Stochastik SS 2019

Dozent: Prof. Dr. Anita Behme

29. Juli 2019

Inhaltsverzeichnis

Vorwort

Was ist Stochastik?

Altgriechisch Stochastikos ($\sigma\tau o\chi\alpha\sigma\tau\iota\kappa\grave{o}\zeta$) und bedeutet sinngemäß "scharfsinnig in Vermuten".
Fragestellung insbesondere aus Glücksspiel, Versicherungs-/Finanzmathematik, überall da wo Zufall/ Risiko / Chance auftauchen.
Was ist Stochastik?

- Beschreibt zufällige Phänomene in einer exakten Spache!
  Beispiel: "Beim Würfeln erscheint jedes sechste Mal (im Schnitt) eine 6." $\longrightarrow$ Gesetz der großen Zahlen ($\nearrow$ später)
- Lässt sich mathematische Stochastik in zwei Teilgebiete unterteilen
  Wahrscheinlichkeitstheorie (Wahrscheinlichkeitstheorie) & Statistik
  - *Wahrscheinlichkeitstheorie*: Beschreibt und untersucht konkret gegebene Zufallssituationen.
  - *Statistik*: Zieht Schlussfolgerungen aus Beobachtungen.

  Statistik benötigt Modelle der Wahrscheinlichkeitstheorie. Wahrscheinlichkeitstheorie benötigt die Bestätigung der Modelle durch Statistik.

In diesem Semester konzentrieren wir uns nur auf die Wahrscheinlichkeitstheorie!

Kapitel I

Grundbegriffe der Wahrscheinlichkeitstheorie

1. Wahrscheinlichkeitsräume

Ergebnisraum

Welche der möglichen Ausgänge eines zufälligen Geschehens interessieren uns?
Würfeln? Augenzahl, nicht die Lage und die Fallhöhe

Definition I.1 (Ergebnisraum)
Die Menge der relevanten Ergebnisse eines Zufallsgeschehens nennen wir Ergebnisraum und bezeichnen diesen mit Ω.

■ **Beispiel**

- Würfeln: $\Omega = \{1, 2, \dots, 6\}$
- Wartezeiten: $\Omega = \mathbb{R}_+ = [0, \infty)$ (überabzählbar!)

Ereignisse

Oft interessieren wir uns gar nicht für das konkrete Ergenis des Zufallsexperiments, sondern nur für das Eintreten gewisser Ereignisse.

■ **Beispiel**

- Würfeln: Zahl ist ≥ 3
- Wartezeit: Wartezeit ≤ 5 Minuten

$\longrightarrow$ Teilmenge des Ereignisraums, also Element der Potenzmenge $\mathscr{P}(\Omega)$, denen eine Wahrscheinlichkeit zugeordnet werden kann, d.h. welche messbar (mb) sind.

Definition I.2 (Ereignisraum, messbarer Raum)
Sei $\Omega \neq \varnothing$ ein Ergebnisraum und $\mathscr{F}$ eine σ-Algebra auf Ω, d.h. eine Familie von Teilmenge von Ω, sodass

1. $\Omega \in \mathscr{F}$
2. $A \in \mathscr{F} \Rightarrow A^C \in \mathscr{F}$
3. $A_1, A_2, \dots \in \mathscr{F} \Rightarrow \bigcup_{i \geq 1} A_i \in \mathscr{F}$

Dann heißt $(\Omega, \mathscr{F})$ Ereignisraum bzw. messbarer Raum.

Wahrscheinlichkeiten

Ordne Ereignissen Wahrscheinlichkeiten zu mittels der Abbildung

$$\mathbb{P} : \mathscr{F} \to [0, 1]$$

sodass

$$\text{Normierung } \mathbb{P}(\Omega) = 1 \tag{N}$$

$$\sigma\text{-Additivität für paarweise disjunkte Ereignisse} A_1, A_2, \dots \in \mathscr{F} \Rightarrow \mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(A_i) \tag{A}$$

(N), (A) und die Nichtnegativität von $\mathbb{P}$ werden als Kolmogorovsche Axiome bezeichnet (nach Kolomogorov: Grundbegriffe der Wahrscheinlichkeitstheorie, 1933)

Definition I.3 (Wahrscheinlichkeitsmaß, Wahrscheinlichkeitsverteilung)
Sei $(\Omega, \mathscr{F})$ ein Ereignisraum und $\mathbb{P} : \mathscr{F} \to [0, 1]$ eine Abbildung mit Eigenschaften (N) und (A). Dann heißt $\mathbb{P}$ Wahrscheinlichkeitsmaß oder auch Wahrscheinlichkeitsverteilung.

Aus der Definition folgen direkt:

Satz I.4 (Rechenregeln für W-Maße)
Sei $\mathbb{P}$ ein W-Maß, Ereignisse $(\Omega, \mathscr{F}), A, B, A_1, A_2, \dots \in \mathscr{F}$. Dann gelten:

1. $\mathbb{P}(\varnothing) = 0$
2. Monotonie: $A \subseteq B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$
3. endliche σ-Additivität: $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$
   und insbesondere $\mathbb{P}(A) + \mathbb{P}(A^C) = 1$
4. σ-Subadditivität:
$$\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) \leq \sum_{i \geq 1} \mathbb{P}(A_i)$$
5. σ-Stetigkeit: Wenn $A_n \uparrow A$ (d.h. $A_1 \subseteq A_2 \subseteq \cdots$ und $A = \bigcup_{i=1}^{\infty}(A_i)$) oder $A_n \downarrow A$, so gilt:
$$\mathbb{P}(A_n) \xrightarrow[n \to \infty]{} \mathbb{P}(A)$$

Beweis. In der Vorlesung wurde auf Schilling MINT Satz 3.3 verwiesen. Ausserdem gab es dazu Präsenzübung 1.3. Der folgende Beweis wurde ergänzt.
Beweise erst Aussage: $A \cap B = \varnothing \implies \mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$.
Es kann σ-Additivität verwendet werden, indem "fehlende" Mengen durch $\varnothing$ ergänzt werden:

$$\mathbb{P}(A \uplus B) = \mathbb{P}(A \uplus B \uplus \varnothing \uplus \varnothing \dots) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(\varnothing) + \dots = \mathbb{P}(A) + \mathbb{P}(B),$$

wobei Maßeigenschaften verwendet werden.

1. Definition des Maßes.

2. Da $A \subseteq B$ ist auch $B = A \uplus (B \setminus A) = A \uplus (B \setminus (A \cap B))$. Wende wieder Aussage von oben an, damit folgt

$$\mathbb{P}(B) = \mathbb{P}(A \uplus (B \setminus A)) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A) \tag{*}$$

3. Zerlege $A \cup B$ geschickt, dann sieht man mit oben gezeigter Aussage und (*)

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

   Im letzten Schritt wurde (*) verwendet.
4. Folgt aus endlicher σ-Additivität, da $\mathbb{P}\left(\bigcap_{i\geq 1} A_i\right) \geq 0$.
5. Definiere $F_1 := A_1, F_2 := A_2 \setminus A_1, \dots, F_{i+1} := A_{i+1} \setminus A_n$. Die F_i Mengen sind paarweise disjunkt und damit folgt für $m \to \infty$

$$A_m = \biguplus_{i=1}^{m} F_i \implies A = \biguplus_{i=1}^{\infty} F_i = \biguplus_{i=1}^{\infty} A_i$$

   und

$$\mathbb{P}(A) = \mathbb{P}\left(\biguplus_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(F_i) = \lim \lim_{m\to\infty} \mathbb{P}\left(\biguplus_{i=1}^{m} F_i\right) = \lim \lim_{m\to\infty} \mathbb{P}(A_m). \qquad \square$$

■ Beispiel I.5

Für ein beliebigen Ereignisraum $(\Omega, \mathscr{F})$ $(\Omega \neq \varnothing)$ und eine beliebiges Element $\xi \in \Omega$ definiere

$$\delta_\xi(A) := \begin{cases} 1 & \xi \in A \\ 0 & \text{sonst} \end{cases}$$

ein (degeneriertes) W-Maß auf $(\Omega, \mathscr{F})$, welches wir als Dirac-Maß oder Dirac-Verteilung bezeichnen.

■ Beispiel I.6

Würfeln mit fairem, 6-(gleich)seitigem Würfel mit Ergebnismenge $\Omega = \{1, \dots, 6\}$ und Ereignisraum $\mathscr{F} = \mathscr{P}(\Omega)$ setzen wir als Symmetriegründen

$$\mathbb{P}(A) = \frac{\#A}{6}.$$

(Wobei $\#A$ oder auch $|A|$ die Kardinalität von A ist.) Das definiert ein W-Maß.

■ Beispiel I.7

Wartezeit an der Bushaltestelle mit Ergebnisraum $\Omega = \mathbb{R}_+$ und Ereignisraum Borelsche σ-Algebra $\mathscr{B}(\mathbb{R}_+) = \mathscr{F}$. Ein mögliches W-Maß können wir dann durch

$$\mathbb{P}(A) = \int_A \lambda e^{-\lambda x} \,\mathrm{d}x$$

für einen Parameter $\lambda > 0$ festlegen. (Offenbar gilt $\mathbb{P}(\Omega) = 1$ und die σ-Additivität aufgrund der

Additivität des Integrals.) Wir bezeichnen diese Maße als Exponentialverteilung. (Warum gerade dieses Maß für Wartezeiten gut geeignet ist $\nearrow$ später)

Satz I.8 (Konstruktion von Wahrscheinlichkeitsmaßen durch Dichten)
Sei $(\Omega, \mathscr{F})$ ein Eriegnisraum.

- Ω abzählbar, $\mathscr{F} = \mathscr{P}(\Omega)$: Sei $\rho = (\rho(\omega))_{\omega\in\Omega}$ eine Folge in $[0,1]$ in $\sum_{\omega\in\Omega}\rho(\omega) = 1$, dann definiert

$$\mathbb{P}(A) = \sum_{\omega\in A}\rho(\omega), A \in \mathscr{F}$$

  ein (diskretes) Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathscr{F})$. ρ wird als Zähldichte bezeichnet.

- Umgekehrt definiert jedes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathscr{F})$ definiert Folge $\rho(\omega) = \mathbb{P}(\{\omega\})$, $\omega \in \Omega$ eine Folge ρ mit den obigen Eigenschaften.

- $\Omega \subset \mathbb{R}^n, \mathscr{F} = \mathscr{B}(\Omega)$: Sei $\rho : \Omega \to [0,\infty)$ eine Funktion, sodass

  1. $\int_\Omega \rho(x)\,\mathrm{d}x = 1$
  2. $\{x \in \Omega\colon \rho(x) \leq c\} \in \mathscr{B}(\Omega)$ für alle $c > 0$

  dann definiert ρ ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathscr{F})$ durch

$$\mathbb{P}(A) = \int_A \rho(x)\,\mathrm{d}x = \int_A \rho\,\mathrm{d}\lambda, \quad A \in \mathscr{B}(\Omega).$$

  Das Integral interpretieren wir stets als Lebesgue-Integral bzw. Lebesgue-Maß λ. ρ bezeichnen wir als Dichte, Dichtefunktion/Wahrscheinlichkeitsdichte von $\mathbb{P}$ und nennen ein solches $\mathbb{P}$ (absolut) stetig (bzgl. denn Lebesgue-Maß).

Anmerkung (English)
Zähldichte heißt im eng. pmf, Dichtefunktion = pdf und später Verteilungsfunktion=cdf - c steht für cummulative.

Beweis.
- Der diskrete Fall ist klar.
- Im stetigen Fall folgt die Bahuptung aus den bekannten Eigenschaften des Lebesgue-Integrals ($\nearrow$ Schilling MINT, Lemma 8.9) □

▶ Bemerkung

- Die eineindeutige Beziehung zwischen Dichte und Wahrscheinlichkeitsmaß überträgt sich nicht auf den stetigen Fall.
  - Nicht jedes Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{B}(\Omega))$ mit $\Omega \subset \mathbb{R}^n$ besitzt eine Dichte.
  - Zwei Dichtefunktionen definieren dasselbe Wahrscheinlichkeitsmaß, wenn sie sich nur auf einer Menge vom Lebesgue-Maß 0 unterscheiden.
- Jede auf $\Omega \subset \mathbb{R}^n$ definierte Dichtefunktion ρ lässt sich auf ganz $\mathbb{R}^n$ fortsetzen durch $\rho(x) =$

0 mit $x \notin \Omega$. Das erzeugte Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathscr{B}(\Omega))$ lässt mit dem Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{B}(\Omega))$ identifizieren.

