beträge vorkommen können wir o.b.d.A. annehmen, dass $X > 0$ und $Y \geq 0$ ist. Sei nun das Wahrscheinlichkeitsmaß $\tilde{P}$ auf Ω definiert als

$$\tilde{P}(A) := \mathbb{P}\left(\mathbb{1}_A \frac{X^p}{\mathbb{E}(X^p)}\right).$$

Sei außerdem $Z := Y/X^{p-1}$. Dann ist

$$\mathbb{E}(XY) = \mathbb{E}(X^p Z) = \tilde{\mathbb{E}}\big(Z\mathbb{E}(X^p)\big) = \tilde{\mathbb{E}}(Z)\mathbb{E}(X^p) \tag{$\star$}$$

Anwenden der Jensen Ungleichung a) mit der Funktion $\phi(x) = x^q \mathbb{1}_{\{q>0\}}$ ergibt

$$\begin{aligned}
(\star) \leq \tilde{\mathbb{E}}(Z^q)^{\frac{1}{q}} \mathbb{E}(X^p) &= \mathbb{E}\left(\frac{Y^q}{X^{q(p-1)} \frac{X^p}{\mathbb{E}(X^p)}}\right) \mathbb{E}(X^p) \\
&= \mathbb{E}(Y^q X^{p-q(p-1)})^{\frac{1}{q}} \mathbb{E}(X^p)^{1-\frac{1}{q}} = \mathbb{E}(Y^q)^{\frac{1}{q}} \mathbb{E}(X^p)^{\frac{1}{p}}
\end{aligned}$$

(Da p, q konjugierte Exponenten sind ist $p - pq + q = 0$.)

c) Spezialfall von b) mit $p = q = 2$

□

Bemerkung 4.29.
- Sei $X, Y \in \mathscr{L}^2$, dann gilt mit 4.28 c), dass auch$X + \alpha Y \in \mathscr{L}^2$. Also ist $\mathscr{L}^2$ ein Vektorraum und $\|X\|_2 = \mathbb{E}(|X|^2)^{\frac{a}{2}}$ ist eine Halbnorm. Dann ist $\mathbb{V}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big)$ die „Länge des optimal Verschobenen Vektors“:

$$\mathbb{V}(X) = \min\{\mathbb{E}\big((X - a)^2\big) : r \in \mathbb{R}\}$$

und

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left(\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\right)$$

ist ein Skalarprodukt der beiden optimal verschobenen Vektoren.

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}$$

gibt den Winkel zwischen optimal verschobenen Vektoren an.

Satz 4.30. *a)* $\operatorname{Cov}(X+a, Y+b) = \operatorname{Cov}(X,Y)$ *für alle* $a, b \in \mathbb{R}$.

b) $\operatorname{Cov}(aX, bY) = ab \operatorname{Cov}(X,Y)$

c) $\mathbb{V}(aX) = a^2\mathbb{V}(X)$

d) Es ist $\sum_{i=1}^n X_i \in \mathscr{L}^2$ *und*

$$\mathbb{V}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{V}(X_i) + \sum_{\substack{i,j=0\\ i\neq j}}^n \operatorname{Cov}(X_i, X_j) \qquad \text{(Binomische Formel)}$$

e) Falls (X_i) *paarweise unkeorreliert sind, dann*

$$\mathbb{V}\left(\sum_{i=1}^\infty X_i\right) = \sum_{i=1}^\infty \mathbb{V}(X_i) \quad \text{(„Gleichheit von Bienarme“/„Pythagoras“)}$$

Satz 4.31. *Seien* $X, Y \in \mathscr{L}^1$ *und* $X \perp\!\!\!\perp Y$*, dann gilt*

a) $XY \in \mathscr{L}^1$ *und* $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

b) $\operatorname{Cov}(X,Y) = 0$

Beweis. a) Da $X \perp\!\!\!\perp Y$ ist folgt mit **??**a), dass das Bildmaß von $\mathbb{P}$ unter $(X,Y) : \Omega \to \mathbb{R}^2$ ein Produktmaß, also $\mathbb{P}_X \otimes \mathbb{P}_Y$ ist.
Nach **??**a)]0417satzist

$$\mathbb{E}\big(|X||Y|\big) = \int |x||y|\mathbb{P}_{(XY)}\ (dx\ dy) = \int_{\mathbb{R}^2} |x||y|\mathbb{P}_X(dx)\mathbb{P}_Y(dy)$$

mit $h : \mathbb{R}^2 \to R, (x,y) \mapsto |x|y|$. Dann folgt mit dem Satz von Fubini

$$= \int_{\mathbb{R}} |x|\mathbb{P}_X(dx) \int_{\mathbb{R}} |y|\mathbb{P}_Y(dy) = \mathbb{E}(|X|)\mathbb{E}(|Y|)$$

Dann folgt dass $XY \in \mathscr{L}^1$ ist. Also darf die gleiche Rechnung auch ohne $|\cdot|$ durchgeführt werden.

b) Es gilt

$$\operatorname{Cov}(X,Y) = \mathbb{E}\left(\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\right) \overset{a)}{=} \mathbb{E}\big(X - \mathbb{E}(X)\big)\mathbb{E}\big(Y - \mathbb{E}(Y)\big) = 0$$

□

Bemerkung 4.32. a) **??**e)]0430rechenregeln gilt also auch wenn $X_i \perp\!\!\!\perp X_j$ für alle $i \neq j$.

b) Es kann aber $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ gelten, ohne dass X, Y unabhängig sind.

Beispiel. Sei $\Omega = [0,1]$, $\mathbb{P} = \frac{1}{2}\lambda_{[-1,1]}$, $X(x) = y$, $Y(x) = x^2$. Dann ist

$$\mathbb{E}(XY) = \frac{1}{2}\int_{-1}^{1} x^3 \; dx = 0 = \mathbb{E}(X) = \mathbb{E}(X)\mathbb{E}(Y)$$

$$\mathbb{E}(X^2Y) = \frac{1}{2}\int_{-1}^{1} x^4 \; dx = \frac{1}{5} \quad \text{aber} \quad \mathbb{E}(\mathbb{E}(X^2))\mathbb{E}(Y) = \left(\frac{1}{3}\right)^3 = \frac{1}{9}$$

Wären $X \perp\!\!\!\perp Y$, dann müsste auch $X^2 \perp\!\!\!\perp Y$ sein, also muss $X \not\perp\!\!\!\perp Y$.

Bemerkung. Orthogonalität in $\mathscr{L}^2$ ist viel schwächer (und weniger stabil) als Unabhängigkeit.

4.A Erzeugende Funktionen

Für viele Wahrscheinlichkeitsmaße auf $\mathbb{N}_0$ besteht die Zähldichte aus den Taylor Koeffizienten einer Funktion entwickelt bei $x = 0$.

Beispiel 4.33. Poisson Verteilung: $e^{-\lambda}\frac{\lambda^k}{k!}$. Dann ist

$$e^{\lambda u - \lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k u^k}{k!}$$

Geometrische Verteilung: $(1-p)^{k-1}p$. Dann ist

$$\sum_{k=0}^{\infty}(1-p)^{k-1}pu^k = \frac{p}{1-p}\sum_{k=1}^{\infty}\big((1-p)\big)^k = \frac{p}{1-p}\frac{1}{1-(1-p)u}$$

Definition 4.34. Sei $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $\mathbb{N}_0$ mit Zähldichte $(p_k)_{k\geq 0}$. Die Funktion

$$\varphi_{\mathbb{P}} : [0,1] \to \mathbb{R} \quad \varphi_{\mathbb{P}}\mathbb{P}(u) = \sum_{k=0}^{\infty} p_k u^k$$

heißt **Erzeugende Funktion** von $\mathbb{P}$ (oder der Folge (p_k)).

Bemerkung 4.35. a) Da $\sum_{k=0}^{\infty} p_k = 1$ ist der Konvergenzradius von ϕ größer gleich als 1.

b) Es gilt

$$\varphi_{\mathscr{B}_{n,p}}(u) = \sum_{k=0}^{n} p^k(1-p)^{n-k}\binom{n}{k}u^k = \big(pu + (1-p)\big)^n$$

Definition 4.36. Sei X Zufallsvariable mit Werten in N_0. Dann heißt

$$\varphi_X(u) = \mathbb{E}(u^X) = \sum_{k=0}^{\infty}\mathbb{P}(X = k)u^k$$

erzeugende Funktion von X.

Satz 4.37. *Sei X eine N_0-wertige Zufallsvariable. Dann*

a) $\mathbb{P}(X = k) = \frac{1}{k!}\, \frac{d^k}{du^k}\varphi_X(u)\Big|_{u=0}$ *für alle* $k \in \mathbb{N}$.

b) $\mathbb{E}(X)$ *existiert* $\Leftrightarrow \lim\limits_{u\uparrow 1} \phi'(u)$ *existiert. Dann ist* $\mathbb{E}(X) = \varphi'_X(1)$.

c) Für alle $p \in \mathbb{N}$ *ist* $X \in \mathscr{L}^p \Leftrightarrow \lim\limits_{u\uparrow 1} \frac{d^p}{du^p}\varphi(u)$ *existiert. Dann ist*

$$\frac{d^p}{du^p}\varphi_X(1) = \mathbb{E}\big(X(X-1)\cdot \ldots \cdot (X-p+1)\big)$$

Beweis. a) Satz von Taylor: Für alle $u \leq 1$(Konvergenzradius)

$$\partial_u^k \varphi(u) = \sum_{n=0}^{\infty} p_n(\partial^n d^k) = \sum_{n=k}^{\infty} p_n \frac{n!}{(n-k)!} u^{k-n}$$

Also gilt bei Auswertung in 0

$$\frac{1}{k!}\, \partial_n^k \varphi(n)\big|_{u=0} = p_k = \mathbb{P}(X = k)$$

b) Für $u < 1$ konvergiert die Reihe absolut

$$\sum_{k=0}^{\infty} k u^k \mathbb{P}(X = k) = u \sum_{k=0}^{\infty} (\partial_u u^k)\mathbb{P}(X = k)$$

also darf man gliedweise Differenzieren

$$= u\partial u \sum_{k=0}^{\infty} u^k \mathbb{P}(X = k) = u\varphi'_X(u)$$

Beide Seiten wachsen Monoton, wenn $u \uparrow 1$:

$$\text{Linke Seite} \underset{u\uparrow 1}{\uparrow} \mathbb{E}(X) \quad \text{Rechte Seite} \underset{u\uparrow 1}{\uparrow} \varphi'(1)$$

c) Anwenden des gleichen Arguments mit

$$\sum_{k\geq p} k(k-1)\ldots(k-p+1)u^k \mathbb{P}(X = k) = u^p \sum_{k=0}^{\infty} \partial_u^p (u^k)\mathbb{P}(X = k)$$

gegen $\mathbb{E}\big(X(X-1)\ldots(X-p+1)\big)$ bzw. $1^p \partial_u^p \varphi_X(1)$.