- Mittels Dirac-Maß δ_x können auch jedes diskrete Wahrscheinlichkeitsmaß auf $\Omega \subset \mathbb{R}^n$ als Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ intepretieren

$$\mathbb{P}(A) = \sum_{\omega \in A} \rho(\omega) = \int_A \mathrm{d}\left(\sum_{\omega\in\Omega} \rho(\omega)\delta_\omega\right)$$

stetige und diskrete Wahrscheinlichkeitsmaße lassen sich kombinieren z.B.

$$\mathbb{P}(A) = \frac{1}{2}\delta_0 + \frac{1}{2}\int_A \mathbb{1}_{[0,1]}(x)\,\mathrm{d}x \quad \forall A \in \mathscr{B}(\mathbb{R})$$

ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Abschließend erinnern wir uns an:

Satz I.9 (Eindeutigkeitssatz für Wahrscheinlichkeitsmaße)
Sei $(\Omega, \mathscr{F})$ Ereignisraum und $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$. Sei $\mathscr{F} = \sigma(\mathscr{G})$ für ein $\cap$-stabiles Erzeugendensystem $\mathscr{G} \subset \mathcal{P}(\Omega)$. Dann ist $\mathbb{P}$ bereits durch seine Einschränkung $\mathbb{P}_{|\mathscr{G}}$ eindeutig bestimmt.

Beweis. ↗ Schilling MINT, Satz 4.5. □

Insbesondere definiert z.B.

$$\mathbb{P}([0,a]) = \int_0^a \lambda e^{-\lambda x}\,\mathrm{d}x = 1 - e^{-\lambda a}, a > 0$$

bereits die Exponentialverteilung aus Beispiel I.7.

Definition I.10 (Gleichverteilung)
Ist Ω endlich, so heißt das Wahrscheinlichkeitsmaß mit konstanter Zähldichte $\rho(\omega) = {}^1/_{|\Omega|}$ die (diskrete) Gleichverteilung auf Ω und wird mit $\mathrm{U}(\Omega)$ notiert (U = Uniform). Ist $\Omega \subset \mathbb{R}^n$ eine Borelmenge mit Lebesgue-Maß $0 < \lambda^n(\Omega) < \infty$, so heißt das Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{B}(\Omega))$ mit konstanter Dichtefunktion $\rho(x) = {}^1/_{\lambda^n(x)}$, die (stetige) Gleichverteilung auf Ω. Sie wird ebenso mit $\mathrm{U}(\Omega)$ notiert.

Wahrscheinlichkeitsräume

Definition I.11 (Wahrscheinlichkeitsraum)
Ein Tripel $(\Omega, \mathscr{F}, \mathbb{P})$ mit $\Omega, \mathscr{F}$ Ereignisraum und $\mathbb{P}$ Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$, nennen wir Wahrscheinlichkeitsraum.

2. Zufallsvariablen

Zufallsvariablen dienen dazu von einem gegebenen Ereignisraum $(\Omega, \mathscr{F})$ zu einem Modellausschnitt $\Omega', \mathscr{F}'$ überzugehen. Es handelt sich also um Abbildungen $X : \Omega \to \Omega'$. Damit wir auch jedem Ereignis in $\mathscr{F}'$ eine Wahrscheinlichkeit zuordnen können, benötigen wir

$$A' \in \mathscr{F}' \Rightarrow X^{-1}(A') \in \mathscr{F}$$

d.h. X sollte messbar sein.

Definition I.12 (Zufallsvariable)
Seien $(\Omega, \mathscr{F})$ und $(\Omega', \mathscr{F}')$ Ereignisräume. Dann heißt jede messbare Abbildung

$$X : \Omega \to \Omega'$$

Zufallsvariable (von $(\Omega, \mathscr{F})$) nach $(\Omega', \mathscr{F}')$/ auf $(\Omega, \mathscr{F})$ oder Zufallselement.

■ **Beispiel I.13**

1. Ist Ω abzählbar und $\mathscr{F} = \mathcal{P}(\Omega)$, so ist jede Abbildung $X : \Omega \to \Omega'$ messbar und damit eine Zufallsvariable.
2. Ist $\Omega \subset \mathbb{R}^n$ und $\mathscr{F} = \mathscr{B}(\Omega)$, so ist jede stetige Funktion $X : \Omega \to \mathbb{R}$ messbar und damit eine Zufallsvariable.

Satz I.14
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable von $(\Omega, \mathscr{F})$ nach $(\Omega', \mathscr{F}')$. Dann definiert

$$\mathbb{P}'(A') := \mathbb{P}\left(X^{-1}(A')\right) = \mathbb{P}\left(\{X \in A'\}\right), \quad A' \in \mathscr{F}'$$

ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathscr{F}')$, welches wir als Wahrscheinlichkeitsverteilung von X unter $\mathbb{P}$ bezeichnen.

Beweis. Aufgrund der Messbarkeit von X ist die Definition sinnvoll. Zudem gelten

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

und für $A'_1, A'_2, \dots \in \mathscr{F}'$ paarweise disjunkt.

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

da auch $X^{-1}A'_1, X^{-1}A'_2, \dots$ paarweise disjunkt

$$\mathbb{P}'\left(\bigcup_{i\geq 1} A'_i\right) = \sum_{i\geq 1} \mathbb{P}'(A'_i).$$

Also ist $\mathbb{P}'$ ein Wahrscheinlichkeitsmaß. □

▶ **Bemerkung**

- Aus Gründen der Lesbarkeit schreiben wir in der Folge $\mathbb{P}(X \in A) = \mathbb{P}(\{\omega\colon X(\omega) \in A\})$
- Ist X die Identität, so fallen die Begriffe Wahrscheinlichkeitsmaß und Wahrscheinlichkeitsverteilung zusammen.
- In der (weiterführenden) Literatur zu Wahrscheinlichkeitstheorie wird oft auf die Angabe eines zugrundeliegenden Wahrscheinlichkeitsraumes verzichtet und stattdessen eine "Zufallsvariable mit Verteilung $\mathbb{P}$ auf Ω" eingeführt. Gemeint ist (fast) immer X als Identität auf $(\Omega, \mathscr{F}, \mathbb{P})$ mit $\mathscr{F} = \mathcal{P}(\Omega)/\mathscr{B}(\Omega)$.
- Für die Verteilung von X unter $\mathbb{P}$ schreibe $\mathbb{P}_X$ und $X \sim \mathbb{P}_X$ für die Tatsache, dass X gemäß $\mathbb{P}_X$ verteilt ist.

Definition I.15 (identisch verteilte, reelle Zufallsvariablen)
Zwei Zufallsvariablen sind identisch verteilt, wenn sie dieselbe Verteilung haben. Von besonderen Interesse sind für uns die Zufallsvariablen, die nach $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ abbilden, sogenannte reelle Zufallsvariablen.

Da die halboffenen Intervalle $\mathscr{B}(\mathbb{R})$ erzeugen, ist die Verteilung eine reelle Zufallsvariable durch die Werte $(-\infty, c], c \in \mathbb{R}$ eindeutig festgelegt.

Definition I.16 ((kumulative) Verteilungsfunktion von $\mathbb{P}$)
Sei $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P})$ Wahrscheinlichkeitsraum, so heißt

$$F : \mathbb{R} \to [0,1] \text{ mit } x \mapsto \mathbb{P}((-\infty, x])$$

(kumulative) Verteilungsfunktion von $\mathbb{P}$.
Ist X eine reelle Zufallsvariable auf beliebigen Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$, so heißt

$$F : \mathbb{R} \to [0,1] \text{ mit } x \mapsto \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x])$$

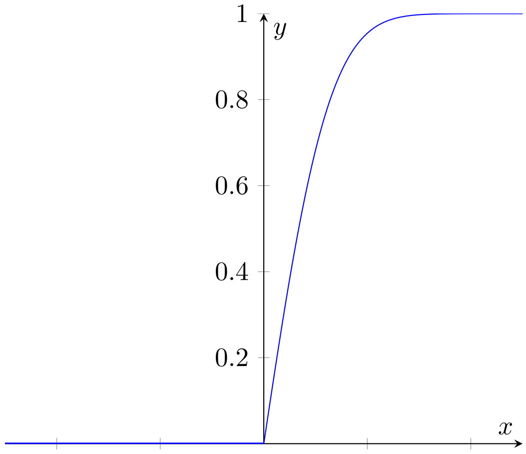
die (kumulative) Verteilungsfunktion von X.

■ **Beispiel I.17**
Sei $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P})$ mit $\mathbb{P}$ Exponentialverteilung mit Parameter $\lambda > 0$

$$\mathbb{P}(A) = \int_{A \cap [0,\infty)} \lambda e^{-\lambda x} \,\mathrm{d}x \quad A \in \mathscr{B}(\mathbb{R}).$$

Dann ist

$$F(x) = \mathbb{P}((-\infty, x)) = \begin{cases} 0 & x \leq 0 \\ \int_0^x \lambda e^{-\lambda y} \,\mathrm{d}y = 1 - e^{-\lambda x} & x > 0 \end{cases}.$$



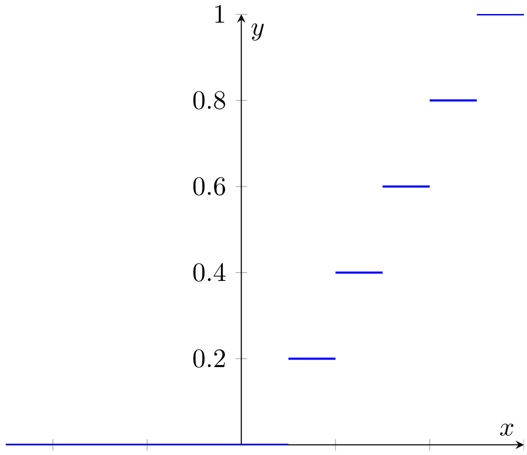
■ **Beispiel I.18**

Das Würfeln mit einem fairen, sechsseitigen Würfel kann mittels einer reellen Zufallsvariablen

$$X : \{1, 2, \dots, 6\} \to \mathbb{R} \text{ mit } x \mapsto x$$

modelliert werden. Es folgt als Verteilungsfunktion

$$\begin{aligned} F(x) &= \mathbb{P}'(X \leq x) = \mathbb{P}(X^{-1}(-\infty, x]) = \mathbb{P}((-\infty, x]) \\ &= \frac{1}{6} \sum_{i=1}^{6} \mathbb{1}_{i \leq x}. \end{aligned}$$



Allgemein:

Satz I.19
Ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ und F die zugehörige Verteilungsfunktion, so gelten

1. F ist monoton wachsend
2. F ist rechtsseitig stetig
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$

Umgekehrt existiert zu jeder Funktion $F : \mathbb{R} \to [0,1]$ mit Eigenschaften 1-3 eine reelle Zufallsvariable auf $((0,1), \mathscr{B}((0,1)), \mathrm{U}((0,1))$ mit Verteilungsfunktion F.

Beweis. Ist F Verteilungsfunktion, so folgt mit Satz I.4

$$x \leq y \Rightarrow F(x) = \mathbb{P}((-\infty, x]) \overset{I.4.3}{\leq} \mathbb{P}((-\infty, y]) = F(y)$$

und

$$\lim_{x \searrow c} F(x) = \lim_{x \searrow c} \mathbb{P}((-\infty, x]) \overset{\sigma\text{-Stetigkeit}}{=} \mathbb{P}((-\infty, c]) = F(c)$$

sowie

$$\lim_{x \to -\infty} F(x) \overset{I.4.5}{=} \mathbb{P}(\varnothing) \overset{I.4.1}{=} 0$$
$$\lim_{x \to \infty} F(x) \overset{I.4.5}{=} \mathbb{P}(\mathbb{R}) = 1.$$

Umgekehrt wähle

$$X(u) := \inf\{x \in \mathbb{R} \colon F(x) \geq u\}, \quad u \in (0,1)$$

Dann ist X eine “linksseitige Inverse” von F (auch Quantilfunktion / verallgemeinerte Inverse). Wegen 3 gilt:

$$-\infty < X(u) < \infty$$

und zudem

$$\{X \leq x\} = (0, F(x)) \cap (0,1) \in \mathscr{B}((0,1)).$$

Da diese halboffene Mengen ein Erzeugendensystem von $\mathscr{B}(\mathbb{R})$ bilden, folgt bereits die Messbarkeit von X, also ist X eine ZV. Insbesondere hat die Menge $\{X \leq x\}$ gerade Lebesgue-Maß $F(x)$ und damit hat X die Verteilungsfunktion F. □

Folgerung I.20
Ist $\mathbb{P}$ Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ und F die zugehörige Verteilungsfunktion. Dann besitzt $\mathbb{P}$ genau eine Dichtefunktion ρ, wenn F stetig differenzierbar ist, denn dann gelten

$$F(x) = \int_{-\infty}^{x} \rho(x) \, \mathrm{d}x, \text{ bzw. } \rho(x) = F'(x)$$

Beweis. Folgt aus Satz I.8, der Satz II.6 der Verteilungsfunktion und dem Eindeutigkeitssatz **???**. □

Kapitel II

Erste Standardmodelle der Wahrscheinlichkeitstheorie

Diskrete Verteilungen

1. Diskrete Gleichverteilungen

Erinnerung:

▶ **Erinnerung (Definition I.I.10)**
Ist Ω endlich, so heißt das Wahrscheinlichkeitsmaß mit Zähldichte

$$\rho(\omega) = \frac{1}{|\Omega|} \quad , \omega \in \Omega$$

(diskrete) Gleichverteilung auf $\Omega \to U(\Omega)$

Es gilt das für jedes $A \in \mathcal{P}(\Omega)$

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

Anwendungsbeispiele sind faires Würfeln, fairer Münzwurf, Zahlenlotto, ...