□

Beispiel 4.38. Sei $X \sim \mathrm{Poi}_\lambda$: Dann ist

$$\mathbb{E}(X) = \partial_u \left(e^{\lambda u - \lambda}\right)\big|_{u=1} = \lambda e^0 = \lambda$$

$$\mathbb{E}\big(X(X-1)\big) = \partial_u^2 \left(e^{\lambda u - \lambda}\right) = \lambda^2$$

Also gilt für die Varianz:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}\big(X(X-1) + \mathbb{E}(X) - \mathbb{E}(X)^2\big) = \lambda$$

Satz 4.39. *Seien X, Y unabhängige $\mathbb{N}_0$-verteilte Zufallsvariablen. Dann gilt*

$$\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u)$$

Beweis. Sei $X \sim \mathrm{Poi}_\lambda$, $Y \sim \mathrm{Poi}_\mu$, $Y \perp\!\!\!\perp X$. Dann ist

$$\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u) = e^{\lambda u - \lambda} e^{\mu u - \mu} = e^{(\mu+\lambda)u - (\mu+\lambda)}$$

Also $X + Y \sim \mathrm{Poi}_{\lambda+\mu}$. $\square$

Bemerkung 4.40 (Ausblick). a) Die rechte Seite in 4.37c) ist etwas unhandlich. Daher definiere für $t > 0$

$$L_X(t) := \varphi_X(e^{-t}) = \mathbb{E}\big(\left(e^{-t}\right)^X \big) = \mathbb{E}\left(e^{-t}\right)$$

(„Laplace Transformierte von $\mathbb{P}_X$")

Dann ist

$$\mathbb{E}(X^p) = \frac{d^p}{dt^p} L_X \bigg|_{t=0}$$

falls $\mathbb{E}(X^p) < \infty$.

b) Das geht auch für reelle Zufallsvariablen. Sei $X : \Omega \to \mathbb{R}$. Es sei $\mathbb{E}\left(e^{-tX}\right) < \infty$ für ein $c > 0$. Dann ist

$$L_X(t) = \mathbb{E}(e^{-tX}) = \int_{-\infty}^{\infty} e^{-t} \mathbb{P}_X(dy)$$

für alle $t \in \{\tilde{t} \in \mathbb{R} \mid \mathbb{E}(e^{-\tilde{t}X}) < \infty\}$.
Wieder gilt

$$\mathbb{E}(X^p) = (-1)^p \, \partial_t^p L_X(t)\big|_{t=0}$$

und es gilt auch L_X bestimmt P_X eindeutig.
Satz 4.39 gilt mit $L_{X+Y} = L_X L_Y$. Außerdem ist $L_{\alpha X}(t) = \mathbb{E}\left(e^{-t\alpha X}\right) = L_X(\alpha t)$, falls alles $< \infty$.

c) Für unabhängige reelle Zufallsvariablen X_i mit $\mathbb{P}_{X_i} = \mathbb{P}_{X_j}$ gilt

$$(**) = L_{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_j}(t) = \prod_{i=1}^n L_{\frac{X_i}{\sqrt{n}}}(t) = \prod_{i=1}^n L_{X_i}\left(\frac{t}{\sqrt{n}}\right) = \left(L_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Es gilt immer $L_X(0) = 1$ und $L_X'(0) = \mathbb{E}(X)$.
Angenommen $\mathbb{E}(X) = 0$ und $L_X''(0) = \mathbb{E}(X^2) < \infty$, dann gilt die Taylor-Gleichung:

$$\ln(L_X(s)) = \ln\left(L_X(0) + sL_X'(0) + \mathscr{O}(s^3)\right) = \frac{1}{2}s^2 L_X''(0) + \mathscr{O}(s^3)$$

Es folgt, dass

$$(**) = e^{n \ln L_X\left(\frac{t}{\sqrt{n}}\right)} = e^{n\left(\frac{1}{2}\frac{t^2}{n} L_X''(0) + \mathscr{O}\left(\frac{t^3}{n^3}\right)\right)} \xrightarrow{n\to\infty} e^{\frac{t^2}{2}\mathbb{E}(X^2)}$$

5 Grenzwertsätze

Wir haben $\mathbb{E}(X) = \sum_{k\in\mathbb{Z}} k\mathbb{P}(X = k)$ über die relative Häufigkeit motiviert:

$$\mathbb{E}(X) \approx \frac{1}{N}\sum_{i=1}^{N} X_i(\omega) \approx \sum_{k\in\mathbb{Z}} \frac{|\{i : X_i(\omega = k)\}|}{N} k$$

Wir werden nu begründen, dass dies Intuition richtig ist.

Definition 5.1. Eine Folge (X_n) von Zufallsvariablen heißt **unabhängig identisch Verteilt (uiv)**, wenn

(i) die X_n sind paarweise unabhängig

(ii) $\mathbb{P}_{X_i} = \mathbb{P}_{X_j}$ für alle $i, j \in \mathbb{N}$.

Satz 5.2 (Schwache Gesetz der großen Zahlen (GGZ), Basisversion)**.** *Seien* (X_i) *uiv, reelle Zufallsvariable mit* $\mathbb{V}(X_i) < \infty$.
Dann ist für alle $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| \geq \epsilon\right) = 0$$

Beweis. Setze $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, dann ist

$$\mathbb{E}(Y_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \mathbb{E}(X_1)$$

$$\mathbb{V}(Y_n) = \mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \stackrel{4.30}{=} \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(X_i) = \frac{1}{n}\mathbb{V}(X_1)$$

Mit 4.28d) (Chebyshev-Markov-Ungleichung) folgt

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| > \epsilon\right) = \mathbb{P}\left(|Y_n - \mathbb{E}(Y_n)| > \epsilon\right) \leq \frac{1}{\epsilon^2}\mathbb{V}(Y_n) = \mathbb{V}(X_1)\frac{1}{n\epsilon^2} \xrightarrow{n\to\infty} 0$$

□

Satz 5.3 (Schwache Gesetz der großen Zahlen, $\mathscr{L}^2$-Version)**.** *Seien* (X_i) *(nicht notwendigerweise uiv) reelle Zufallsvariablen für die gilt:*

(i) ***Paarweise unkorreliert*** $\mathrm{Cov}(X_i, X_j) = 0$ *für alle* $i \neq j$.

(ii) ***Beschränkte Varianzen*** $v := \sup\{\mathbb{V}(X_i), i \in \mathbb{N}\} < \infty$

Dann gilt für alle $n \in \mathbb{N}$ *und alle* $\epsilon_n > 0$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} \big(X_i - \mathbb{E}(X_i)\big)\right| \geq \epsilon_n\right) \leq \frac{v}{n\epsilon_n^2}$$

Beweis. Analog, Übung □

Satz 5.4 (Schwache Gesetz der großen Zahlen, $\mathscr{L}^1$-Version)**.** *Seien* (X_i) *paarweise undabhängige, identisch Verteilte Zufallsvariablen und* $\mathbb{E}(|X_1|) < \infty$. *Dann gilt für alle* $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}(X_1)\right| \geq \epsilon\right) = 0$$

Beweis. Zerlege

$$X_i(\omega) = \underbrace{X_i(\omega)\mathbb{1}_{\{|X_i(\omega)|<\sqrt[4]{i}\}}(\omega)}_{Y_i(\omega)} + \underbrace{X_i(\omega)\mathbb{1}_{\{|X_i(\omega)|\geq\sqrt[4]{i}\}}(\omega)}_{Z_i(\omega)}$$

Dann gilt

a) $\mathbb{E}(X_i) = \mathbb{E}(Y_i) + \mathbb{E}(Z_i)$

b) mit der Dreiecksungleichung

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) - \mathbb{E}(X_1)\right| &= \frac{1}{n}\left|\sum_{i=1}^{n}(Y_i(\omega) - \mathbb{E}(Y_i) + Z_i(\omega) - \mathbb{E}(Z_i))\right| \\ &\leq \frac{1}{n}\left|\sum_{i=1}^{n}(Y_i(\omega) - \mathbb{E}(Y_i))\right| + \frac{1}{n}\left|\sum_{i=1}^{n}(Z_i(\omega) - \mathbb{E}(Z_i))\right|\end{aligned}$$

Daher ist

$$\begin{aligned}\underbrace{\left\{\omega\in\Omega : \left|\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) - \mathbb{E}(X_i)\right| \geq \epsilon\right\}}_{A_n} \subset &\underbrace{\left\{\omega\in\Omega : \left|\frac{1}{n}\sum_{i=1}^{n} Y_i(\omega) - \mathbb{E}(Y_i)\right| \geq \frac{\epsilon}{2}\right\}}_{B_n} \\ &\bigcup \underbrace{\left\{\omega\in\Omega : \left|\frac{1}{n}\sum_{i=1}^{n} Z_i(\omega) - \mathbb{E}(Z_i)\right| \geq \frac{\epsilon}{2}\right\}}_{C_n}\end{aligned}$$

und damit $\mathbb{P}(A_n) \leq \mathbb{P}(B_n) + \mathbb{P}(C_n)$.
Wegen **??**a)]0318satz sind auch die Y_i paarweise unabhängig. Sei nun $\overline{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$, dann ist

$$\begin{aligned}\mathbb{P}(B_n) = \mathbb{P}\left(\left|\overline{Y}_n - \mathbb{E}(\overline{Y}_n)\right| \geq \frac{\epsilon}{2}\right) &\overset{4.28d)}{\leq} \frac{4}{\epsilon^2}\mathbb{V}(\overline{Y}_n) = \frac{4}{\epsilon^2}\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(Y_i) \\ &\leq \frac{4}{n^2\epsilon^2}\sum_{i=1}^{n}\mathbb{E}(Y_i^2) \leq \frac{4}{n^2\epsilon^2}\sum_{i=1}^{n}\sqrt{i} \leq \frac{4}{\epsilon^2 n^2}n\sqrt{n} = 4\frac{\epsilon^2}{\sqrt{n}} \xrightarrow{n\to\infty} 0\end{aligned}$$

Analog für C_n mit Sei $\overline{Z_n} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. Dann ist

$$\begin{aligned}\mathbb{P}(C_n) = \mathbb{P}\left(\left|\overline{Z_n} - \mathbb{E}(\overline{Z_n})\right| \geq \frac{\epsilon}{2}\right) &\overset{\substack{4.28d) \\ f(x)=|x|}}{\leq} \frac{2}{\epsilon}\mathbb{E}\left(\left|\overline{Z}_n - \mathbb{E}(\overline{Z}_n)\right|\right) = \frac{2}{n\epsilon}\mathbb{E}\left(\left|\sum_{i=1}^{n} Z_i - \mathbb{E}(Z_i)\right|\right) \\ &\leq \frac{2}{n\epsilon}\sum_{i=1}^{n}\mathbb{E}\left(|Z_i - \mathbb{E}(Z_i)|\right) \leq \frac{2}{n\epsilon}\sum_{i=1}^{n} 2\mathbb{E}(Z_i)\end{aligned}$$

Da aber falls $a_i \xrightarrow{i\to\infty} 0$ geht auch $\frac{1}{n}\sum_{i=1}^n a_i \xrightarrow{in\to\infty}$. Also

$$\mathbb{P}(C_n) \leq \frac{2}{n\epsilon}\sum_{i=1}^n 2\mathbb{E}(Z_i) \xrightarrow{n\to\infty} 0$$