2. Urnenmodelle

Ein "Urnenmodell" ist eine abstrakte Darstellung von Zufallsexperimenten, bei denen zufällig Stichproben aus einer gegebenen Menge "gezogen" werden.

Definition (Urne)
Eine Urne ist ein Behältnis in welchem sich farbige/nummerierte Kugeln befinden, die ansonsten ununterscheidbar sind.

Aus der Urne ziehe man blind/zufällig eine oder mehrere Kugeln und notiere Farbe/Zahl.

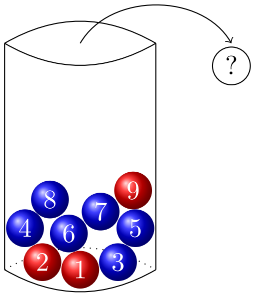


Abbildung II.1: Urnenmodell

2.1. Urnenmodell mit Zurücklegen: Multinomial-Verteilung

Gegeben: Urne mit N Kugeln, verschiedenfarbig mit Farben aus E, $|E| \geq 2$

Ziehe: n Stichproben/Kugeln, wobei nach jedem Zug die Kugel wieder zurückgelegt wird. Uns interessiert die Farbe in jedem Zug, setze also

$$\Omega = E^n \text{ und } \mathscr{F} = \mathcal{P}(\Omega)$$

Zur Bestimmung eines geeigneten Wahrscheinlichkeitsmaßes, nummerieren wir die Kugeln mit $1, \dots, N$, so dass alle Kugeln der Farbe $a \in E$ eine Nummer aus $F_a \subset \{1, \dots, N\}$ tragen. Würden wir die Nummern notieren, so wäre

$$\bar{\Omega} = \{1, \dots, N\}^n \text{ und } \overline{\mathscr{F}} = \mathcal{P}(\overline{\Omega})$$

und wir könnten die Gleichverteilung $\overline{\mathbb{P}} = \mathrm{U}(\overline{\Omega})$ als Wahrscheinlichkeitsmaß für einen einzelnen Zug verwenden. Für den Übergang zu Ω konstruieren wir Zufallsvariablen. Die Farbe im i-ten Zug wird beschrieben durch

$$X_i : \overline{\Omega} \to E \text{ mit } \overline{\omega} = (\overline{\omega}_1, \dots, \overline{\omega}_n) \mapsto a \text{ falls } \overline{\omega}_i \in F_a.$$

Der Zufallsvektor

$$X = (X_1, \dots, X_n) : \overline{\Omega} \to \Omega$$

beschreibt dann die Abfolge der Farben. Für jedes $\omega \in \Omega$ gilt dann

$$\{X = \omega\} = F_{\omega_1} \times \dots \times F_{\omega_n} = \bigtimes_{i=1}^{n} F_{\omega_i}$$

und damit

$$\begin{aligned} \mathbb{P}(\{\omega\}) &= \overline{\mathbb{P}}(X^{-1}(\{\omega\})) = \mathbb{P}(X = \omega) \\ &= \frac{|F_{\omega_1}| \cdots |F_{\omega_n}|}{\left|\overline{\Omega}\right|} \\ &= \prod_{i=1}^{n} \frac{|F_{\omega_i}|}{N} =: \prod_{i=1}^{n} \rho(\omega_i) \end{aligned}$$

Zähldichten, die sich als Produkte von Zähldichten schreiben lassen, werden auch als Produktdichten bezeichnet (↗ Abschnitt 2). Sehr oft interessiert bei einem Urnenexperiment nicht die Reihenfolge der gezogenen Farben, sondern nur die Anzahl der Kugeln in Farbe $a \in E$ nach n Zügen. Dies entspricht

$$\hat{\Omega} = \left\{ k = (k_a)_{a \in E} \in \mathbb{N}_0^{|E|} : \sum_{a \in E} k_a = n \right\} \text{ und } \hat{\mathscr{F}} = \mathcal{P}\left(\hat{\Omega}\right)$$

Den Übergang $\Omega \to \hat{\Omega}$ beschreiben wir durch die Zufallsvariablen

$$Y_a(\omega) : \Omega \to \mathbb{N}_0 \text{ mit } \omega = (\omega_1, \ldots, \omega_n) \mapsto \sum_{a \in E} \mathbb{1}_{\{a\}}(\omega_i)$$

und

$$Y = (Y_a)_{a \in E} : \Omega \to \hat{\Omega} = \left\{ k = (k_a)_{a \in E} : \sum_{a \in E} k_a = n \right\}$$

Wir erhalten

$$\begin{aligned} \mathbb{P}(Y = k) &= \mathbb{P}(Y_a = k_a, \ \ a \in E) \\ &= \sum_{\omega \in \Omega : Y(\omega) = k} \prod_{i=1}^{n} \rho(\omega_i) \\ &= \sum_{\omega \in \Omega : Y(\omega) = k} \prod_{a \in E} \rho(a) \\ &= \binom{n}{(k_a)_{a \in E}} \prod_{a \in E} \rho(a)^{k_a}, \end{aligned}$$

wobei

$$\binom{n}{(k_1, \ldots, k_l)} = \begin{cases} \frac{n!}{k_1! \, k_2! \cdots k_l!} & \sum_{i=1}^{l} k_i = n \\ 0 & \text{sonst} \end{cases}$$

der Multinomialkoeffizient ist, welcher die Anzahl der Möglichkeiten beschreibt, n Objekte in l Gruppen aufzuteilen, so dass Gruppe i gerade k_i Objekte beinhaltet.

Definition II.1 (Multinomialverteilung)
Sei $l > 2, p = (p_1, \ldots, p_l)$ eine Zähldichte und $n \in \mathbb{N}$, dann heißt die Verteilung auf $\left\{ k = (k_i)_{i=1,\ldots,l} \in \mathbb{N}_0^l : \sum_{i=1}^{l} k_i = n \right\}$ mit Zähldichte

$$m((k_1, \ldots, k_l)) = \binom{n}{k_1, \ldots, k_l} \prod_{i=1}^{l} p_i^{k_i}$$

Multinomialverteilung mit Parametern n und p. Wir schreiben auch $\text{Multi}(n, p)$.

■ **Beispiel II.2**
Eine Urne enthalte nur schwarze “1” und weiße “0” Kugeln, d.h. $E = \{0, 1\}$, und es sei $\rho(1) = p$ gerade die Proportion der schwarzen Kugeln (= Wahrscheinlichkeit bei einem Zug schwarz zu ziehen), dann ist Wahrscheinlichkeit in n Zügen k-mal schwarz zu ziehen:

$$\binom{n}{k} \prod_{i=0,1} \rho(i)^{k_i} = \binom{n}{k} p^k (1-p)^{n-k}.$$

Ein solches (wiederholtes) Experiment mit nur zwei möglichen Ereignissen und fester Wahrscheinlichkeit $p \in [0, 1]$ für eines der Ergebnisse nennen wir auch (wiederholtes) Bernoulliexperiment.

Definition II.3 (Binomialverteilung, Bernoulliverteilung)
Sei $p \in [0,1]$ und $n \in \mathbb{N}$, dann heißt die Verteilung mit Zähldichte

$$\rho(k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ mit } k \in \{0, 1, \dots, n\}.$$

Binomialverteilung auf $\{0, \dots, n\}$ mit Parameter p (auch Erfolgswahrscheinlichkeit). Wir schreiben auch $\operatorname{Bin}(n,p)$. Im Fall $n = 1$ nennen wir die Verteilung mit Zähldichte

$$\rho(0) = 1 - p \text{ und } \rho(1) = p$$

auch Bernoulliverteilung mit Parameter p und schreiben $\operatorname{Bernoulli}(p)$.

Urnenmodell ohne Zurücklegen: Hypergeometrische Verteilung
Gegeben: Urne mit N Kugeln verschiedener Farben aus E,

$$|E| \geq 2.$$

Es werden $n \leq N$ Stichproben entnommen, wobei die gezogenen Kugeln nicht in die Urne zurückgelegt werden.

2.2. Urnenmodell ohne Zurücklegen: Hypergeometrische Verteilung

Gegeben: Urne mit N Kugeln verschiedener Farben aus E, $|E| \geq 2$. Es werden $n \leq N$ Stichproben entnommen, wobei die gezogenen Kugeln nicht in die Urne zurückgelegt werden.

■ **Beispiel II.4**
Eine Urne enthalte S schwarze "1" und W weiße Kugeln "0" Kugeln, $(E = \{0,1\}, S + W = N)$. Dann ist die Wahrscheinlichkeit in n Zügen ohne Zurücklegen gerade s schwarze und w weiße Kugeln zu ziehen

$$\rho(w) = \frac{\binom{W}{w}\binom{S}{s}}{\binom{N}{n}}, \quad 0 \leq s \leq S, 0 \leq w \leq W, s + w = n, S + W = N.$$

Beweis. Hausaufgabe 2.3! □

Definition II.5 (Hypergeometrische Verteilung)
Seien $N \in \mathbb{N}, W \leq N, n \leq N$, dann heißt die Verteilung auf $\{0, \dots, n\}$ mit Zähldichte

$$\rho(w) = \frac{\binom{W}{w}\binom{N-W}{n-w}}{\binom{N}{n}}, \quad w = \max\{0, n = N + W\}, \dots, \min\{W, n\},$$

die Hypergeometrische Verteilung mit Parametern N, W, n. Wir schreiben $\operatorname{Hyper}(N, W, n)$.

3. Poisson-Approximation und Poisson-Verteilung

$\operatorname{Bin}(n,p)$ ist zwar explizit und elementar definiert, jedoch für große n mühsam auszuwerten. Für seltene Ereignisse (n groß, p klein) verwende daher:

Satz II.6 (Poisson-Approximation)
Sei $\lambda > 0$ und $(p_n)_{n\in\mathbb{N}}$ eine Folge in $[0,1]$ mit

$$np_n \to \lambda, \quad n \to \infty.$$

Dann gilt $\forall k \in \mathbb{N}_0$ für die Zähldichte der $\mathrm{Bin}(n, p_n)$-Verteilung

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Beweis. Sei $k \in \mathbb{N}_0$ fix, dann

$$\begin{aligned}\binom{n}{k} = \frac{n!}{k!(n-k)!} &= \frac{n^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k} \\ &= \frac{n^k}{k!} \cdot 1 \cdot (1 - \frac{1}{n} \cdots \frac{k-1}{n}) \\ &\overset{n\to\infty}{\sim} \frac{n^k}{k!},\end{aligned}$$

wobei $a(l) \overset{l\to\infty}{\sim} b(l) \Leftrightarrow \frac{a(l)}{b(l)} \xrightarrow{l\to\infty} 1$. Damit

$$\begin{aligned}\binom{n}{k} p^k (1-p)^{n-k} &\overset{n\to\infty}{\sim} \frac{n^k}{k!} p_n^k (1-p_n)^{n-k} \\ &\overset{n\to\infty}{\sim} \frac{\lambda^k}{k!}(1-p_n)^n \\ &= \frac{\lambda^n}{k!}\left(1 - \frac{np_n}{n}\right)^n \\ &\xrightarrow{n\to\infty} \frac{\lambda^n}{k!} e^{-\lambda}.\end{aligned}$$

Der letzte Schritt hat eine verwandte Ungleichung zur Bernoulli-Ungleichung genutzt

$$\lim_{n\to\infty} (1 - \frac{np_n}{n})^n \leq \lim_{n\to\infty} e^{\frac{-n^2 p_n}{n}} = \lim_{n\to\infty} e^{-np_n} \sim e^{-\lambda}$$

(Falls jemand eine Idee hat wie man die Ungleichung beweist, bitte bescheid sagen, oder einfach hinzufügen.) □

Der erhaltene Grenzwert liefert die Zähldichte auf $\mathbb{N}_0$, denn

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

Definition II.7 (Poissonverteilung)
Sei $\lambda > 0$. Dann heißt das auf $(\mathbb{N}_0, \mathbb{P}(\mathbb{N}_0))$ definierte Wahrscheinlichkeitsmaß mit

$$\mathbb{P}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k \in \mathbb{N}_0,$$

Poissonverteilung mit Parameter λ. Schreibe $\mathrm{Poisson}(\lambda)$.

Die Poissonverteilung ist ein natürliches Modell für die Anzahl von zufälligen, seltenen Ereignissen (z.B. Tore im Fußballspiel, Schadensfälle einer Versicherung, ...).

Kapitel III

Bedingte Wahrscheinlichkeiten und (Un-)abhängigkeit

1. Bedingte Wahrscheinlichkeiten

■ **Beispiel III.1**

Das Würfeln mit zwei fairen, sechsseitigen Würfeln können wir mit

$$\Omega = \{(i,j) \colon i,j \in \{1,\dots,6\}\}$$

und $\mathbb{P} = \mathrm{U}(\Omega)$. Da $|\Omega| = 36$ gilt also

$$\mathbb{P}(\{\omega\}) = \frac{1}{36} \quad \forall \omega \in \Omega.$$

Betrachte das Ereignis

$$A = \{(i,j) \in \Omega \colon i + j = 8\},$$

dann folgt

$$\mathbb{P}(A) = \frac{5}{36}.$$

Werden die beiden Würfe nacheinander ausgeführt, so kann nach dem ersten Wurf eine Neubewertung der Wahrscheinlichkeit von A erfolgen.
Ist z.B.

$$B = \{(i,j) \in \Omega, i = 4\}$$

eingetreten, so kann die Summe 8 nur durch eine weitere 4 realisiert werden, also mit Wahrscheinlichkeit

$$\frac{1}{6} = \frac{|A \cap B|}{|B|}.$$

Das Eintreten von B führt also dazu, dass das Wahrscheinlichkeitsmaß $\mathbb{P}$ durch ein neues Wahrscheinlichkeitsmaß $\mathbb{P}_B$ ersetzt werden muss. Hierbei sollte gelten:

Renormierung: $\mathbb{P}_B(\Omega) = 1$ (R)

Proportionalität: Für alle $A \in \mathscr{F}$ mit $A \subseteq B$ gilt $\mathbb{P}_B(A) = c_B \mathbb{P}(A)$ mit einer Konstante c_B. (P)

Lemma III.2
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $B \in \mathscr{F}$ mit $\mathbb{P}(B) > 0$. Dann gibt es genau ein Wahrscheinlichkeitsmaß $\mathbb{P}_B$ auf $(\Omega, \mathscr{F})$ mit den Eigenschaften (R) und (P). Dieses ist gegeben durch

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \forall A \in \mathscr{F}.$$

Beweis. Offenbar erfüllt $\mathbb{P}_B$ wie definiert (R) und (P). Umgekehrt erfüllt $\mathbb{P}$ (R) und (P). Dann folgt für $A \in \mathscr{F}$:

$$\mathbb{P}_B(A) = \mathbb{P}_B(A \cap B) + \underbrace{\mathbb{P}_B(A \setminus B)}_{=0,\text{ wegen (R)}} \overset{(P)}{=} c_B \mathbb{P}(A \cap B).$$

Für $A = B$ folgt zudem aus (R)

$$1 = \mathbb{P}_B(B) = c_B \mathbb{P}(B)$$

also $c_B = \mathbb{P}(B)^{-1}$. □

Definition III.3 (Bedingte Wahrscheinlichkeit)
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $B \in \mathscr{F}$ mit $\mathbb{P}(B) > 0$. Dann heißt

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ mit } A \in \mathscr{F}$$

die bedingte Wahrscheinlichkeit von A gegeben B. Falls $\mathbb{P}(B) = 0$, setze

$$\mathbb{P}(A \mid B) = 0 \qquad \forall A \in \mathscr{F}$$

■ **Beispiel III.4**
In der Situation Beispiel III.1 gilt

$$A \cap B = \{(4, 4)\}$$

und damit

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

Aus Definition III.3 ergibt sich

Lemma III.5 (Multiplikationsformel)
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \mathscr{F}$. Dann gilt

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1) \dots \mathbb{P}(A_n \mid A_1 \cap \dots \cap A_{n-1})$$

Beweis. Ist $\mathbb{P}(A_1 \cap \dots \cap A_n) = 0$, so gilt auch $\mathbb{P}(A_n \mid \bigcap_{i=1}^{n-1} A_i) = 0$. Andernfalls sind alle Faktoren der rechten

Seite ungleich Null und

$$\begin{aligned}\mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\dots\mathbb{P}(A_n \mid \bigcap_{i=1}^{n-1} A_i) &= \mathbb{P}(A_1)\cdot\frac{\mathbb{P}(A_1\cap A_2)}{\mathbb{P}(A_1)}\cdots\frac{\mathbb{P}(\bigcap_{i=1}^{n} A_i)}{\mathbb{P}(\bigcap_{i=1}^{n-1} A_i)}\\ &= \mathbb{P}(\bigcap_{i=1}^{n} A_i)\end{aligned}$$

□

Stehen die A_i in Lemma III.5 in einer (zeitlichen) Abfolge, so liefert Formel einen Hinweis wie Wahrscheinlichkeitsmaße für Stufenexperimente konstruiert werden können. Ein Stufenexperiment aus n nacheinander ausgeführten Teilexperimenten lässt sich als Baumdiagramm darstellen.

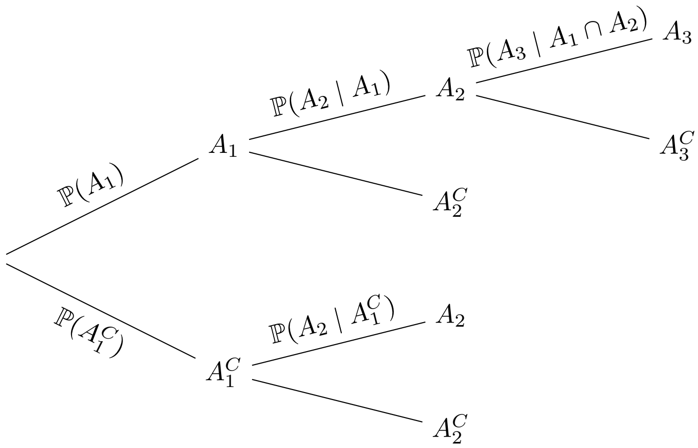


Abbildung III.1: Lemma III.5

Satz III.6 (Konstruktion des Wahrscheinlichkeitsmaßes eines Stufenexperiments)
Gegeben seinen n Ergebnisräume $\Omega_i = \{\omega_i(1),\dots,\omega_i(k)\}, k \in \mathbb{N}\cup\{\infty\}$ und es sei $\Omega = \bigtimes_{i=1}^n \Omega_i$ der zugehörige Produktraum. Weiter seinen $\mathscr{F}_i$ σ-Algebren auf Ω_i und $\mathscr{F} = \bigotimes_{i=1}^n \mathscr{F}_i$ die Produkt-σ-Algebra auf Ω. Setze $\omega = (\omega_1,\dots,\omega_n)$ und

$$\begin{aligned}[\omega_1,\dots,\omega_m] := \{\omega_1\}\times\dots\times\{\omega_m\}\times\Omega_{m+1}\times\dots\times\Omega_n, \quad m \le n\\ \mathbb{P}(\{\omega_m\}[\omega_1,\dots,\omega_{m-1}])\end{aligned}$$

für die Wahrscheinlichkeit in der m-ten Stufe des Experiments ω_m zu beobachten, falls in den vorausgehenden Stufen $\omega_1,\dots,\omega_{m-1}$ beobachten wurden. Dann definiert

$$\mathbb{P}(\{\omega\}) := \mathbb{P}(\{\omega_1\})\prod_{m=2}^{n}\mathbb{P}(\{\omega_m\} \mid [\omega_1,\dots,\omega_{m-1}])$$

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

Beweis. Nachrechnen! □

■ **Beispiel III.7 (Polya-Urne)**
Gegeben sei eine Urne mit s schwarzen und w weißen Kugeln. Bei jedem Zug wird die gezogene

Kugel zusammen mit $c \in \mathbb{N}_0 \cup \{-1\}$ weiteren Kugeln derselben Farbe zurückgelegt.

- $c = 0$: Urnenmodell mit Zurücklegen
- $c = -1$: Urnenmodell ohne Zurücklegen

Beide haben wir schon in Kapitel 2.2 gesehen.
Sei deshalb $c \in \mathbb{N}$. (Modell für zwei konkurrierende Populationen) Ziehen wir n-mal, so haben wir ein n-Stufenexperiment mit

$$\Omega = \{0,1\}^n \text{ mit } 0 = \text{“weiß”}, 1 = \text{“schwarz”} \quad (\Omega_i = \{0,1\})$$

Zudem gelten im ersten Schritt

$$\mathbb{P}(\{0\}) = \frac{w}{s+w} \text{ und } \mathbb{P}(\{1\}) = \frac{s}{s+w}$$

sowie

$$\mathbb{P}(\{\omega_m\} \mid [\omega_1, \dots \omega_{m-1}]) = \begin{cases} \frac{w+c\left(m-1-\sum_{i=1}^{m-1}\omega_i\right)}{s+w+c(m-1)} & \omega_m = 0 \\ \frac{s+c\sum_{i=1}^{m-1}\omega_i}{s+w+c(m-1)} & \omega_m = 1 \end{cases}$$

Mit Satz III.6 folgt als Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{P}(\Omega))$

$$\begin{aligned} \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) &= \mathbb{P}(\{\omega_1\}) \prod_{m=2}^{n} \mathbb{P}(\{\omega_m\} \mid [\omega_1, \dots, \omega_{m-1}]) \\ &= \frac{\prod_{i=0}^{l-1}(s+c\cdot i)\prod_{i=0}^{n-l-1}(w+c\cdot j)}{\prod_{i=0}^{n}(s+w+c\cdot i)} \text{ mit } l = \sum_{i=1}^{n}\omega_i. \end{aligned}$$

Definiere wir nun die Zufallsvariable

$$S_n : \Omega \to \mathbb{N}_0 \text{ mit } (\omega_1, \dots, \omega_n) \mapsto \sum_{i=1}^{n}\omega_i$$

welche die Anzahl der gezogenen schwarzen Kugeln modelliert, so folgt

$$\mathbb{P}(S_n = l) = \binom{n}{l} \frac{\prod_{i=0}^{l-1}(s+c\cdot i)\prod_{j=0}^{n-l-1}(w+c\cdot j)}{\prod_{i=0}^{n}(s+w+c\cdot i)}$$

Mittels $a := {}^{s}/{}_{c}, b := {}^{w}/{}_{c}$ folgt

$$\begin{aligned} \mathbb{P}(S_n = l) &= \binom{n}{l} \frac{\prod_{i=0}^{l-1}(-a-i)\prod_{j=0}^{n-l-1}(-b-j)}{\prod_{i=0}^{n}(-a-b-i)} = \frac{\binom{-a}{l}\binom{-b}{n\cdot l}}{\binom{-a-b}{n}} \\ &\text{mit } l \in \{0, \dots, n\} \end{aligned}$$

Dies ist die Polya-Verteilung auf $\{0, \dots, n\}, n \in \mathbb{N}$ mit Parametern $a, b > 0$.

■ **Beispiel III.8**
Ein Student beantwortet eine Multiple-Choice-Frage mit 4 Antwortmöglichkeiten, eine davon ist

richtig. Er kennt die richtige Antwort mit Wahrscheinlichkeit $2/3$. Wenn er diese kennt, so wählt er diese aus. Andernfalls wählt er zufällig (gleichverteilt) eine Antwort. Betrachte

$$\begin{aligned} W &= \{\text{richtige Antwort gewusst}\} \\ R &= \{\text{Richtige Antwort gewählt}\} \end{aligned}$$

Dann gilt

$$\mathbb{P}(W) = \frac{2}{3}, \mathbb{P}(R \mid W) = 1, \mathbb{P}(R \mid W^C) = \frac{1}{4}$$

Angenommen, der Student gibt die richtige Antwort. Mit welcher Wahrscheinlichkeit hat er diese gewusst? $\longrightarrow \mathbb{P}(W \mid R) = \ ?$

Satz III.9
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $\Omega = \bigcup_{i \in I} B_i$ eine höchstens abzählbare Zerlegung in paarweise disjunkte Ereignisse $B_i \in \mathscr{F}$.