□

Beispiel 5.5 (Satz von Weierstraß). Sei $f : [0,1] \to \mathbb{R}$ stetig.
Zu $p \in [0,1]$ seien $(X_i)_{i\in\mathbb{N}}$ uiv Zufallsvariablen mit $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$. Dann ist $\sum_{i=1}^n X_i \sim \mathscr{B}_{n,p}$ und

$$f_n(p) := \mathbb{E}_p\left(f\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right) = \sum_{k=0}^n f\left(\frac{k}{n}\right)\mathbb{P}_p\left(\sum_{i=1}^n X_i = k\right) = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k}p^k(1-p)^{n-k}$$

ist Polynom n-ten Grades in p („Bernstein-Polynome“).
f ist gleichmäßig stetig in $[0,1]$, also $\forall\epsilon > 0 \exists\delta > 0$, sodass für alle $|x-y| < \delta$ immer $|f(x) - f(y)| < \epsilon$ gilt.
Außerdem ist

$$f_n(p) - f(p) = \mathbb{E}_p\left(f\left(\frac{1}{n}\sum_{i=1}^n X_i\right) - f(p)\right)$$

Also kann man abschätzen:

$$\begin{aligned}|f_n(p) - f(p)| &\leq \left|\mathbb{E}_p\left(f\left(\frac{1}{n}\sum_{i=1}^\infty\right) - f(p)\right)\right| \leq \mathbb{E}_p\left(\left|f\left(\frac{1}{n}\sum_{i=1}^\infty\right) - f_p\right|\right)\\ &\leq \epsilon + 2\|f\|_\infty \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - p\right| \geq \delta\right) \overset{5.3}{\leq} \epsilon + 2\|f\|_\infty \frac{\mathbb{V}(X_i)}{n\delta^2}\end{aligned}$$

Also ist $\limsup_{n\to\infty}|f_n(p) - f(p)| \leq \epsilon$ für alle $\epsilon > 0$.
Also ist $\lim\limits_{n\to\infty}|f_n(p) - f(p)| = 0$.

Bemerkung.

Beispiel 5.6. Sei $\Omega = [0,1)$ mit dem Lebesgue-Maß, Sei $X_i(\omega) = \mathbb{1}_{I_i}(\omega)$. für $i = 1, ..., 15$ mit

$$\begin{aligned}I_1 &= [0, \tfrac{1}{2}) & & & I_1 &= [\tfrac{1}{2}, 1)\\ I_3 &= [0, \tfrac{1}{4}) & &\cdots & I_6 &= [\tfrac{3}{4}, 1)\\ I_7 &= [0, \tfrac{1}{8}) & &\cdots & I_{15} &= [\tfrac{7}{8}, 1)\end{aligned}$$

Dann ist

$$\mathbb{P}(|X_n - 0| > \epsilon) = \begin{cases}\frac{1}{2} & n = 1,2\\ \frac{1}{4} & n = 3,4,5,6\\ \frac{1}{8} & n = 7,...,15\end{cases} \quad \leq \frac{2}{n} \xrightarrow{n\to\infty} 0$$

Aber $\limsup_{n\to\infty}|X_n(\omega) - 0| = 1\ \forall\omega\in\Omega$.
Also $X_n(\omega) \overset{n\to\infty}{\not\longrightarrow} 0$.

Definition 5.7. Seien (S_n) Zufallsvariablen. Man sagt für $c \in \mathbb{R}$, dass S_n konvergiert $\mathbb{P}$ fast sicher ($S_n \xrightarrow{n\to\infty} c$), wenn

$$\mathbb{P}\left(\lim_{n\to\infty} S_n = c\right) = 1 \qquad \text{(Gl. 5.2)}$$

Bemerkung 5.8. Setze $\Omega_m := \{\omega \in \Omega : \limsup_{n\to\infty} |S - n - c| < \frac{1}{m}\}$.
Da $\Omega_{m+1} \subseteq \Omega_m$, also

$$\Omega_\infty := \left\{\omega \in \Omega : \limsup_{n\to\infty} S_n = c\right\}$$

mit **??**e). Dann gilt

$$\mathbb{P}\left(\limsup_{n\to\infty} S_n = c\right) = 1 \Leftrightarrow \forall \epsilon >=: \mathbb{P}\left(\limsup_{n\to\infty} |S_n - c| < \epsilon\right) = 1 \qquad (\star\star\star)$$

Daher besteht der Unterschied zwischen **??** und Gl. 5.2 in der Platzierung der Grenzwerts: $\forall \epsilon, c > 0$

- **??** besagt $\lim\limits_{n\to\infty} \mathbb{P}\left(|S_n - c| < \epsilon\right) = 1$
- Gl. 5.2 besagt $\mathbb{P}\left(\limsup_{n\to\infty} |S_n - c| < \epsilon\right) = 1$

Außerdem gilt das Lemma von Fatou 4.15: $\mathbb{E}\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$, falls $X_n \geq 0$.
Setze nun $X_n = \mathbb{1}_{\{|S_n - c| < \epsilon\}}$. Dann ist

$$\begin{aligned}\liminf_{n\to\infty}(X_n\omega) = 1 &\Leftrightarrow \exists m \in \mathbb{N} : \forall n > m \mathbb{1}_{\{|S_n - c| < \epsilon\}}(\omega) = 1 \\ &\Leftrightarrow |S_n - c| > \epsilon \text{ unendlich oft} \Leftrightarrow \limsup_{n\to\infty} |S_n - c| < \epsilon\end{aligned}$$

Dann folgt (Fatou)

$$\liminf_{n\to\infty} \mathbb{P}(|S_n - c| < \epsilon) = \liminf_{n\to\infty} \mathbb{E}(X_n) \geq \mathbb{E}\left(\liminf_{n\to\infty} X_n\right) = \mathbb{E}\left(\limsup_{n\to\infty} |S_n - c| < \epsilon\right)$$

Daher folgt aus $\mathbb{P}\left(\lim\limits_{n\to\infty} |S_n - c| = 0\right)$ mit $\star\star\star$, dass $\forall \epsilon > 0 . \mathbb{P}\left(\limsup_{n\to\infty} |S_n - c| < \epsilon\right)$.
Dann folgt mit Fatou: $\forall \epsilon > 0 \lim\limits_{n\to\infty} \mathbb{P}\left(|S_n - c| < \epsilon\right) = 1$.
Also ist Gl. 5.2 stärker als Gl. 5.2.

Satz 5.9 (Starkes Gesetz der großen Zahlen, $\mathscr{L}^4$-Version)**.** *Seien X_n Zufallsvariablen mit $\mathbb{E}(|X_n|^4) < \infty$, dann ist*

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}(X_1)\right) = 1$$

Beweis. Wegen $\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}(X_1) = \frac{1}{n}\sum_{i=1}^n \left(X_i - \mathbb{E}(X_i)\right)$ kann man $\mathbb{E}(X_1) = 0$ annehmen.

(i) Sei

$$A_n = \left\{ \omega \in \Omega \mid \left| \frac{1}{n} \sum_{i=1}^n X_i(\omega) \right| \geq n^{-\frac{1}{8}} \right\}$$

Es gilt

$$\begin{aligned}
\mathbb{P}(A_n) &= \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq n^{-\frac{1}{8}} \right) \\
&\underset{\text{„Chebychev } f(x)=x^4\text{“}}{\overset{4.28}{\leq}} \frac{\mathbb{E}\left(\left(\frac{1}{n} \sum_{i=1}^n X_i \right)^4 \right)}{\left(n^{-\frac{1}{8}} \right)^4} = n^{\frac{1}{2}} n^{-4} \mathbb{E}\left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n X_i X_j X_k X_l \right) \\
&= n^{\frac{1}{2}-4} \sum_{i,j,k,l=1}^n \mathbb{E}\left(X_i X_j X_k X_l \right)
\end{aligned}$$

$\mathbb{E}(...) \neq 0$, nur wenn je zwei oder alle Indizes gleich sind.

$$\begin{aligned}
&= n^{\frac{1}{2}-4} \left(\sum_{i=1}^n \mathbb{E}(X_i^4) + \sum_{i=1}^n \sum_{\substack{i=1 \\ i \neq j}}^n \mathbb{E}(X_i^2 X_j^2) \right) \\
&= n^{\frac{1}{2}-4} \left(n\mathbb{E}(X_1^4) + n(n-1)\mathbb{E}(X_1^2)^2 \right) \leq cn^{-\frac{3}{2}}
\end{aligned}$$

Also ist $\sum_{i=1}^\infty \mathbb{P}(A_n) \leq \infty$.

(ii) Dann folgt mit dem Lemma von Borel-Cantelli (3.35):

$$\begin{aligned}
0 = \mathbb{P}\left(\limsup_{n\to\infty} A_n \right) &= \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq n^{-\frac{1}{8}} \text{ unendlich oft} \right) \\
&= 1 - \underbrace{\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq n^{-\frac{1}{8}} \text{ nur endlich oft oft} \right)}_{\geq \mathbb{P}\left(\lim_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^n X_i \right| = 0 \right)} \\
&\geq 1 - \mathbb{P}\left(\lim_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^n X_i \right| = 0 \right)
\end{aligned}$$

Dann ist also $\mathbb{P}\left(\lim \lim\inf_{n\to\infty} \left| \frac{1}{n} \sum_{i=1}^n X_i \right| = 0 \right) \geq 1$, also $= 1$.

$\square$

Bemerkung 5.10. Die aussage von 5.9 gilt auch falls $X \in \mathscr{L}^1$.

Bemerkung 5.11. Rückblick auf 5.3:
Falls X_i uiv, dann ist

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}(X_1) \right| \geq \epsilon_n \right) \leq \frac{V}{n\epsilon_n^2}$$

Also gilt $\mathbb{P}(...) \xrightarrow{n\to\infty} 0$, falls $\epsilon_n \gg \frac{1}{\sqrt{n}}$ (Also $\limsup_{n\to\infty} \sqrt{n}\epsilon_n = \infty$).
Das Bedeutet1: Der Abstand zwischen $\frac{1}{n}\sum_{i=1}^n X_i$ und $\mathbb{E}(X_1)$ schrumpft „fast" wie $\frac{1}{\sqrt{n}}$.

Frage: Was ist $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n \left(X_i - \mathbb{E}(X_i)\right)\right| \geq \frac{1}{\sqrt{n}}\right)$?
Für $\alpha > \frac{1}{2}$ gilt

$$\mathbb{V}\left(n^{\alpha-1}\sum_{i=1}^n \left(X_i - \mathbb{E}(X_i)\right)\right) = n^{2\alpha-2}\sum_{i=1}^n \mathbb{V}(X_1)$$

Sei $v = n^{2\alpha-1}$, dann $v \xrightarrow{n\to\infty}$, falls $\alpha > \frac{1}{2}$ und $n > 0$. Also ist zu erwarten, dass für $\alpha > \frac{1}{2}$ gilt

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}(X_i)\right| \leq n^{-\alpha}\right) \not\to 0$$

Tatsächlich gilt

Satz 5.12 (zentraler Grenzwertsatz)**.** *Sei X_n eine Folge von unabhängigen, identisch verteilten Zufallsvariablen, $\mathbb{E}(X_i) = m$, $\mathbb{V}(X_i) = v \leq \infty$ für alle $i \in \mathbb{N}$.*
Dann gilt

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - m) \leq c\right) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi v}} e^{-\frac{x^2}{2v}}\ dx$$

Bemerkung 5.13. Die Gleichung aus 5.12 bedeutet, dass die Folge der Bildmaße $\mathbb{P}_{\overline{S}_n(\omega)}$ von $\mathbb{P}$ unter $\omega \mapsto \overline{S_n(\omega)} := \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i(\omega) - m)$ konvergiert gegen $\mathcal{N}_{0,v}$.