1. Satz von der totalen Wahrscheinlichkeit: Für alle $A \in \mathscr{F}$ gilt

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i) \qquad \text{(totale Wahrscheinlichkeit)}$$

2. Satz von Bayes: Für alle $A \in \mathscr{F}$ mit $\mathbb{P}(A) > 0$ und alle $k \in I$

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{i \in I} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)} \qquad \text{(Bayes)}$$

Beweis. 1. Es gilt:

$$\sum_{i \in I} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i) \overset{\text{Def}}{=} \sum_{i \in I} \frac{\mathbb{P}(A \cap B_i)}{\cancel{\mathbb{P}(B_i)}} \cancel{\mathbb{P}(B_i)} = \sum_{i \in I} \mathbb{P}(A \cap B_i) \overset{\sigma-Add.}{=} \mathbb{P}(A)$$

2.

$$\mathbb{P}(B_k \mid A) \overset{\text{Def}}{=} \frac{\mathbb{P}(A \cap B_k)}{\mathbb{P}(A)} \overset{\text{Def}}{=} \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\mathbb{P}(A)}$$

also folgt (b) aus (a). □

■ **Beispiel III.10**
In der Situation von Definition III.3 folgt mit Satz III.9 (totale Wahrscheinlichkeit)

$$\begin{aligned} \mathbb{P}(R) &= \mathbb{P}(R \mid W)\mathbb{P}(W) + \mathbb{P}(R \mid W^C)\mathbb{P}(W^C) \\ &= 1 \cdot \frac{2}{3} + \frac{1}{4}\frac{1}{3} = \frac{3}{4} \end{aligned}$$

und mit Satz III.9 (Bayes)

$$\mathbb{P}(W \mid R) = \frac{\mathbb{P}(R \mid W)\mathbb{P}(W)}{\mathbb{P}(R)} = \frac{1 \cdot \frac{2}{3}}{\frac{3}{4}} = \frac{8}{9} \text{ für die gesuchte Wahrscheinlichkeit.}$$

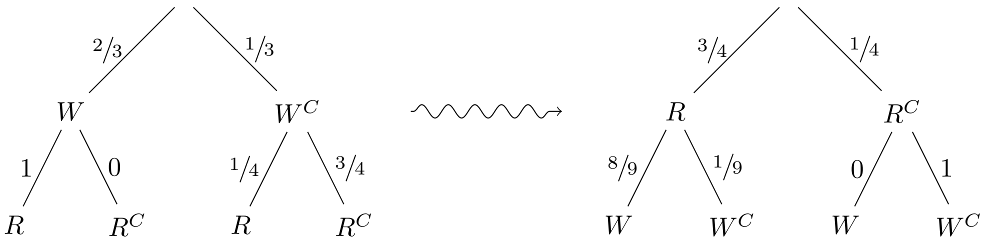


Abbildung III.2: Satz III.3

2. (Un)abhängigkeit

In vielen Fällen besagt die Intuition über verschiedene Zufallsexperimente / Ereignisse, dass diese sich nicht gegenseitig beeinflussen. Für solche $A, B \in \mathscr{F}$ mit $\mathbb{P}(A) > 0, \mathbb{P}(B) > 0$ sollte gelten

$$\mathbb{P}(A \mid B) = \mathbb{P}(A), \quad \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

Definition III.11 ((Stochastische) Unabhängigkeit)
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \in \mathscr{F}$ heißt (stochastisch) unabhängig bezüglich $\mathbb{P}$, falls

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

Wir schreiben auch $A \perp\!\!\!\perp B$.

■ **Beispiel III.12**
Würfeln mit 2 fairen, sechsseitigen Würfeln:

$$\Omega = \{(i,j) \mid i,j \in \{1,\dots,n\}\}, \quad \mathscr{F} = \mathcal{P}(\Omega), \quad \mathbb{P} = \mathrm{U}(\Omega)$$

Betrachte

$$A := \{(i,j) \in \Omega, i \text{ gerade}\}$$
$$B := \{(i,j) \in \Omega, j \leq 2\}.$$

In diesem Fall, erwarten wir intuitiv Unabhängigkeit von A und B.
In der Tat ist

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{1}{3} \text{ und } \mathbb{P}(A \cap B) = \frac{1}{6}$$

was

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

erfüllt. Betrachte nun

$$C := \{(i,j) \in \Omega \mid i + j = 7\}$$
$$D := \{(i,j) \in \Omega \mid i = 6\}$$

dann gilt

$$\mathbb{P}(C) = \frac{1}{6}, \quad \mathbb{P}(D) = \frac{1}{6}$$

und wegen $C \cap D = \{(6,1)\}$ folgt

$$\mathbb{P}(C \cap D) = \frac{1}{36} = \frac{1}{6}\frac{1}{6} = \mathbb{P}(C) \cdot \mathbb{P}(D)$$

C und D sind also stochastisch unabhängig, obwohl eine kausale Abhängigkeit vorliegt!

Definition III.13 (Unabhängigkeit bezüglich $\mathbb{P}$)
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $I \neq \varnothing$ endliche Indexmenge. Dann heißt die Familie $(A_i)_{i \in I}$ von Ereignissen in $\mathscr{F}$ unabhängig bezüglich $\mathbb{P}$, falls für alle $J \subseteq I, J \neq \varnothing$ gilt:

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

Offensichtlich impliziert die Unabhängigkeit einer Familie die paarweise Unabhängigkeit je zweier Familienmitglieder nach Definition III.11. Umgekehrt gilt dies nicht!

■ **Beispiel III.14 (Abhängigkeit trotz paarweiser Unabhängigkeit)**
Betrachte zweifaches Bernoulliexperiment mit Erfolgswahrscheinlichkeit $1/2$, d.h.

$$\Omega = \{0,1\}^2, \quad \mathscr{F} = \mathcal{P}(\Omega), \quad \mathbb{P} = \mathrm{U}(\Omega)$$

sowie

$$A = \{1\} \times \{0,1\} \qquad \text{(Münzwurf: erster Wurf ist Zahl)}$$
$$B = \{0,1\} \times \{1\} \qquad \text{(Münzwurf: zweiter Wurf ist Zahl)}$$
$$C = \{(0,0),(1,1)\} \qquad \text{(beide Würfe haben selbes Ergebnis)}$$

Dann gelten $\mathbb{P}(A) = \frac{1}{2} = \mathbb{P}(B) = \mathbb{P}(C)$ und

$$\begin{aligned}\mathbb{P}(A \cap B) &= \mathbb{P}(\{(1,1)\}) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B)\\ \mathbb{P}(A \cap C) &= \mathbb{P}(\{(1,1)\}) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C)\\ \mathbb{P}(B \cap C) &= \mathbb{P}(\{(1,1)\}) = \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C)\end{aligned}$$

also paarweise Unabhängigkeit. Aber

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{(1,1)\}) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$$

und A, B, C sind nicht stochastisch unabhängig.

Definition III.15 (Unabhängige σ-Algebren)
Seien $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $I \neq \varnothing$ Indexmenge und $(E_i, \mathcal{E}_i)$ Messräume

1. Die Familie $\mathscr{F}_i \subset \mathscr{F}, i \in I$, heißen unabhängig, wenn für die $J \subseteq I, J \neq \varnothing, |J| < \infty$ gilt

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \qquad \text{für beliebige } A_i \in \mathscr{F}_i, i \in J$$

2. Die Zufallsvariable $X_i : (\Omega, \mathscr{F}) \to (E_i, \mathcal{E}_i), i \in I$, heißen unabhängig, wenn die σ-Algebren

$$\sigma(X_i) = X^{-1}(\mathcal{E}_i) = \{\{X_i \in F\} \colon F \in \mathcal{E}_i\}, \quad i \in I$$

unabhängig sind.

Lemma III.16 (Zusammenhang der Definitionen)
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $I \neq \varnothing, A \in \mathscr{F}, i \in I$. Die folgenden Aussagen sind äquivalent:

1. Die Ereignisse $A_i, i \in I$ sind unabhängig.
2. Die σ-Algebren $\sigma(A_i), i \in I$ sind unabhängig.
3. Die Zufallsvariablen $\mathbb{1}_{A_i}, i \in I$ sind unabhängig.

Beweis. Da die Unabhängigkeit über endliche Teilemengen definiert ist, können wir oBdA $I = \{1, \dots, n\}$ annehmen.

- Da $\sigma(\mathbb{1}_{A_i}) = \sigma(A_i)$ folgt die Äquivalenz von 2. und 3. direkt aus Definition III.15.
- Zudem ist 2. $\to$ 1. klar!
- Für 1 $\to$ 2. genügt es zu zeigen, dass

$$A_1, \dots, A_n \text{ unabhängig } \Rightarrow B_1, \dots, B_n \text{ unabhängig mit } B_i \in \left\{\varnothing, A_i, A_i^C, \Omega\right\}.$$

Rekursiv folgt dies bereits aus

$$A_1, \dots, A_n \text{ unabhängig } \Rightarrow B_1, A_2, \dots, A_n \text{ unabhängig mit } B_1 \in \left\{\varnothing, A_1, A_1^C, \Omega\right\}.$$

Für $B_1 \in \{\varnothing, A_1, \Omega\}$ ist dies klar.
Sei also $B_1 = A_1^C$ und $J \subseteq I, J \neq \varnothing$. Falls $1 \notin J$, ist nichts zu zeigen. Sei $1 \in J$, dann gilt mit

$$A = \bigcap_{i \in J, i \neq 1} A_i$$

sicherlich

$$\begin{aligned}
\mathbb{P}\big(A_1^C \cap A\big) &= \mathbb{P}(A \setminus (A_1 \cap A)) \\
&= \mathbb{P}(A) - \mathbb{P}(A_1 \cap A) \\
&= \prod_{i \in J \setminus \{1\}} \mathbb{P}(A_i) - \prod_{i \in J} (A_i) \\
&= (1 - \mathbb{P}(A_1)) \prod_{i \in J \setminus \{1\}} \mathbb{P}(A_i) \\
&= \mathbb{P}\big(A_1^C\big)) \prod_{i \in J \setminus \{1\}} \mathbb{P}(A_i)
\end{aligned}$$

□

Insbesondere zeigt Lemma III.16, dass wir in einer Familie unabhängiger Ereignisse beliebig viele Ereignisse durch ihr Komplement, $\varnothing$ oder Ω ersetzen können, ohne die Unabhängigkeit zu verlieren.

Satz III.17
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $\mathscr{F}_i \subseteq \mathscr{F}, i \in I$, seien $\cap$-stabile Familien von Ereignissen. Dann gilt

$$\mathscr{F}_i, i \in I \text{ unabhängig} \iff \sigma(\mathscr{F}_i), i \in I \text{ unabhängig.}$$

Beweis. oBdA sei $I = \{1, \ldots, n\}$ und $\Omega \in \mathscr{F}_i, i \in I$.

- $\Leftarrow$: trivial, da $\mathscr{F}_i \subseteq \sigma(\mathscr{F}_i)$ und das Weglassen von Mengen erlaubt ist.
- $\Rightarrow$: zeigen wir rekursiv
  1. Wähle $F_i \in \mathscr{F}_i, i = 2, \ldots, n$ und defniere für $F \in \sigma(\mathscr{F}_i)$ die endlichen Maße

     $$\mu(F) = \mathbb{P}(F \cap F_2 \cap \cdots \cap F_n) \text{ und } \nu(F) = \mathbb{P}(F)\,\mathbb{P}(F_2)\,\ldots\,\mathbb{P}(F_n)$$

  2. Da die Familien $\mathscr{F}_i$ unabhängig sind, gilt $\mu\mid_{\mathscr{F}_1} = \nu\mid_{\mathscr{F}_1}$. Nach dem Eindeutigkeitssatz für Maße (Satz I.I.9) folgt $\mu\mid_{\sigma(\mathscr{F}_1)} = \nu\mid_{\sigma(\mathscr{F}_1)}$ also

     $$\mathbb{P}(F \cap F_2 \cap \cdots \cap F_n) = \mathbb{P}(F)\mathbb{P}(F_2)\ldots\mathbb{P}(F_n)$$

     für alle $F \in \sigma(\mathscr{F}_i)$ und $F_i \in \mathscr{F}_i, i = 1, \ldots, n$. Da $\Omega \in \mathscr{F}_i$ für alle i gilt die erhaltene Produktformel für alle Teilemengen $J \subseteq I$.
     Also sind

     $$\sigma(\mathscr{F}_1), \mathscr{F}_2, \ldots, \mathscr{F}_n \text{ unabhängig}$$

  3. Wiederholtes Anwenden von 1 und 2 liefert den Satz. □

Mit Satz III.17 folgen:

Folgerung III.18
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und

$$\mathscr{F}_{i,j} \subseteq \mathscr{F}, \quad 1 \leq i \leq n, 1 \leq j \leq m(i)$$

unabhängige, $\cap$-stabile Familien. Dann sind auch

$$\mathscr{G}_i = \sigma(\mathscr{F}_{i,1}, \dots, \mathscr{F}_{i,m(i)}), \quad 1 \leq i \leq n$$

unabhängig.

Folgerung III.19
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und

$$X_{ij} : \Omega \to E, \quad 1 \leq i \leq n, 1 \leq j \leq m(i)$$

unabhängige Zufallsvariablen. Zudem seien $f_i : E^{m(i)} \to \mathbb{R}$ messbar. Dann sind auch die Zufallsvariablen

$$f_i(X_{i,1}, \dots, X_{i,m(i)}), \quad 1 \leq i \leq n$$

unabhängig.

■ **Beispiel III.20**
$X_1, \dots, X_n$ unabhängige reelle Zufallsvariablen. Dann sind auch

$$Y_1 = X_1, Y_2 = X_2 + \dots + X_n$$

unabhängig.

Beweis (Folgerung III.18). OBdA sei $\Omega \in \mathscr{F}_{i,j} \forall i, j$. Dann sind die Familien:

$$\mathscr{F}_i^{\cap} := \left\{ F_{i,1} \cap \dots \cap F_{i,m(i)} \mid F_{i,j} \in \mathscr{F}_{i,j}, 1 \leq j \leq m(i) \right\}, 1 \leq i \leq n$$

$\cap$-stabil, unabhängig und es gilt: $\mathscr{F}_{i,1}, \dots, \mathscr{F}_{i,m(i)} \subseteq \mathscr{F}_i^{\cap}$ ($\nearrow$ HA)! Nach Satz III.17 sind auch $\sigma(\mathscr{F}_i^{\cap})$ unabhängig. Damit folgt die Behauptung, da $\sigma(\mathscr{F}_i^{\cap}) = \mathscr{G}_i$:

$$\begin{aligned} \mathscr{F}_{i,1}, \dots, \mathscr{F}_{i,m(i)} \subseteq \mathscr{F}_i^{\cap} \subseteq \sigma(\mathscr{F}_{i,1}, \dots, \mathscr{F}_{i,m(i)}) = \mathscr{G}_i \\ \Rightarrow \mathscr{G} = \sigma(\mathscr{F}_{i,1}, \dots, \mathscr{F}_{i,m(i)}) \subseteq \sigma(\mathscr{F}_i^{\cap}) \subseteq \mathscr{G}_i. \end{aligned}$$ □

Beweis (Folgerung III.19). Setze $\mathscr{F}_{i,j} = \sigma(X_{i,j})$ und $\mathscr{G}_i = \sigma(\mathscr{F}_{i,1}, \dots, \mathscr{F}_{i,m(i)})$, dann sind nach Folgerung III.18 die $\mathscr{G}_i, i = 1, \dots, n$ unabhängig. Zudem ist

$$Y_i := f_i(X_{i,1}, \dots, X_{i,m(i)})$$

$\mathscr{G}_i$ messbar, also $\sigma(Y_i) \subseteq \mathscr{G}_i$. Damit erben die Y_i die Unabhängigkeit der $\mathscr{G}_i$. □

Satz III.21 (Unabhängigkeit von Zufallsvariablen)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $X_1, \ldots, X_n : (\Omega, \mathscr{F}) \to (E, \mathcal{E})$ Zufallsvariablen. Dann sind die folgenden Aussagen äquivalent:

1. $X_1, \ldots, X_n$ sind unabhängig
2. $\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \quad \forall A_1, \ldots, A_n \in \mathcal{E}$.
3. Die gemeinsame Verteilung der X_i entspricht dem Produktmaß der einzelnen Verteilungen

$$\mathbb{P}_{X_1,\ldots,X_n} = \bigotimes_{i=1}^n \mathbb{P}_{X_i}$$

Beweis. Per Ringschluss:

1 ⇒ 2: Seien $A_1, \ldots, A_n \in E$ beliebig, dann gilt per Definition

$$\begin{aligned}
\mathbb{P}_{X_1,\ldots,X_n}(A_1 \times \cdots \times A_n) &= \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) \\
&= \mathbb{P}\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) \\
&\overset{\text{unabh}}{=} \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \\
&= \prod_{i=1}^n \mathbb{P}_{X_i}(A_i) = \left(\bigotimes_{i=1}^n \mathbb{P}_{X_i}\right)(A_1 \times \cdots \times A_n)
\end{aligned}$$

2 ⇒ 3: Aus der obigen Rechnung sehen wir, dass 2 bereits 3 impliziert für alle Rechtecke: $\bigtimes_{i=1}^n A_i$. Da die Familie der Rechtecke $\cap$-stabil ist und $\mathbb{E}^{\otimes n}$ erzeugt, folgt die Aussage aus dem Eindeutigkeitssatz für Maße **???**.

3 ⇒ 1: Sei $J \subseteq \{1, \ldots, n\}$ und setze

$$A_i := \begin{cases} \text{beliebig} & \text{in } \mathbb{E}, i \in J \\ E & i \notin J. \end{cases}$$

Dann

$$\begin{aligned}
\mathbb{P}(X_i \in A_i, i \in J) &= \mathbb{P}(X_i \in A_i, i = 1, \ldots, n) \\
&= \prod_{i=1}^n \mathbb{P}(X_i \in A_i) \\
&= \prod_{i \in J} \mathbb{P}(A_i \in A_i).
\end{aligned}$$

□

■ **Beispiel III.22**
Im Urnenmodell mit Zurücklegen hat der Vektor $X = (X_1, \ldots, X_n)$ mit X_i = Farbe im i-ten Zug als Zähldichte die Produktdichte der X_i. Die $X_1, \ldots, X_n$ sind also unabhängig.

2.1. Konstruktion unabhängiger Zufallsvariablen

Kapitel I: Zu beliebiger Wahrscheinlichkeitsverteilung $\mathbb{P}_X$ existiert Wahrscheinlichkeitsraum mit Zufallsvariable X auf diesem Wahrscheinlichkeitsraum, so dass $X \sim \mathbb{P}_X$.

1. Seien $\mathbb{P}_{X_1}, \dots, \mathbb{P}_{X_n}$ Wahrscheinlichkeitsverteilungen auf $(E, \mathbb{E})$. Gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$ und Zufallsvariablen X_1, X_2 unabhängig, so dass $X_1 \sim \mathbb{P}_{X_1}$?
2. Wie kann ich beliebig (unendlich) viele unabhängige Zufallsvariablen konstruieren?

Wir beginnen mit 1:
Konstruiere zwei Wahrscheinlichkeitsräume $(\Omega_i, \mathscr{F}_i, \mathbb{P}_i), i = 1, 2$ und Zufallsvariablen X_1, X_2 mit $X_i \sim \mathbb{P}_{X_i}$. Auf dem Produktraum

$$\Omega = \Omega_1 \times \Omega_2, \quad \mathscr{F} := \mathscr{F}_1 \otimes \mathscr{F}_2 \text{ und } \mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$$

definiere

$$\begin{aligned} X_1' &: \Omega_1 \times \Omega_2 \to E\colon (\omega_1, \omega_2) \mapsto X_1(\omega_1) \\ X_2' &: \Omega_1 \times \Omega_2 \to E\colon (\omega_1, \omega_2) \mapsto X_2(\omega_2) \end{aligned}$$

Dann gilt für beliebige Ereignisse: $F_1, F_2 \in \mathbb{E}$

$$\underbrace{\{X_1' \in F_1\} \cap \{X_2' \in F_2\}}_{\supseteq \Omega = \Omega_1 \times \Omega_2} = \underbrace{\{X_1 \in F_1\}}_{\supseteq \Omega_1} \times \underbrace{\{X_2 \in F_2\}}_{\supseteq \Omega_2} \in \mathscr{F}_1 \times \mathscr{F}_2$$

und damit folgt die Messbarkeit der Abbildungen X_1', X_2', d.h. X_1', X_2' sind Zufallsvariablen auf $(\Omega, \mathscr{F})$. Zudem gilt

$$\begin{aligned} \mathbb{P}(X_1' \in F_1, X_2' \in F_2) &= \mathbb{P}_1 \otimes \mathbb{P}_2(\{X_1 \in F_1\} \times \{X_2 \in F_2\}) \\ &= \mathbb{P}_1(X_1 \in F_1)\mathbb{P}_2(X_2 \in F_2), \end{aligned}$$

also

$$\mathbb{P}(X_i' \in F_i) = \mathbb{P}_i(X_i' \in F_i)$$

sowie nach Satz III.23 $X_1' \perp\!\!\!\perp X_2'$.
Wenn $(\Omega_2, \mathscr{F}_1, \mathbb{P}_1) = (\Omega_2, \mathscr{F}_2, \mathbb{P}_2)$, so liefert die obige Konstruktion zwei unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum. Andernfalls können wir auf den Produktraum ausweichen und X_i' anstelle von X_i betrachten. Die obige Konstruktion lässt sich direkt auf <u>endlich</u> viele Zufallsvariablen übertragen.
Zu 2:

Satz III.23 (Satz von Kolmogorov)
Sei I beliebige Indexmenge und $(\Omega_i, \mathscr{F}_i, \mathbb{P}_i), i \in I$ Wahrscheinlichkeitsräume. Setze

$$\Omega_I := \bigtimes_{i\in I} \Omega_i = \left\{ \omega : I \to \bigcup_{i\in I} \Omega_i, \omega_i \in \Omega_i, i \in I \right\}$$
$$\mathscr{F}_I := \sigma(\pi^{-1}(\mathscr{F}_i), i \in I)$$

wobei $\pi_i : \Omega_I \to \Omega_i$ mit $\omega \longmapsto \omega_i$ die Projektionsabbildung. Dann existiert auf $(\Omega_I, \mathscr{F}_I)$ genau ein Maß $\mathbb{P}_I$, sodass für alle $H \subseteq I$ mit $0 < |H| < \infty$ gilt

$$\pi_H(\mathbb{P}_I) = \bigotimes_{i\in H} \mathbb{P}_i,$$

wobei $\pi_H : \Omega_I \to \Omega_H$ wiederum die Projektionsabbildung.

Beweis. ↗ Schilling Maß und Integral, Satz 17.4. □

Sind auf den Wahrscheinlichkeitsräumen $(\Omega_i, \mathscr{F}_i, \mathbb{P}_i), i \in I$, nun Zufallsvariablen $X_i : \Omega_i \to E$ gegeben, so definieren wir wie im Satz von Kolmogorov (Satz III.23)

$$(\Omega, \mathscr{F}, \mathbb{P}) := \left(\Omega_I, \mathscr{F}_I, \mathbb{P}_I = \bigotimes_{i\in I} \right) \text{ mit } \omega = (\omega_i)_{i\in I}$$

und wie im endlichen Fall

$$X_i' : \Omega \to E \text{ mit } X_i'(\omega) = X_i(\omega_i).$$

Da die Unabhängigkeit der Zufallsvariablen über endliche Teilfamilien definiert ist, folgt diese wie im endlichen Fall.

2.2. Faltungen

Seien X, Y zwei reelle und unabhängige Zufallsvariablen mit

$$X \sim \mathbb{P}_X \text{ und } Y \sim \mathbb{P}_Y.$$

Dann hat (X, Y) die Verteilung $\mathbb{P}_X \otimes \mathbb{P}_Y$ auf $\mathbb{R}^2$. Andernfalls ist auch $X + Y$ eine reelle Zufallsvariable, dann

$$X + Y = A(X, Y) \text{ mit } A : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto x + y.$$

A ist stetig, also messbar. Die Verteilung von $X + Y$ ist dann $(\mathbb{P}_X \otimes \mathbb{P}_Y) \circ A^{-1}$.

Definition III.24 (Faltung)
Seien $\mathbb{P}_1, \mathbb{P}_2$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. Das durch

$$\mathbb{P}_1 \star \mathbb{P}_2(F) = \iint \mathbb{1}_F(x+y)\mathbb{P}_1(\mathrm{d}x)\mathbb{P}_2(\mathrm{d}y)$$

definierte Wahrscheinlichkeitsmaß $\mathbb{P}_1 \star \mathbb{P}_2 = (\mathbb{P}_1 \otimes \mathbb{P}_2) \circ A^{-1}$ auf $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ heißt Faltung von $\mathbb{P}_1$ und $\mathbb{P}_2$.

Satz III.25
Seien $X, Y : \Omega \to \mathbb{R}^n$ unabhängige Zufallsvariablen mit Verteilungen $\mathbb{P}_X, \mathbb{P}_Y$. Dann ist

$$\mathbb{P}_{X+Y} = \mathbb{P}_X \star \mathbb{P}_Y,$$

die Verteilung von $X + Y$.