Im Sinne der Maßkonvergenz

$$\mathbb{P}_{\overline{S}_n}\big((-\infty, c]\big) \xrightarrow{n\to\infty} \mathbb{P}_{\mathcal{N}_{0,v}}\big((-\infty, c]\big) \forall c$$

Definition 5.14. Seine Folge μ_n von endlichen Maßen auf $\mathbb{R}$ heißt **schwach konvergent** gegen ein Maß μ, falls $\forall f \in C_0(\mathbb{R})$

$$\lim_{n\to\infty} \mathbb{E}_{\mu_n}(f) = \mathbb{E}_{\mu}(f) \quad \text{bzw.} \quad \lim_{n\to\infty} \int f\ d\mu_n = \int f\ d\mu$$

Bemerkung. In 5.12 verwenden wir $f(x) = \mathbb{1}_{(-\infty,c]}(x)$. Da f unstetig ist, gilt z.b. mit $\mu_n = \delta_{\frac{1}{n}}$, $g = \mathbb{1}_{(-\infty,0]}$, $\mu = \delta_0$. Dann ist

$$\int f\ d\mu_n = f\left(\frac{1}{n}\right) \xrightarrow{n\to\infty} f(0)$$

falls f stetig differenzierbar bei 0, aber $\int g\ d\mu_n = 0$, $\int g\ d\mu = 1$.

Proposition 5.15. Seien μ_n Wahrscheinlichkeitsmaße auf $\mathbb{R}$, μ Wahrscheinlichkeitsmaß auf $\mathbb{R}$ und μ habe Lebesgue-Dichte ρ und sei für alle $k \in \mathbb{N}$

$$\lim_{n\to\infty} \mathbb{E}_{\mu_n}(f) = \mathbb{E}_{\mathbb{P}_\mu}(f) \forall f \in C_b^k$$

Dann ist

$$\lim_{n\to\infty} \mathbb{P}_{\mu_n}\big((-\infty, c]\big) = \mathbb{P}_{\mu}\big((-\infty, c]\big) \forall c \in \mathbb{R}$$

Beweis. (i) Zu zeigen: $\forall x \in \mathbb{R} \forall \epsilon > 0 \exists \delta > 0$

$$\limsup_{n\to\infty} \mathbb{P}_{\mu_n}\big([x-\delta, x+\delta]\big) < \epsilon$$

Wähle dazu $f \in C_b^k$ mit $0 \leq f \leq 1$, $f(y) = 1 \forall y \in [x-\delta, x+\delta]$ und $f(y) = 0 \forall y \notin [x-2\delta, x+2\delta]$.
Dann gilt

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}_{\mu_n}\big([x-\delta, x+\delta]\big) \leq \limsup_{n\to\infty} \mathbb{E}_{\mu_n}(f) &= \mathbb{E}_{\mathbb{P}_X}\mu(f) \\ &= \int_{x-2\delta}^{x+2\delta} f(y)\rho(y)\ dy \qquad \leq \int_{x-2\delta}^{x+2\delta} \rho(y)\ dy \xrightarrow{\delta\to 0} 0\end{aligned}$$

mit 4.11, da $\rho \in \mathscr{L}^1$.

(ii) Zu $\epsilon > 0$, $c \in \mathbb{R}$ wähle $f(x) = 1 \forall x \leq c$.
Dann ist

$$\mathbb{P}_{\mu_n}\big((-\infty, c]\big) - \mathbb{P}_{\mu}\big((-\infty, c]\big) = \mathbb{E}_{\mu_n}\left(\mathbb{1}_{(-\infty,c]-f}\right) + \mathbb{E}_{\mu_n}(f) - \mathbb{E}_{\mu}(f) - \mathbb{E}_{\mu}\underbrace{\left(\mathbb{1}_{(-\infty,c]} - f\right)}_{\substack{|\cdot| \leq \mathbb{1}_{(-\infty,c]} \\ (\star)}}$$

Mit (i) folgt, δ existiert, sodass $|(\star)| \leq 2\epsilon + |\mathbb{E}_{\mu_n}(f) - \mathbb{E}_{\mathbb{P}_X}\mu(f)|$. Also ist $\limsup_{n\to\infty} |(\star)| \leq 2\epsilon \forall \epsilon > 0$.

□

Beweis von 5.12. Setze $\tilde{X}_i = \frac{1}{\sqrt{v}}(X_i - m)$. Dann ist $\mathbb{E}(\tilde{X}_i) = 0$ und $\mathbb{V}(\tilde{X}_i) = 1$ und

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - m) \geq c\right) = \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i \geq c\sqrt{v}\right)$$

Also kann oBdA angenommen werden, dass bereits $\mathbb{E}(X-i) = 0$ und $\mathbb{V}(X_i) = 1$ gilt.
Sei nun $\overline{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$.
Mit 5.15 folgt, dass es bereits ausreicht, dass zu prüfen, ob (mit $k = 3$) $\lim\limits_{n\to\infty} \mathbb{E}\big(f(\overline{S}_n)\big) = \mathbb{E}_{\mathscr{N}_{0,1}}(f)$ für alle $f \in C_b^3(\mathbb{R})$ gilt.

Sei dazu (Y_n) eine weiter Folge von unabhängig identisch verteilen Zufallsvariablen mit

a) (Y_n) unabhängig von (X_n)

b) $(Y_n \sim \mathscr{N}_{0,1}) \forall n$

Dann gilt für $\overline{T}_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \sim \mathscr{N}_{0,1}$, dass

$$\mathbb{E}\big(f(\overline{T}_n)\big) = \mathbb{E}_{\mathscr{N}_{0,1}}(f)$$

Also gilt es zu Ziegen, dass

$$\forall f \in \mathbb{C}_b^3 \quad \mathbb{E}\big(f(\overline{S}_n) - f(\overline{T}_n)\big) \xrightarrow{n\to\infty} 0$$

Dabei gilt

$$
\begin{aligned}
&- f\big(\overline{S}_n(\omega)\big) - f\big(\overline{T}_n\big)\\
=&\sum_{i=1}^{n}\left[f\left(\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{i}X_j(\omega)+\sum_{j=i+1}^{n}Y_j(\omega)\right)\right)-f\left(\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{i-1}X_j(\omega)+\sum_{j=i}^{Y_j(\omega)}\right)\right)\right]\\
=&\sum_{i=1}^{n} f
\end{aligned}
$$

□

6 Grundzüge der Statistik

6.A Fragestellung und Modellbildung

Bemerkung 6.1. a) Ausgangspunkt: Daten $x = (x_1, ..., x_n) \in \mathbb{R}$ oder allgemein $x \in \Omega$. Diese Daten sind fest vorgegeben, nicht zufällig und nicht verhandelbar (z.B. Messwerte).

b) Modellannahme: $x_1, ..., x_n$ sind das Ergebnis eines (unbekannten) zufälligen Geschehens. Also gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$ mit $x \in \Omega$, sodass x das Ergebnis von „einmal würfeln mit $\mathbb{P}$“ ist.

c) Ziel: Finde heraus welches $\mathbb{P}$ (äquivalent: welche Zufallsvariable X) die Daten x erzeugt haben könnte.

d) Häufige zusätzliche Annahme: Es kommt in b) nur eine Familie $(P_\theta)_{\theta\in\Theta}$ mit $\mathbb{P}_\theta = \mathscr{N}(\mu, \sigma^2)$ in Frage.
θ heißt Parameter. Daher spricht man von „parametrische Statistik“.

Beispiel 6.2 (Feuerwerk). Großpackung 10^6 Raketen. Davon werden 100 getestet und 5 davon sind defekt.
Wie viele der 10^6 Raketen sind voraussichtlich Defekt (Ausschussquote)?

a) Daten: $x = 5$

b) Modell (Approximativ) ziehen mit zurücklegen (eigentlich: ohne) mit Erfolgswahrscheinlichkeit $\theta = \frac{|\{\text{defekte}\}|}{10^6}$.

d) $\Theta = [0, 1]$, $\mathbb{P}_\theta = \mathscr{B}_{N,\theta}$ binomialverteilt $N = 100$, Erfolgswahrscheinlichkeit θ.

c) Welches θ hat die Daten erzeugt.

Beispiel 6.3 („6.2“ abstrahiert). Ein Ratespiel. Der Spielleiter hat für jedes $\theta \in \Theta$ einen Würfel mit der Verteilung $\mathbb{P}_\theta$. Er nimmt einen davon, würfelt einmal und sagt das Ergebnis.
Nun muss man möglichst gut Raten welcher Würfel genommen wurde.

Definition 6.4. Ein **statistisches Modell** (sM) ist ein Tripel $(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, wobei $(\Omega, \mathscr{F})$ ein Messraum ist und Θ die sog. „**Parametermenge**“ und $\mathbb{P}_\theta$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$ für alle $\theta \in \Theta$ ist.

Definition 6.5. Ein statistisches Modell heißt **Produktmodell**, falls $n \in \mathbb{N}$ existiert, sodass für $(\tilde{\Omega}, \tilde{\mathscr{F}})$ Messraum, $\tilde{\mathbb{P}}_\theta$ Familie von Wahrscheinlichkeitsmaßen auf $\tilde{\Omega}$

$$\left(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta}\right) = \left(\tilde{\Omega}^n, \tilde{\mathscr{F}}^n, \left(\tilde{\mathbb{P}}_\theta^{\otimes n}\right)_{\theta\in\Theta}\right)$$

ein n-faches unabhängiges Würfeln Modelliert.

Beispiel 6.6. In Beispiel 6.2 alternativ:

$$\Omega = \{0,1\}^{100}, \quad \mathscr{F} = \mathscr{P}(\{0,1\})^{\otimes 100}, \quad \mathbb{P}_\theta = \left(\operatorname{Ber}_{\{0,1\},\theta}\right)$$

Typische Daten: $x = (0, 0, 1, 0, ..., 1)$. Interssant ist die Anzahl der Einsen.

Definition 6.7. $\left(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta}\right)$ sei statistisches Modell. Eine messbare Abbildung $T : \Omega \to \Theta$ heißt **Schätzer** für den Parameter $\theta \in \Theta$.
Eine messbare Abbildung $\Omega \to \Omega'$ heißt Ω'-wertiger Schätzer.
Bedeutung: Bei Daten $x \in \Omega$ „schätzt“ man dass $\theta = T(x)$. Ein Ω'-wertiger Schätzer ist sinnvoll, falls z.B. $\Theta = \mathbb{R} \times \mathbb{R}^+$, $\theta = (\mu, \sigma^2)$, $\mathbb{P}_\theta = \mathscr{N}(\mu, \sigma^2)$ man aber nur μ wissen will. Dann ist $\Omega' = \mathbb{R}$.

Bemerkung. 6.7 verlangt nicht,dass der geschätzte Wert $T(x)$ etwas mit dem wahren θ zu tun hat. Im falle eines Produktmodells hat man aber folgendes Gütekriterium:

Definition 6.8. Zu jedem $n \in \mathbb{N}$ sei $(\Omega^n, \mathscr{F}^{\otimes n}, \left(\mathbb{P}_{(X,Y)}\theta\right)_{\theta\in\Theta}^{\otimes n})$ ein Produktmodell $T_N : \Omega \to \Omega'$ ein Schätzer, $f : \Theta \to \Omega'$ eine Funktion: $(T_n)_{n\in\mathbb{N}}$ heißt **schwach konsistente Familie von Schätzern**, falls

$$\forall \theta \in \Theta, \forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}_{\theta^{\otimes n}}\left(|T_n - f(\theta)| > \epsilon\right) = 0$$

Bemerkung 6.9. 6.8 bedeutet nicht, dass für gegebene Daten x und $\epsilon > 0$ und großem n der Wert $T(x)$ mit hoher Wahrscheinlichkeit höchstesn Abstand ϵ von $f(\theta)$ hat.
Denn x und θ sind fest (obwohl θ unbekannt ist). Hier kommt kein Zufall in Spiel.
Stattdessen: Egal welches θ das richtige ist liegt $T(x)$ nur dann weit von $f(\theta)$ weg (für große n) wenn wir beim Erzeugen der Daten „Pech hatten“.
Mit anderen Worten: Würden wir die Messung häufig wiederholen und Datensätze $x^{(1)}, ..., x^{(m)}$ erzeugen, so wäre für die meisten dieser Datensätze die Ungleichung $|T(x^{(k)}) - f(\theta)| \le \epsilon$ wahr.

Beispiel 6.10. In 6.2 wählt man $T(x) = \frac{x}{100}$ bzw in der Form **??**: $T(x) = \frac{1}{100}\sum_{i=1}^{100} x_i$.
Speziell für $x = 5$ ist $T(x) = \frac{1}{20}$, d.h. man schätzt 5% Ausschuss.

Definition 6.11. Sei $(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\theta)_\theta^{\otimes n})$ für alle n ein Produktmodell. Sei $\mathbb{P}_\theta$ die Verteilung einer reelle Zufallsvariable X mit $\mathbb{E}(|X|) < \infty$ und $\mathbb{E}(X) = \theta$. Dann heißst $T_n : \Omega^n \to \mathbb{R}, x \mapsto \frac{1}{n}\sum_{i=1}^n x_i$ heißt **Mittelwertschätzer**.

Satz 6.11. *Der Mittelwertschätzer ist schwach konsistent*

Beweis. Durch GGZ. □

Definition 6.12. Sei T Schätzer für eine Größe $f(\theta)$ dann

a) ist der sog **Bias** definiert als

$$\mathrm{Bias}_\theta := \mathbb{E}_\theta(T) - f(\theta)$$

b) heißt T **Erwartungstreu**, falls $\mathrm{Bias}_\theta = 0$.

c) der sog **mean square error** definiert als

$$\mathrm{MSE}_\theta(T) := \mathbb{E}_{\mathbb{P}_\theta^{\otimes n}}\left(\left(T - \mathbb{E}_{\mathbb{P}_\theta^{\otimes n}}(T)\right)^2\right) = \mathbb{V}_{\mathbb{P}_\theta^{\otimes n}}(T)$$

d) heißt T **konsistent im quadratischen Mittel** (im Produktmodell), falls $\mathrm{MSE}_\theta \xrightarrow{n\to\infty} 0$.

Lemma 6.13. *Sei T erwartungstreu und $\mathbb{P}_{(X,Y)}\theta^{\otimes n}(T)$ konsistent im quadratischen Mittel, dann ist T schwach konsistent.*

Definition 6.14. Sei $(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\theta)_\theta^{\otimes n})$ für alle n ein Produktmodell, sei $\mathbb{P}_{(X,Y)}\theta$ die Verteilung einer rellen Zufallsvariable X mit $\mathbb{E}(|X|) < \infty$, $\mathbb{V}_\theta(X) = \theta$.
Dann ist der **Varianzschätzer** gegeben durch

$$x \mapsto \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \frac{1}{n}\sum_{i=1}^{n} x_j)^2 \qquad \text{(„Empirische Varianz“)}$$

Korollar 6.14. *T ist erwartungstreu*

Korollar 6.14. *Wenn (X_i) Gaußverteilt ist, dann ist T konsistent im quadratischen Mittel.*

Bemerkung. In der Theorie der Statistik untersucht man „optimale“ Schätzer (z.B. mit minimalem MSE). Sie zu finden ist nicht immer einfach.

Beispiel 6.15. Sei $\Omega = [0,\infty)$, $\mathbb{P}_{(X,Y)}\theta = \frac{1}{\theta}\mathbb{1}_{[0,\theta]}\cdot\lambda$. (Gleichverteilung auf $[0,\theta]$).
Modell: $(\Omega^n, \mathscr{F}^{\otimes n}, \mathbb{P}_\theta^{\otimes n})$. Schätze θ aus Daten $x_1, ..., x_n$.
Mögliche Sätzer

a) $T_1(x) = 2\frac{1}{n}\sum_{i=1}^n x_i$

b) $T_2(x) = \frac{n+1}{n}\max\{x_1, ..., x_n\}$

c) $T_3(x) = \max(x_1, ..., x_n) + \min\{x_1, ..., x_n\}$

6.B Maximum-Likelihood-Schätzer

Definition 6.16. a) Auf diskreten Wahrscheinlichkeitsräumen: Sei $(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ statistisches Modell, Ω abzählbar, $\rho_\theta(x) := \mathbb{P}(\{x\})$ sei Zähldichte von $\mathbb{P}_\theta$.

Simpled Idee (i) Zu den Daten x Berechne $\mathbb{P}_\theta(\{x\})$

(ii) Rak-Vorschrift: das „richtige“ θ ist jenes für das $\mathbb{P}_\theta(\{x\})$ maximal ist.

Formal Zu $x \in \Omega$ heißt die Abbildung $\theta \mapsto \rho_\theta(x)$ Likelyhood- (oder Plausibilitäts-)Funktion zum Beobachtungswert x.
Eine Abbildung $\Omega \to \Theta, x \mapsto \arg\max \rho_x$ heißt **maximum-Likelyhood Schätzert**.

b) Stetige Wahrscheinlichkeitsräume mit Dichte: Sei $(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ statistisches Modell, $\Omega = \mathbb{R}^d$, ρ_θ Dichtefunktion zu $\mathbb{P}_\theta$.
Dann heißt die Abbildung $\Omega \to \Theta, x \mapsto \arg\max \rho_\theta$ **ML-Schätzer**.

Bemerkung 6.17. Berechnen von ML Schätzern ist meist durch Ableiten von $\theta \mapsto \rho_\theta(x)$ und Nullsetzen möglich. Da Schätzer häufig Produktstruktur haben ist es oft einfacher $\theta \mapsto \ln \rho_\theta(x)$.

Definition 6.18. Sei $(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ statistisches Modell, $f : \Theta \to \Omega'$ messbar, $\alpha > 0$.
Eine Abbildung $K : \Omega \to \mathscr{P}(\Omega')$ heißt **Konfidenzbereich für** f zum **Irrtumsniveau** α (**Sicherheitsniveau** $1 - \alpha$), wenn gilt

(i) $\{\omega \in \Omega : f(\theta) \in K(\omega)\} \in \mathscr{F}$

(ii) $\mathbb{P}_\theta\big(f(\theta \in K(\cdot))\big) \geq 1 - \alpha$

Bemerkung 6.19. es wird zu jedem Datensatz x eine Menge $K(x) \subset \Omega'$ bestimmt, sodass $f(\theta)$ nur mit sehr geringer Wahrscheinlichkeit (beim Erzeugern der Daten) nicht in $K(x)$ liegt, falls θ der wahre Wert war.
Genauer: Macht man das Experiment

a) Würfele x mit $\mathbb{P}_\theta$ aus

b) rate, dass $\theta \in K(x)$

c) Schaue nach ob richtig geraten wurde

Häufiges unabhängiges durchführen: Man hat in mehr als $(1-\alpha)N$ Fällen recht.

Definition 6.20. Sei (vgl **??**) $(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ statistisches Modell, $f : \Theta \to \Omega'$ messbar, $\Omega' = \mathbb{R}$ und $K : \Omega \to \mathscr{P}(\Omega')$. Dann heißt die Abbildung $x \mapsto K(x)$ **Konfidenzintervall**.

Satz 6.21 (Berechnung von Konfidenzintervallen)**.** *Sei $(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ statistisches Modell, $\Omega' = \mathbb{R}$, $f : \Theta \mapsto \mathbb{R}$ Zielgröße, $x \in \Omega$ Daten.*
Gesucht: $KI : [R^-, R^+]$.

Einzige Möglichkeit *Sage R_-, R_+ aus den Daten vorher. Gesucht sind Schätzer $x \mapsto R_-(x), x \mapsto R_+(x)$ für die Intervallenden.*

Benötigte Eigenschaft

$$\forall\theta : \mathbb{P}_\theta\big(f(\theta) \in [R_-, R_+)\big) \geq 1-\alpha \Leftrightarrow \mathbb{P}_\theta\big(R_- > f(\theta)\big) \textit{ oder } \mathbb{P}_\theta\big(R_+ < f(\theta)\big) < \alpha$$

Hinreichend *(und in der Praxis auch immer Benutzt)*

$$\mathbb{P}_\theta\big(R_- > f(\theta)\big) \leq \alpha/2 \quad \textit{und} \quad \mathbb{P}_\theta\big(R_+ < f(\theta)\big) \leq \alpha/2$$

Finde also $\forall\theta \in \Theta$ eine Funktion $\omega \mapsto R_{-,\theta}(\omega)$, sodass

$$\mathbb{P}_\theta\big(R_{-,\theta} > f(\theta)\big) \leq \alpha/2 \Leftrightarrow \mathbb{P}_\theta \circ R_{-,\theta}^{-1}\big([f(\theta, \infty))\big) \leq \alpha/2$$

Da man θ nicht kennt muss man annehmen $R_-(x) = \inf\{R_{-,\theta(x)} | \theta \in \Theta\}$ und $R_+(x) = \sup\{R_{+,\theta(x)} | \theta \in \Theta\}$.

Beispiel 6.22. Sei $\mathbb{P}_\theta = \mathscr{N}_{\theta,\sigma^2}$, σ^2 bekannt und fest.
Finde $f(\theta) = \theta$, d.h. schätze den Erwartungswert.
$(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta})$ Wähle dann

$$R_-(x) = \frac{1}{n}\sum_{i=1}^n x_i - s \quad R_+ = \frac{1}{n}\sum_{i=1}^n x_i + s$$

Dann ist , da$X_i \sim \theta, \sigma^2$, $Y_i \sim \mathscr{N}_{0,\sigma}$, $Z \sim \mathscr{N}_{0,\sigma^2}$, $W \sim \mathscr{N}_{0,1}$.

$$\begin{aligned}
\mathbb{P}^{\otimes n}(R_- \geq \theta) &= \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n (X_i - \theta) \geq s\right) \\
&= \mathbb{P}\left(\frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i\right) \geq s\right) \\
&= \mathbb{P}\left(\frac{1}{\sqrt{n}} Z \geq s\right) \\
&= \mathbb{P}(Z \geq s\sqrt{n}) = \mathbb{P}\left(W \geq \frac{s}{\sigma}\sqrt{n}\right)
\end{aligned}$$

Also muss s (mindestens) als Lösung der Gleichung

$$\int_{\frac{s}{\sigma}\sqrt{n}}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \alpha/2$$

wählen.