Beweis. Siehe Herleitung Faltung. □

Faltung von Wahrscheinlichkeitsmaßen und Dichten besitzen wieder eine Dichte.

Satz III.26
Seien $\mathbb{P}_1, \mathbb{P}_2$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}^n))$

1. Diskreter Fall: Sind $\mathbb{P}_1, \mathbb{P}_2$ de facto Wahrscheinlichkeitsmaße auf $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$ mit Zähldichte ρ_1, ρ_2. Dann ist die Faltung $\mathbb{P}_1 \star \mathbb{P}_2$ Wahrscheinlichkeitsmaß auf $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$ mit Zähldichte

$$\rho_1 \star \rho_2(k) = \sum_{l \in \mathbb{Z}} \rho_1(l)\rho_2(k-l).$$

2. Stetiger Fall: Besitzt $\mathbb{P}_1, \mathbb{P}_2$ Dichtefunktionen ρ_1, ρ_2, so besitzt die Faltung $\mathbb{P}_1 \star \mathbb{P}_2$ die Dichtefunktion

$$\rho_1 \star \rho_2(x) = \int_{\mathbb{R}} \rho_1(y)\rho_2(x-y)\,\mathrm{d}y \quad x \in \mathbb{R}$$

Beweis. 1. Diskrete Fall: Sei $k \in \mathbb{Z}$

$$\begin{aligned}(\mathbb{P}_1 \otimes \mathbb{P}_2)(A = k) &= \sum_{\substack{l_1, l_2 \in \mathbb{Z} \\ l_1 + l_2 = k}} \rho_1(l_1)\rho_2(l_2) \\ &= \rho_1 \star \rho_2(k)\end{aligned}$$

2. Stetiger Fall: Sei $c \in \mathbb{R}$

$$\begin{aligned}
\mathbb{P}_1 + \mathbb{P}_2((-\infty, c]) &= (\mathbb{P}_1 \otimes \mathbb{P}_2)(A \leq c) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{(\infty,c]}(x+y)\rho_1(x)\rho_2(y)\,\mathrm{d}x\,\mathrm{d}y \\
&\overset{y=z-x}{=} \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{(\infty,c]}(z)\rho_1(x)\rho_2(z-x)\,\mathrm{d}x\,\mathrm{d}z \\
&= \int_{-\infty}^{c} \underbrace{\int_{\mathbb{R}} \rho_1(x)\rho_2(z-x)\,\mathrm{d}x}_{\rho_1 \star \rho_2(z)}\ \mathrm{d}z.
\end{aligned}$$

□

■ **Beispiel III.27**

Seien $X \sim \text{Poisson}(\lambda), Y \sim \text{Poisson}(\mu)$ zwei unabhängigen reellen Zufallsvariablen (mit Werten in $\mathbb{N}_0$). Dann ist $X+Y$ eine Zufallsvariable mit Werten in $\mathbb{N}_0$ und Zähldichte

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{l\in\mathbb{Z}} \mathbb{P}(X=l)\mathbb{P}(Y=k-l) \\
&= \sum_{l\in\mathbb{Z}} \frac{\lambda^l}{l!}e^{-\lambda}\frac{\mu^{k-l}}{(k-l)!}e^{-\mu} \\
&= e^{-(\lambda+\mu)}\frac{1}{k!}\sum_{l=0}^{k}\binom{k}{l}\lambda^l\mu^{k-l} \\
&= e^{-(\lambda+\mu)}\frac{1}{k!}(\lambda+\mu)^k \quad \forall k \in \mathbb{N}_0,
\end{aligned}$$

so dass

$$X+Y \sim \text{Poisson}(\lambda+\mu).$$

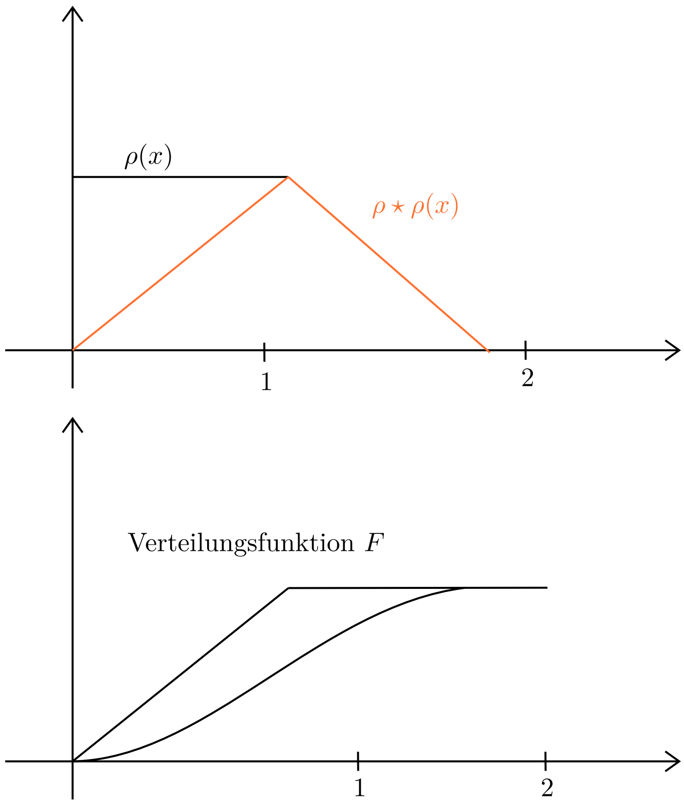
D.h. der Typ der Verteilung ist bei der Faltung erhalten geblieben.

Hinweis: Das ist aber nicht immer der Fall!

■ **Beispiel III.28**

Seien $X, Y \sim \text{U}([0,1])$ zwei unabhängige Zufallsvariablen mit Dichten $\rho(x) = \mathbb{1}_{[0,1]}(x)$. Dann ist $X+Y$ eine Zufallsvariable mit Werten in $[0,2]$ und Dichte

$$\begin{aligned}
\rho \star \rho(x) &= \int_{\mathbb{R}} \rho(y)\rho(x-y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}} \mathbb{1}_{[0,1]}(y)\mathbb{1}_{[0,1]}(x-y)\,\mathrm{d}y \\
&= \int_{0\vee(x-1)}^{1\wedge x} \mathrm{d}y = \begin{cases} x & 0 \leq x \leq 1 \\ 2-x & 1 \leq x \leq 2 \\ 0 & \text{sonst.} \end{cases}
\end{aligned}$$


$\rho(x)$
$\rho \star \rho(x)$
1
2
Verteilungsfunktion F
1
2

Kapitel IV

Weitere Standardmodelle der Wahrscheinlichkeitstheorie

1. Stetige Gleichverteilung

▶ **Erinnerung**

$\Omega \subset \mathbb{R}^n$ Borel-messbar mit LEBESGUE-Volumen $0 < \lambda(\Omega) < \infty$. Wahrscheinlichkeitsmaß ist $(\Omega, \mathscr{B}(\Omega))$ mit Dichte

$$\rho(x) = \frac{1}{\lambda(\Omega)}$$

heißt stetige Gleichverteilung auf Ω: $\mathrm{U}(\Omega)$.
Für alle $A \in \mathscr{B}(\Omega)$ gilt:

$$\mathbb{P}(A) = \int_A \rho(x)\,\mathrm{d}x = \frac{\lambda(A)}{\lambda(\Omega)}.$$

Meist verwenden wir $\mathrm{U}([a,b]), a < b$ (Gleichverteilung auf Intervall) mit $\rho(x) = {}^1\!/\!{}_{(b-a)}, a \leq x \leq b$ und Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x < a \\ \int_a^x \frac{1}{b-a}\,\mathrm{d}x = \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b. \end{cases}$$

2. Wartezeitverteilungen

Negative Binomialverteilung:
Wir wiederholen ein Bernoulliexperiment mit Erfolgswahrscheinlichkeit $p \in [0,1]$ unendlich oft. Gesucht ist die Anzahl der Misserfolge bis zum r-ten Erfolg, $r \in \mathbb{N}$. Ein passender Ergebnisraum ist $\Omega = \mathbb{N}_0$. Für Modellierung ist es jedoch leichter in jedem Versuch erfolgt ("1") oder Misserfolg ("0") festzuhalten und i mit dem unendlichen Produktmaß des Bernoullimaßes auf $\{0,1\}^{\mathbb{N}}$ zu arbeiten.
Als Zufallsvariable

$$X_r : \{0,1\}^{\mathbb{N}} \to \Omega$$

welche die Anzahl der Misserfolge bis zum r-ten Erfolg darstellt, setze

$$X_r(\omega) = \min\left\{\sum_{i=1}^{k} \omega_i = r\right\} = r.$$

Dann

$$\mathbb{P}(X_r = k) = \sum_{\substack{\omega \in \{0,1\}^{\mathbb{N}} \\ X_r(\omega)=k}} \prod_{i=1}^{\infty} \rho(\omega_i)$$

mit $\rho(0) = 1 - p, \rho(1) = 1$ (Zähldichte der Bernoulliverteilung), also

$$\mathbb{P}(X_r = k) = \binom{r+k-1}{k}(1-p)^k p^r \quad r \in \mathbb{N}_0.$$

Definition IV.1 (negative Binomialverteilung, geometrische Verteilung)
Sei $p \in [0,1]$ und $r \in \mathbb{N}$, dann heißt die Verteilung auf $\mathbb{N}_0$ mit Zähldichte

$$\rho(k) = \binom{r+k-1}{k} p^r (1-p)^k$$

die negative Binomialverteilung mit Parametern (r, p). Schreibe $\mathrm{negBin}(r, p)$. Im Fall $r = 1$ nennen wir die Verteilung mit Zähldichte

$$\rho(k) = p(1-p)^k \quad k \in \mathbb{N}_0$$

geometrische Verteilung mit Parametern p. Schreibe $\mathrm{Geom}(p)$.

2.1. Exponential- und Gammaverteilung

1. Ziel: Modelliere die Wartezeit auf r Ereignisse in kontinuierlicher Zeit.
2. Wähle: $(\Omega, \mathscr{F}) = (\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$
3. Annahmen:
   - Jedes Ereignis geschieht zu einer zufälligen Zeit
   - Die Anzahl der Ereignisse bis zur Zeit t sei $\mathrm{Poisson}(\lambda t)$ verteilt.

   Die zweite Annahme macht Sinn, denn

   - Poissonverteilung ist Modell für Anzahl seltener Ereignisse
   - Nach Beispiel 3.26:

     $$\mathrm{Poisson}(\lambda t) \star \mathrm{Poisson}(\lambda s) = \mathrm{Poisson}(\lambda(t+s))$$

     Die Linearität des Parameters entspricht also einer Stationaritätsvorrausetzung:

Modelliert

$$\begin{aligned} X &\sim \text{Poisson}(\lambda t) \text{ die Ereignisse in } (0,t], \\ Y &\sim \text{Poisson}(\lambda s) \text{ die Ereignisse in } (t,t+s] \end{aligned}$$

so modelliert

$$X+Y \sim \text{Poisson}(\lambda(t+s)) \text{ die Ereignisse in } (0,t+s].$$

Unter diesen Annahmen folgt für die Wahrscheinlichkeit in $(0,t]$ mindestens r Ereignisse zu beobachten

$$\mathbb{P}((0,t)) = 1 - \sum_{k=0}^{r-1} \underbrace{e^{-\lambda t}\frac{(\lambda t)^k}{k!}}_{\text{Zähldichte Poisson}(\lambda t) \text{ in } t}$$

Wegen:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(1-\sum_{k=0}^{r-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!}\right) &= -(-\lambda)e^{-\lambda t}\sum_{k=0}^{r-1}\frac{(\lambda t)^k}{k!} - e^{-\lambda t}\sum_{k=0}^{r-1}\frac{\lambda^k t^{k-1}}{(k-1)!} \\ &= e^{-\lambda t}\left(\sum_{k=0}^{r-1}\frac{\lambda^k t^{k-1}}{k!} - \sum_{l=0}^{r-1}\frac{\lambda^k t^{l-1}}{(l-1)!}\right) \end{aligned}$$

gilt

$$\mathbb{P}((0,t)) = \int_0^t e^{-\lambda t}\frac{\lambda^r x^{r-1}}{(r-1)!}\,\mathrm{d}x.$$

Wir definieren allgemeiner:

Definition IV.2 (Gammaverteilung, Gammafunktion)
Seien $\lambda > 0, r > 0$, dann heißt die Verteilung auf $(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ mit Dichte

$$f(x) = \frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x} \text{ mit } \Gamma(r) = \int_0^\infty y^{r-1}e^{-y}\,\mathrm{d}y, r > 0$$

Gammafunktion, Gammaverteilung mit Parametern λ, r. Schreibe $\text{Gamma}(\lambda, r)$.