Definition 6.23. Sei $\mathbb{P}$ Wahrscheinlichkeitsmaß auf $\mathbb{R}$, $0 < \alpha < 1$.
Eine Zahl $q_a \in \mathbb{R}$ heißt **α-quantil** von $\mathbb{P}$, falls $\mathbb{P}\left((-\infty, q_\alpha)\right) \leq \alpha$ und $\mathbb{P}\left((-\infty, q_\alpha]\right) \geq \alpha$.

Lemma 6.24. *In 6.23 sei F die Verteilungsfunktion von $\mathbb{P}$. Dann gilt*

a) $\lim\limits_{x\uparrow q_\alpha} \leq \alpha$ *und* $F(q_\alpha) \geq \alpha$ *für jedes α-Quantil q_α.*

b) Jede zahl im Intervall $(\sup\{y \in \mathbb{R} : F(y) \leq \alpha\}, \inf\{y \in \mathbb{R} : F(y) \geq \alpha\})$ *ist ein α-Quantil*

c) Falls F stetig und streng monoton ist, dann ist q_α die eindeutige Zahl mit $F(q\alpha) = \alpha$.

Proposition 6.25. Sei $\left(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})_{\theta\in\mathbb{R}}\right)$ statistisches Modell mit $\mathbb{P}_\theta = \mathscr{N}(\theta, \sigma^2)$, wobei σ^2 bekannt und fest. Sei $q_{1-\frac{\alpha}{2}}$ das $1-\frac{\alpha}{2}$-Quantil von $\mathscr{N}(0,1)$.
Dann ist die Abbildung

$$K : \Omega^n \to \mathscr{P}(\mathbb{R}) : x \mapsto \left[M(x) - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}, M(x) + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\right] = \left[M(x) + \frac{\sigma}{\sqrt{n}} q_{\alpha/2}, M(x) + \frac{\sigma}{\sqrt{n}} q_{1-\alpha/2}\right]$$

(Dabei ist $M(x) := \frac{1}{n}\sum_{i=1}^n x_i$ der Mittelwert) ein Konfindenzintervall zum Irrtumniveau α.

Beweis. Wähle $s = \frac{\sigma}{\sqrt{n}} q_{1-\alpha}$ in6.22 $\square$

Beispiel 6.26. Sei $\Theta = \mathbb{R}\times\mathbb{R}^+$, $\theta = (\mu, \sigma^2)$, $\mathbb{P}_\theta = \mathscr{N}(\mu, \sigma^2)$ und $\big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})$ statistisches Modell.
Gesucht wird nun das Konfidezintervall für die Varianz σ^2.
Der einzige vernünftige Schätzer ist dann

$$S^2 : \Omega^n \to \mathbb{R}^+,\ S^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}\big(x_i - M(x)\big)^2$$

Für $X_i \sim \mathscr{N}(\mu, \sigma^2)$ ist dann

$$\begin{aligned} S^2\big((X_1, ..., X_n)\big) &= \frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu) - \frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)\right]^2 \\ &= \frac{\sigma^2}{n-1}\sum_{i=1}^{n}\left[\frac{X_1-\mu}{\sigma} - \frac{1}{n}\sum_{j=1}^{n}\frac{X_i-\mu}{\sigma}\right]^2 \end{aligned}$$

Dabei ist $\frac{X_i-\mu}{\sigma} \sim \mathscr{N}(0,1)$, also mit $Y_1, ..., Y_n \sim \mathscr{N}(0,1)$ uiv

$$\sim \sigma^2 S^2\big((Y_1, ..., Y_n)\big)$$

Da Bildmaß von $\frac{S^2(x)}{\sigma^2}$ ist unabhängig von θ! Also folgt

$$\begin{aligned} \mathbb{P}_\theta(S^2 > a\sigma^2) &= \mathbb{P}_\theta \circ \left(\frac{S^2}{\sigma^2}\right)^{-1}\big((a,\infty]\big) =: \nu\big(a,\infty\big) \\ \mathbb{P}_\theta(S^2 < b\sigma^2) &= \mathbb{P}_\theta \circ \left(\frac{S^2}{\sigma^2}\right)^{-1}\big((-\infty,b)\big) =: \nu\big(-\infty,b\big) \end{aligned}$$

Falls man a als $(1-\frac{\alpha}{2})$-Quantil von ν und b als $\alpha/2$-Quantil von μ wählt, dann gilt

$$\mathbb{P}_\theta\big(\sigma^2 \in [\frac{1}{a}S^2, \frac{1}{b}S^2]\big) \geq 1-\alpha$$

Definition 6.27. Inhalt...

Satz 6.28. *Seien $X-1, ..., X_n$ uiv und $X_1 \sim \mathscr{N}(\mu,\sigma)$. Dann ist $M(x) = \dfrac{1}{n}\sum_{i=1}^n X_i(\omega)$, $S^2(\omega) = \dfrac{1}{1-n}\sum_{i=1}^n \big(X_i(\omega) - M(\omega)\big)$.*

Bemerkung. t_{n-1} ist ähnlich wie $\mathscr{N}(0,1)$ und für große n gilt $\mathbb{P}_{t_n} \to \mathscr{N}(0,1)$

Proposition 6.29. Sei $\big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})$ statistisches Modell mit $\mathbb{P}_\theta = \mathscr{N}(\mu,\sigma^2)$, $f(\theta) = f\big((\mu,\sigma^2)\big) = \sigma^2$.
Seien q_α das α-Quantil der χ^2_{n-1}-Verteilung. Dann ist die Abbildung

$$K : \Omega^n \to \mathscr{B}\big([0,\infty]\big), \quad x \mapsto \left[(n-1)\frac{S^2(x)}{q_{1-\alpha/2}}, (n-1)\frac{S^2(x)}{q_{\alpha/2}}\right]$$

ein Konfidenzintervall von $f(\theta)$ zum Irrtumsniveau α.

Beweis. Es gilt

$$\begin{aligned}\mathbb{P}_\theta\left(f(\theta < \frac{(n-1)S^2}{q_{1-\alpha/2}})\right) &= \mathbb{P}\left((n-1)S^2 > \sigma^2 q_{1-\alpha/2}\right) \\ &= \mathbb{P}_\theta\left(\frac{n-1}{\sigma^2}S^2 > q_{1-\alpha/2}\right) \\ &= \mathbb{P}_{\chi^2_{n-1}}\left(Y > q_{1-\alpha/2}\right) = \alpha/2\end{aligned}$$

und $\mathbb{P}_\theta\big(f(\theta) > (n-1)^2 \frac{S^2}{q_{\alpha/2}}\big) = ... = \mathbb{P}_{\mathbb{1}^2_{n-1}}(Y < q_{\alpha/2}) = \frac{\alpha}{2}$. □

Bemerkung 6.30. 6.29 erscheint so also ob das Konfidenzintervall mit wachsendem n größer wird. Kontraintuitiv!
Es gilt

$$\mathbb{P}_{\mathbb{1}^2_n}(\mathrm{id} \leq c) = \mathbb{P}\left(\sum_{j=1}^n X_i \leq c\right)$$

mit $X - i \sim \Gamma_{\frac{1}{2},\frac{1}{2}}$ (siehe **??**c).
Dann gilt $\mathbb{E}(X_i) = 1$, $\mathbb{E}(X_i^2) = 3$, also $\mathbb{V}(X_i) = 2$. Somit ist

$$\begin{aligned}\mathbb{P}_{\mathbb{1}^2_{n-1}}((-\infty, c]) &= \mathbb{P}\left(\sum_{j=1}^n (X_j - 1) \leq c - (n-1)\right) \\ &= \mathbb{P}\left(\frac{1}{\sqrt{n-1}}\sum_{j=1}^n (X_j) \leq \frac{c}{\sqrt{n-1}} - \sqrt{n-1}\right)\end{aligned}$$

ist nach dem Gesetz der Großen Zahlen $\sim \mathscr{N}(0,2)$ für große n.
Beispiel 6.31. Sei $\Theta = \mathbb{R} \times R^+$, $\theta = (\mu, \sigma^2)$, $\mathbb{P}_\theta = \mathscr{N}(\mu, \sigma^2)$. Sei $\big(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta}\big)$ statistisches Modell. Nun sind μ und σ^2 unbekannt, gesucht ist ein Konfidenzintervall für μ:
Naiver Versuch: Berechne das Konfidenzintervall gemäß 6.22 für alle σ^2 und maximiere über σ^2. Dann egit sich

$$K_{\sigma^2}(x) = \left[M(x) + \frac{\sigma}{\sqrt{n}} q_{\alpha/2}, M(x) + \frac{\sigma}{\sqrt{n}} q_{(1-\alpha)/2}\right]$$

Dann ergibt das Maximieren über $\sigma \in \mathbb{R}$ ergibt $K(x) = \mathbb{R}$.

Beweis. Für alle $\theta \in \Theta$ gilt

$$\mathbb{P}_\theta\left(M - \frac{\sqrt{S^2}}{\sqrt{n}} q_{1-\alpha/2} \geq \mu\right) = \mathbb{P}_\theta\left(M - \mu \geq \frac{\sqrt{S^2}}{\sqrt{n}} q_{1-\alpha/2}\right) = \mathbb{P}_\theta\left(\sqrt{n}\frac{M-\mu}{\sqrt{S^2}} \geq q_{1-\alpha/2}\right)$$

□

6.C Tests

Bemerkung 6.32. Wichtigstes Werkzeug der Statistik: wissenschaftliche Studien, Medikamentenzulassung, Nachweis von Produktmängeln, ect. Dient zum Treffen von Ja-Nein-Entscheidungen anhand „verrauschter" Daten.
Sei $\big(\Omega, \mathscr{F}, (\mathbb{P}_\theta)_{\theta\in\Theta}\big)$ statistisches Modell. Wir teilen Θ in zwei Bereiche: $A_0, A_1 \subset \Theta$, mit $A_0 \dot\cup A_1 = \Theta$. Dann beschreibt die

- **Nullhypothese**: $\theta \in A_0$ (Schreibe H_0 gilt)
- **Alternative**: $\theta \in A_1$ (Schreibe H_1 gilt)

Testproblem: Kann man aufgrund der Daten $x \in \Omega$ bestimmen, ob $\sigma \in A_0$ oder $\in A_1$ war?
Dabei soll $\theta \in A_1$ nur sehr unwahrscheinlich (bwzüglich der Daten) fälschlicherweise angenommen werden.

Definition 6.33. Eine Abbildung $T : \Omega \to \{0, 1\}$ heißt **Test**. Dabei bedeute

$T(x) = 0$ als „θ liegt in A_0“.
Wir sagen **Nullhypothese wird nicht abgelehnt**

$T(x) = 1]$ als „θ liegt in A_1“.
Wir sagen **Nullhypothese wird abgelehnt**

Falls $\theta \in A_0$ aber $T(x) = 1$ so sagen wir die Daten führen zu einem **Fehler 1. Art**.
Falls $\theta \in A_1$ aber $T(x) = 0$ so sagen wir die Daten führen zu einem **Fehler 2. Art**.

Beispiel 6.34.1 (Feuerwerkskörper). Behauptung des Herstellers: höchstens 3% Ausschuss,
H_1:$\theta > 0.03$, $H_0 : \theta \leq 0.03$.
Falls die Daten suggerieren, dass H_1 gilt, dann gibt es Ärger für alle Beteiligten (Reklamation, Untersuchung, Prozess). Falls H_1 dagegen nicht erkannt wird, sind die Verluste gering.

Beispiel 6.34.2 (Wissenschaftlicher Durchbruch). H_0 entspricht der „Lehrmeinung“, H_1 entspricht „Skandal, Sensation“ (z.B. „Kalte Kernfusion“,„Stammzellentherapie“,...).
Ein Falsches Behaupten von H_1 kostet Reputation, Karriere.
Nicht erkennen von H_1 ist eine verpasste Chance aber nicht fatal.

Definition 6.35. Ein Test T hält das **Irrtumsniveau** $\alpha > 0$ ein, falls gilt:

$$\forall \theta \in A_0 : \mathbb{P}_\theta(T = 1) < \alpha$$

Zu $\theta \in A_1$ heißt die Zahl $G_T(\theta) := \mathbb{P}_\theta(T = 1)$ die **Güte** (oder Schärfe, Machte) des Test bei θ.

Bemerkung 6.36 (Konstruktion von Tests). mit Konfidenzintervallen

a) Zweiseitiger Test: angenommen man hat ein Konfidenzintervall $K : \Omega \to \Omega'$ für die Größe $f(\theta) \in \Omega'$ zum Nievau α gegeben. Dann setzte man

$$T : \Omega \to \{0, 1\},\, T(x) = \begin{cases} ! & \text{falls } K(x) \cap \{f(\theta : \theta \in A_0)\} = \emptyset \\ 0 & \text{sonst} \end{cases}$$

Dann gilt für alle $\theta_o \in A_0$

$$\mathbb{P}_\theta(T = 1)\mathbb{P}_{\theta_0}\big(\forall \theta \in A_0 : f(\theta) \notin K\big) \leq \mathbb{P}_{\theta_0}\big(f(\theta_0) \notin K\big) = 1 - \mathbb{P}_{\theta_0}\big(f(\theta_0) \in K\big) \leq \alpha$$

Interpretation, falls $T(x) = 1$: Die Daten x können nur mit sehr geringer Wahrscheinlichkeit von $\theta_0 \in A_0$ erzeugt werden, sodass wir $\theta_0 \in A_0$ ausschließen.

b) einseitige Tests: mit Asymmetrischen Konfidenzintervallen. Oft ist $\Omega' = \mathbb{R}$, $A_0 = (-\infty, c]$ oder $A_0 = [c, \infty)$. Dann ist es besser statt wie in a) symmetrischer auch asymmetrische Konfidenzintervalle der Form $K(x) = [S(x), \infty)$ oder $K(x) = (-\infty, S(x)]$ zu wählen.

Beispiel 6.37.1 (Gaußtest, Zweiseitig). Inhalt...

7 Markovketten

Bemerkung (Motivation). Sei E eine Menge (=„Spielfeld") mit abzählbar vielen Feldern („Zuständen") =Elemente von E.
Auf jemden Feld $x \in E$ liegt ein Würfel bereit, der $N_r \leq |E|$ Seiten hat und mit Wahrscheinlichkeit $p(x, y)$ die Seite y würfelt.
Man geht dann auf das Feld y.

Beispiel 7.1.1 (Irrfahrt). $E = \mathbb{Z}$, $p(x, x+1) = p(x, x-1) = \frac{1}{2}$.

Beispiel 7.1.2 (Galton-Watson-Populationsmodell). $E = \mathbb{N}_0, p(x, y) = \mathbb{P}\left(\sum_{j=1}^{n} Z_j = y\right)$, wobei Z_j eine Zufallsvariable ist, die die Nachkommen des j-ten Individuum modelliert. Normalerweise sind die Z_j uiv.

Beispiel 7.1.3 (Sammelbilder). Sei $E = \{1, ..., N\}$ („Anzahl der noch fehlenden Bilder")und

$$p(x, y) = \begin{cases} x/N & \text{falls } y = x - 1 \\ 1 - \frac{x}{N} & \text{falls } y = x \\ 0 & \text{sonst} \end{cases}$$

Bemerkung. Wie hoch ist die Wahrscheinlichkeit in 2 Schriteenvon x nach z zu gehen?

$$\begin{aligned} \mathbb{P}(x \xrightarrow{\text{2 Schritte}} z) &= \mathbb{P}(x \to y, y \to z) \\ &= \sum_{y \in E} \mathbb{P}(x \to y)\mathbb{P}(y \to z) \\ &= \sum_{y \in E} p(x, y)p(y, z) \end{aligned}$$

Vergleiche: Matrixmultiplkation

$$= P^2(x, z)$$

mit $P^2 = P \cdot P$, $P = \big(p(x, y)\big)_{x, y \in E}$ (Matrix)
Beachte dabei:

$$\sum_{y \in E} p(x, y) = \sum_{y \in E} \mathbb{P}(x \to y) = \mathbb{P}(x \to E) = 1$$

(Zeilensumme =1)

Definition 7.2. Eine Matrix $P = \big(p(x, y)\big)_{x, y \in E}$ heißt **stochastisch**, wenn gilt

(i) $p(x, y) \geq 0 \ \forall x, y \in E$

(ii) $\sum_{y \in E} p(x,y) = 1 \ \forall x \in E$

Definition 7.3. Sei E höchstens abzählbar, P stochastische Matrix.
Eine Folge (X_n) von E-wertigen Zufallsvariablen heißt **Markovkette** (Mk) mit Übergangsmatrix P (ÜM), falls für alle $n \in \mathbb{N}$ und für alle $x_1, ..., x_n \in E$ gilt

$$\mathbb{P}\left(X_{n+1} = x_{n+1} | X_0 = x_0, ..., X_n = x_n\right) = p(x_n, x_{n+1})$$

Bemerkung. Falls $\mathbb{P}(X_0 = x) = \mu(x)$ für ein Wahrscheinlichkeitsmaß μ, dann schreibt man auch $\mathbb{P}^\mu$ statt $\mathbb{P}$.
Falls

$$\mathbb{P}(X_0 = x) = \delta_y(x) = \begin{cases} 1 & \text{falls } y = x \\ 0 & \text{sonst} \end{cases}$$

schreibt man $\mathbb{P}^y$ statt $\mathbb{P}$.

Proposition 7.4. Sei (X_n) Markovkette mit Übergangsmatrix P. Dann gilt für alle $x \in \mathbb{N}$, $x, y \in E$

$$\mathbb{P}^x(X_n = y) = P^n(x,y)$$

wobei P^n die n-te Potenz von P ist.

Beweis. Induktiv □

Bemerkung. Es besteht eine starke Verbindung zur linearen Algebra. Falls $|E| = \infty$ wird P trotzdem Matrix genannt.

Proposition 7.5. Sei P stochastische Matrix. Dann gilt

a) $P1 = 1$, wobei $1(x) = 1 \forall x$.
Das heißt, 1 ist Rechts-Eigenvektor von P_y zum Eigenwert 1.

b) Falls $|E| < \infty$, dann Existiert ein Zeilenvektor $\mu \neq 0$ mit $\mu P = \mu$. D.h. μ ist Links-Eigenvektor zu P,zum Eigenwert 1.

Beweis. a) $P1(x) = \sum_{y \in E} p(x,y)1(y) = 1$ (Zeilensumme)

b) $P1 = 1$ genau dann wenn $\langle w, P1 \rangle = \langle w, 1 \rangle$ für alle $w \in \mathbb{R}^E$.
Genau dann wenn $\langle P^* w, w, 1 \rangle = 0$ für alle $w \in R^E$.
Genau dann wenn $\mathrm{Im}(P^* w - 1) \perp \langle \{1\} \rangle$.
Dann ist also $\dim(\mathrm{Kern}(P^* - 1)) > 0$.
Aus dem Eigenwert 1 folgt, dass P Zeileneigenwerte zum Eigenwert 1.

□

Definition 7.6. Sei (X_i) Markovkette auf E mit Übergangsmatrix P.
Ein Wahrscheinlichkeitsmaß $\hat{\mu}$ auf E heißt **invariante Verteilung** von (X_n), falls $\mathbb{P}^{\hat{\mu}}(X_1 = y) = \hat{\mu}(\{y\})$. Dabei ist

$$\begin{aligned} \mathbb{P}^{\hat{\mu}}(X_1 = y) &= \sum_{y \in E} \hat{\mu}(\{x\}) \mathbb{P}^x(X_1 = y) \\ &= \sum_{y \in E} \hat{\mu}(\{x\}) p(xy) = (\mu P)(y) \end{aligned}$$

(und μ die Zähldichte von $\hat{\mu}$).

Bemerkung 7.7. Invarianz von Verteilungen bedeutet, dass die Zähldichte μ von $\hat{\mu}$ ein Linkseigenvektor von P ist.
Nach 7.5 existiert für $|E| < \infty$ immer irgendein Linkseigenvektor.
Ist dies auch die Zähldichte einer Verteilung?
Dazu muss der Linkseigenvektor zusätzlich nur nichtnegative Einträge haben und Zeilensumme 1 haben.
Der Satz von Perron-Frobenius („Sei A eine Matrix mit nichtnegativen Einträgen, $\|A\| = \sup\{|Av| : |v| = 1\}$, dann ist w ein Eigenvektor von A zum Eigenwert $\|A\|$. Dann kann w nicht-negativ gewählt werden.“)

Bemerkung. Markovketten sind „gedächtnislose“ Prozesse.
Ist man zur Zeit $x \in X$, dann spielt es für die weitere Entwicklung keine Rolle, wie man dort hin gekommen ist.

Satz 7.8 (Markov-Eigenschaft)**.** *Sei X_n Markovkette mit Übergangsmatrix P. Dann gilt für alle $x, z \in E$, $n \in \mathbb{N}$, $A \in \mathscr{P}(E)^{\otimes \mathbb{N}_0}$ und $B \in E^n$, falls $\mathbb{P}\big((X_1, ..., X_n) \in B, X_n = x\big) > 0$, dann ist*

$$\begin{aligned}\mathbb{P}^z\big((X_n, X_{n+1}, X_{n+2}, ...) \in A | (X_0, ..., X_{n-1}) \in B, X_n = x\big) &= \mathbb{P}^z\big((X_n, X_{n+1}, ...) \in A | X_n = x\big) \\ &= \mathbb{P}\big((X_1, X_2, ...) \in A\big)\end{aligned}$$

Beweis. Sei $k \in \mathbb{N}$, $y_1, ..., y_k \in E$. Dann ist

$$\begin{aligned}&\mathbb{P}\big((x_0, ..., x_{n-1}) \in B,\ x_n = x,\ x_{n+1} = y_i \text{ für alle } i \geq k\big) \\ =& \sum_{\omega_1, ..., \omega_{n-1} \in B} p(z, \omega_1) p(\omega_1, \omega_2) ... p(\omega_{n-2}, \omega_{n-1}) p(\omega_{n-1}, x) p(x_1, y_1) p(y_1, y_2) .... p(y_{k-1}, y_k) \\ =& \mathbb{P}^k\big((x_1, ..., x_{n-1}) \in B,\ x_n = x\big) \mathbb{P}^x\big((x_1, ..., x_k) = (y_1, ..., y_k)\big)\end{aligned}$$

Daraus folgt die Behauptung für $A = \{v \in E^{\mathbb{N}_0} : (v_1, ... = y_1, ..., v_{k-1})\}$.
Nach der Definition der bedingten Wahrscheinlichkeit sind das alle Zylindermengen. Es folgt mit **??** die Behauptung. ☐

Beispiel 7.9 (Ruin des Spielers). Sei X_n = Gewinng nach dem $n-$tenn Spiel, $E = \{-a, ..., 0, ..., b\}$, $p(n, n+1) = p$, $p(n, n-1) = 1 - p$, falls $-a < n < b$.

- $p(-a, -a) = 1$ („Ruin“)
- $p(b, b) = 1$ („Gewinnmitnahme“)

Definition 7.10. $z \in E$ heißt **absorbierend** bezüglich $\mathbb{P}$, falls $p(z, z) = 1$.
Dann heißt

$$h_z(x) := \mathbb{P}^x\left(\exists N \in \mathbb{N} : \forall n \geq N : X_n = z\right) = \mathbb{P}\left(\bigcup_{N \geq 0} \bigcap_{n \geq N} \{x_n = z\}\right)$$

die **Absorptionswahrscheinlichkeit in z bei Startwert x.**

Definition 7.11. Sei (X_n) eine Markovkette, $z \in E$. Die Zufallsvariable

$$\tau_z : \Omega \to \mathbb{N} \cup \{\infty\},\ \omega \mapsto \infty\{n \geq 1 : X_n(\omega) = z\}$$

heißt Trefferzeit (oder Eintrittszeit) bei z.

Lemma 7.12. *Sei (X_n) Markovkette, τ_z Trefferzeit. Dann ist*

$$\{\omega \in \Omega : \tau_z = n\} \in \sigma(X_1, ..., X_n) \tag{Gl. 7.4}$$

Beweis. Es gilt

$$\begin{aligned}\{\omega \in \Omega : \tau_z(\omega) = n\} &= \{\omega \in \Omega : X_1(\omega) \neq z, ..., X_{n-1}(\omega) \neq z, X_n(\omega) = z\} \\ &= \bigcap_{k=1}^{n-1} X_k^{-1}(E \setminus \{z\}) \cap X_n^{-1}\big(\{z\}\big) \in \sigma(X_1, ..., X_n)\end{aligned}$$

□

Bemerkung. Die Messbarkeitseigenschaft Gl. 7.4 macht τ_z zu einer sogenannten **Stoppzeit**. Um zu entscheiden, dass „gestoppt“ wird musss man nciht „in die Zukunft Blicken“.

Satz 7.13. *Sei (X_n) Markovkette, $z \in E$ absorbierend, $x \in E$ beliebig. Dann ist*

$$h_z(x) = \mathbb{P}^x(\tau_z < \infty) = \mathbb{P}^x\left(\lim_{n\to\infty} X_n = z\right) \tag{Gl. 7.5}$$

und

$$h_z = \min\left\{h : E \to \mathbb{R}^+ \text{ mit } h(z) = 1 \text{ und } \sum_{y\in E} p(x,y)h(y) = h(x)\right\} \tag{Gl. 7.6}$$

(wobei $h \leq g \Leftrightarrow h(x) \leq g(x) \forall x \in E$)
Da Minimum ist eindeutig.

Bemerkung. h_z ist also Rechts-Eigenvektor von P, mit $h_z(z) = 1$. Auch 1 mit $1(x) = 1$ ist Rechts-Eigenvektor, aber h_z kann kleiner sein.

Beweis. Zu Gl. 7.5:

$$\mathbb{P}^z(X_n = z\forall >\geq 1) = \lim_{k\to\infty} \mathbb{P}^z(X_n = z\forall n \leq k)$$

Da $\{X_n = z\forall n \leq k\}_k$ eine Folge absteigender Mengen ist

$$= \lim_{k\to\infty} P^k(z,z) ? \lim_{k\to\infty} p(z,z)^k = 1$$

Also gilt für alle $x \in E$

$$\begin{aligned}\mathbb{P}^x(X_n = z) &= \mathbb{P}^x(X_n = z)\mathbb{P}^z(X_i = z\forall i \geq 1) \\ &\overset{7.8}{=} \mathbb{P}^x(X_n = zX_{n+1} = z\forall i \geq 1) \\ &\xrightarrow{n\to\infty} \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty}\{X_j = z\}\right) = h_z(x)\end{aligned}$$

Weiter gilt für alle $k \leq n$

$$\begin{aligned}\mathbb{P}^x(X_n = z|\tau_z = k) &= \mathbb{P}^x(X_n = z|\tau_z = k, X_k = z) \\ &\overset{7.8}{=} \mathbb{P}^x(X_n = z|X_k = z) = \mathbb{P}^z(X_{n-k} = z) = 1\end{aligned}$$

Dann folgt mit 7.12, dass

$$\underbrace{\mathbb{P}^x(X_n = z)}_{\xrightarrow{n\to\infty} h_z(x)} = \sum_{k=1}^{n} \mathbb{P}^x(\tau_z = k)\overbrace{\mathbb{P}^x(X - n = z|\tau_z = k)}^{=1}$$
$$= \mathbb{P}(\tau_n \leq n) \xrightarrow{n\to\infty} \mathbb{P}(\tau_z < \infty)$$

Zu Gl. 7.6:

a) $h_z \in \{h : E \to \mathbb{R}^+, h(z) = 1, Ph = h\}$, denn $h_z(x) \geq 0$ als Wahrscheinlichkeit, $h(z) = 1$ und

$$\sum_{y\in E} p(x,y)h_z(y) = \sum_{y\in E} \mathbb{P}^x(X_1 = y)\mathbb{P}^y(\{\exists N \in \mathbb{N} : \forall n \geq N : X_n = z\})$$
$$= \sum_{y\in E} \mathbb{P}^x(X_n = y)\mathbb{P}(\exists N \in \mathbb{N}_0 : \forall n$$
$$\overset{3.4a)}{=} \mathbb{P}^x(\exists N \in \mathbb{N} : \forall n \geq N : X_{n+1}$$

b) h_z ist maximale. Sei $\tilde{h} >\geq 0$, $\tilde{h}(x) = 1$ und $P\tilde{h} = \tilde{h}$, dann ist

$$\tilde{h}(x) = P^n\tilde{h}(x) = \sum_{x\in E} P^n(x,y)\tilde{h}(y) \geq ...$$

□

Bemerkung 7.14. für $z \in E_0 \subset E$ (ϕ beliebige Funktion) und $Ph(y) = h(y)\forall y \notin E_0$.

Beispiel 7.15 (Galton-Watson-Verzweigungsprozess). Sei n die Anzahl der Individuen, jedes davon bekommt k Kinder mit Wahrscheinlichkeit $\rho(k)$ und stirbt dann. Dann ist also $E = \mathbb{N}_0$ und

$$p(n,m) = \sum_{\substack{k_1,\dots,k_n=1 \\ \sum_{i=1}^{\infty} k_i = m}}^{\infty} \rho(k_1)...\rho(k_n) = \sum_{k_1=0}^{m}\sum_{k_2=0}^{m-k_1} ... \sum_{k_n=0}^{m-\sum_{i=1}^{n-1} k_i} \rho(k_1)...\rho(k_n)$$

Da $\rho(0,0) = 1$ ist 0 absorbierend.

Frage: Wie groß ist $\mathbb{P}$(Population stibt aus | Anfangspopulation=1). Also was ist $\mathbb{P}(\tau_0 < \infty)$?
Triviale Fälle:

$$\rho(0) = \begin{cases} 1 & \text{dann ist } \mathbb{P}^1(\tau_0 < \infty) = 1 \\ 0 & \text{dann ist } \mathbb{P}^1(\tau_0 < \infty) = 0 \end{cases}$$

Für $0 < \rho(0) < 1$ gilt:

a) Für alle $k <, n \geq 0$ ist $\mathbb{P}^k(X_n = 0) = \left[\mathbb{P}^1(X_n = 0)\right]^k$.

Beweis. Induktiv

$n = 0$ stimmt

Induktionsschritt :

$$
\begin{aligned}
&= \sum_{\substack{l_1,\dots,l_k=0 \\ \sum_{i=1}^k l_i = l}}^{\infty} \rho(l_1)\dots\rho(l_k) \underbrace{\mathbb{P}^l(X_n = 0)}_{\text{IndV.}=\mathbb{P}^1(X_n=0)^l} \\
&= \sum_{l_1,\dots,l_k=0}^{\infty} \rho(l_1)\dots\rho(l_n) \left(\mathbb{P}^1(X_n = 0)\right)^{l_1+l_2+\dots+l_k} \\
&= \left(\sum_{l=0}^{\infty} p(1,l)\mathbb{P}^l(X_n = 0)\right)^k \\
&= \left(\mathbb{P}(X_{n+1} = 0)\right)^k
\end{aligned}
$$

□

b) Sei $q(k) := \mathbb{P}^k(\tau_0 < \infty)$.

$$
\begin{aligned}
&= \lim \lim_{n\to\infty} \mathbb{P}^k(\tau_0 \le n) = \lim_{n\to\infty} \mathbb{P}^k(X_n = 0) \\
&\overset{a)}{=} \lim_{n\to\infty} \left(\mathbb{P}^1(X_n = 0)\right)^k = \left[\lim_{n\to\infty} \mathbb{P}^1(X_n = 0)\right]^k = q(1)^k
\end{aligned}
$$

c) $q(1) = h_0(1)$ ist harmonisch. d.h.

$$
\begin{aligned}
q(1) = Pq(1) &= \sum_{k=0}^{\infty} p(1,k)q(k) \\
&\overset{b)}{=} \sum_{k=0}^{\infty} \rho(k)q(1)^k \\
&= \varphi_\rho(q(1))
\end{aligned}
$$

Dabei ist φ die erzeugenden FUnktion von ρ (**??**).
Also ist $q(1)$ ein **Fixpunkt** von φ_ρ un zwar der kleinste. Denn falls s ein beliebiger Fixpunkt von φ_ρ ist, dann gilt $\tilde{h}(k) := s^k : P(\tilde{h}) = \tilde{h}$.
Also ist $\tilde{h}$ harmonisch bezüglich P. Da aber $q(1)$ die minimal harmonsiche Funktion ist gilt $s \ge q(1)$.
Die Funktion $s \mapsto \varphi_\rho(s) = \sum_{k=0}^{\infty} \varrho(k)s^k$ hat immer den Fixpunkt $s = 1$.
Falls $\rho(0) + \rho(1) = 1$, dann ist $\varphi_\rho''(x) = 0$. Also