Insbesondere ist $\text{Gamma}(\lambda, 1)$ gerade die Exponentialverteilung (vgl. Bsp 17?).
Die Gammaverteilung ist reproduktiv: Die Wartezeit auf $r+s$ Ereignisse entspricht der Wartezeit auf r Ereignisse + s (weitere) Ereignisse:

Lemma IV.3
Seien $X \sim \text{Gamma}(\lambda, r), Y \sim \text{Gamma}(\lambda, s)$ unabhängig, dann impliziert das

$$X+Y \sim \text{Gamma}(\lambda, r+s)$$

Beweis. Hier nur für $r, s \in \mathbb{N}$, allgemein später mit momenterzeugende Funktionen.

Seien $\rho(x), \rho(y)$ Dichten von X, Y. Nach Satz III.25 folgt

$$\begin{aligned}
\rho_{(X+Y)} &= \rho_X \star \rho_Y(x) = \int_{\mathbb{R}} \rho_X(y)\rho_Y(x-y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}} \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y} \frac{\lambda^s}{\Gamma(s)} (x-y)^{s-1} e^{-\lambda(x-y)}\,\mathrm{d}y \\
&= \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda x} \int_0^x y^{r-1}(x-y)^{s-1}\,\mathrm{d}y \\
&\overset{\text{P.I.}}{=} \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(s)} e^{-\lambda x} \left(\underbrace{\left[\frac{1}{r} y^r (x-y)^{s-1}\right]_{y=0}^{x}}_{=0} + \frac{s-1}{r} \int_0^x y^r (x-y)^{s-2}\,\mathrm{d}y \right) \\
&\overset{\text{Ind}}{=} \frac{\lambda^{r+s}}{\Gamma(r)\Gamma(r)} e^{-\lambda x} \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \frac{\lambda^{r+s}}{\Gamma(r+s)} e^{-\lambda x}.
\end{aligned}$$

□

Exponentialverteilungen sind zudem gedächtnislos:

Lemma IV.4 (Gedächtnislosigkeit der Exponentialverteilung)
Sei $X \sim \operatorname{Exp}(\lambda)$, dann gilt

$$\mathbb{P}(X > t) = \mathbb{P}(X > t + s \mid X > s) \quad t, s \geq 0. \tag{$*$}$$

Beweis.

$$\begin{aligned}
\mathbb{P}(X > t + s \mid X \geq s) &= \frac{\mathbb{P}(X > t + s, X \geq s)}{\mathbb{P}(X > s)} \\
&= \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda t} = \mathbb{P}(X > t).
\end{aligned}$$

□

■ **Beispiel IV.5**
Eine Studentin wartet morgens eine $\operatorname{Exp}(1/5)$ verteilte Zeit X auf den Bus zur Uni. Die Wahrscheinlichkeit einer Wartezeit ≥ 5 Minuten

$$\mathbb{P}(X \geq 5) = e^{-\frac{1}{5}\cdot 5} = e^{-1} \approx 0,37.$$

An einen kalten, stürmischen Frühlingstag hat die Studentin bereits 10 Minuten gewartet. Die Wahrscheinlichkeit mindestens 5 weitere Minuten zu warten ist

$$\mathbb{P}(X \geq 15 \mid X > 10) = \mathbb{P}(X \geq 5) = e^{-1} \approx 0,37.$$

Hinweis: Man kann sogar zeigen, dass die Exponentialverteilung die einzige absolutstetige Verteilung mit $(*)$ ist.

Kapitel V

Erwartungswerte & Varianz

1. Frage:
   Beispiel IV.5 Durchschnittliche Wartezeit? $\rightsquigarrow$ Erwartungswert
   Wie stark ist die Streuung um den Durchschnitt? $\rightsquigarrow$ Varianz

1. Der Erwartungswert

Definition V.1 (Erwartungswert)
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariable. Dann ist

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(\,\mathrm{d}\omega) = \int_\mathbb{R} x\mathbb{P}(X \in \,\mathrm{d}x)$$

der Erwartungswert von X.

Hinweis: Der Erwartungswert von X existiert, genau dann wenn

$$\int_\Omega |X(\omega)|\mathbb{P}(\,\mathrm{d}\omega) < \infty \text{ bzw. } \mathbb{E}[|X|] < \infty$$

d.h. genau dann wenn $X \in \mathscr{L}^1(\mathbb{P})$.
Für nichtnegative Zufallsvariablen ist der Erwartungswert immer definiert, wenn wir $+\infty$ als zulässigen Wert annehmen, was wir in der Folge auch tun.

■ **Beispiel V.2**
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $A \in \mathscr{F}$ und sei $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ die Indikatorvariable

$$X(\omega) = \mathbb{1}_A(\omega)$$

Dann gilt: $X \in \mathscr{L}^1(\mathbb{P})$ und

$$\mathbb{E}[X] = \int_\Omega \mathbb{1}_A(\omega)\mathbb{P}(\,\mathrm{d}\omega) = \int_A \mathbb{P}(\,\mathrm{d}\omega) = \mathbb{P}(A).$$

Satz V.3
Sei $X : (\Omega, \mathscr{F}) \to (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n)$ Zufallsvariable und $f : \mathbb{R}^n \to \mathbb{R}$ Borel-messbar. Dann

$$\mathbb{E}[f(X)] = \int f(X)\,\mathrm{d}\mathbb{P} = \int_\Omega \mathbb{E}(X(\omega))\,\mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}^n} f(X)\mathbb{P}(X \in \,\mathrm{d}x).$$

Beweis. Sei $f(X)$ eine reelle Zufallsvariable. Die Formel folgt direkt auf dem Transformationssatz für Bildmaße (↗ Schilling MINT 18.1). □

Satz V.4 (Erwartungswerte bei Existenz einer (Zähl-)dichte)
Sei $X : (\Omega, \mathscr{F}) \to (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ Zufallsvariable und

$$f : \mathbb{R}^n \to \mathbb{R} \text{ Borel-messbar.}$$

1. diskreter Fall: Ist $\mathbb{P}_X$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$ und der Zähldichte ρ, so
$$\mathbb{E}[f(X)] = \sum_{x \in \mathbb{Z}} f(x)\rho(x).$$
2. stetiger Fall: Besitzt $\mathbb{P}_X$ eine Dichte ρ (bzgl Lebesguemaß), so
$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f(x)\rho(x)\,\mathrm{d}x$$

Beweis. Klar aus Definition V.1 und Satz V.3. □

■ **Beispiel V.5**
Sei $X \sim \mathrm{Bin}(n,p)$. Dann gilt

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \underbrace{\binom{n-1}{k-1} p^{k-1} (1-p)^{n-1-(k-1)}}_{\text{Zähldichte } \mathrm{Bin}(n-1,p) \text{ in } k-1} \\
&= np.
\end{aligned}$$

Da der Erwartungswert ein Integral ist, übertragen sich viele Eigenschaften.

Satz V.6 (Eigenschaften des Erwartungswertes)
Seien $X, Y, X_n : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R})), n \in \mathbb{N}$ Zufallsvariablen in $\mathscr{L}^1(\mathbb{P})$ und $a, b \in \mathbb{R}$ konstant.

1. Linearität: $\mathbb{E}[aX + bY] = a\mathbb{E}X + b\mathbb{E}Y$
2. Monotonie: $X \le Y$, d.h. $X(\omega) \le Y(\omega), \forall \omega \in \Omega$. Dann gilt
$$\mathbb{E}[X] \le \mathbb{E}[Y] \text{ und insbesondere gilt } X \ge 0 \implies \mathbb{E}X \ge 0.$$
3. Lemma von Fatou:
$$\mathbb{E}[\liminf_{n\to\infty} X_n] \le \liminf_{n\to\infty} \mathbb{E}[X_n]$$
4. Satz von Beppo-Levi: Wenn $X_n \ge 0$ und $X_n \uparrow X$ so gilt:
$$\mathbb{E}[X] = \sup_{n\in\mathbb{N}} \mathbb{E}[X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$$
5. Dominierte Konvergenz/ Satz von Lebesgue: Sei $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ und
$$\mathbb{P}(\{\omega : |X_n(\omega)| \le Y(\omega)\}) = 1 \qquad (|X| \le Y \mathbb{P} \text{ fast sicher})$$
für $Y \in \mathscr{L}^1(\mathbb{P}) \implies X \in \mathscr{L}^1(\mathbb{P})$ und $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$
6. Markov-Ungleichung: Sei $\varepsilon > 0$, dann gilt
$$\mathbb{P}(|X| \ge \varepsilon) \le \frac{1}{\varepsilon}\mathbb{E}[|X|]$$
7. Hölder-Ungleichung: Sei $1 \le p, q \le \infty$, $\frac{1}{p} + \frac{1}{q} = 1$
$$\mathbb{E}[|XY|] \le (\mathbb{E}[|X|^p])^{\frac{1}{p}} (\mathbb{E}[|Y|^q])^{\frac{1}{q}}$$
($p = q = 2$ Cauchy-Schwarz-Ungleichung)
8. Jensen'sche Ungleichung: Sei $X \ge 0$ und $\varphi : [0, \infty) \to [0, \infty)$ konvex, messbar
$\varphi(\mathbb{E}[X]) \le \mathbb{E}[\varphi(X)]$.

Beweis. ↗ Schilling MINT. □

■ Beispiel V.7

Da für $X_1, \dots, X_n \sim \text{Bernoulli}(p)$ unabhängig gilt, dass $\underbrace{X_1 + \cdots + X_n}_{=X} \sim \text{Bin}(n,p)$ folgt

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_1 + \cdots + X_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = n\mathbb{E}[X_1] \\ &= n(1 \cdot \underbrace{\mathbb{P}(X_1 = 1)}_{=p} + 0 \cdot \mathbb{P}(X_1 = 0)) \\ &= n \cdot p.\end{aligned}$$

Satz V.8 (Produktformel für Erwartungswerte)

Seien $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $X_1, \dots, X_n : \Omega \to \mathbb{R}^d$ unabhängige Zufallsvariablen und $f_1, \dots, f_n : \mathbb{R}^d \to \mathbb{R}$ messbar. Wenn $f_i(X_i) \geq 0, i = 1, \dots, n$ sei mit $f_i(X_i) \in \mathscr{L}^1(\mathbb{P}), i = 1, \dots, n$, dann gilt

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]$$

Für den Beweis von Satz V.8 benötigen wir:

Lemma V.9

$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X, Y : \Omega \to \mathbb{R}^d$ unabhängige Zufallsvariablen und $h : \mathbb{R}^{2d} \to \mathbb{R}$ messbar. Falls $h \geq 0$ oder $h(X,Y) \in \mathscr{L}^1(\mathbb{P})$, dann

$$\begin{aligned}\mathbb{E}[h(X,Y)] &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x,y) \mathbb{P}(X \in \mathrm{d}x) \mathbb{P}(Y \in \mathrm{d}y) \\ &= \mathbb{E}\left[\int_{\mathbb{R}^d} h(X,y) \mathbb{P}(Y \in \mathrm{d}y)\right] \\ &= \mathbb{E}\left[\int_{\mathbb{R}^d} h(x,Y) \mathbb{P}(X \in \mathrm{d}x)\right]\end{aligned}$$

Beweis. Sei $h(X,Y)$ eine reelle Zufallsvariable und

$$\begin{aligned}\mathbb{E}[h(X,Y)] &= \int_\Omega h(X(\omega), Y(\omega)) \,\mathrm{d}\mathbb{P}(\omega) \\ &\overset{???}{=} \int_{\mathbb{R}^d} h(x,y) \mathbb{P}(X \in \mathrm{d}x, Y \in \mathrm{d}y) \\ &\overset{X \perp\!\!\!\perp Y, ???}{=} \int_{\mathbb{R}^{2d}} h(x,y) \mathbb{P}_X \otimes \mathbb{P}_Y(\,\mathrm{d}x, \,\mathrm{d}y) \\ &\overset{\text{Fubini}}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x,y) \mathbb{P}_X(\,\mathrm{d}x) \mathbb{P}_Y(\,\mathrm{d}y) \\ &\overset{???}{=} \int_\Omega \int_{\mathbb{R}^d} h(x, Y(\omega)) \mathbb{P}_X(\,\mathrm{d}x) \mathbb{P}(\omega) \\ &= \mathbb{E}[\int_{\mathbb{R}^d} h(x,Y) \mathbb{P}_X(\,\mathrm{d}x)].\end{aligned}$$

□

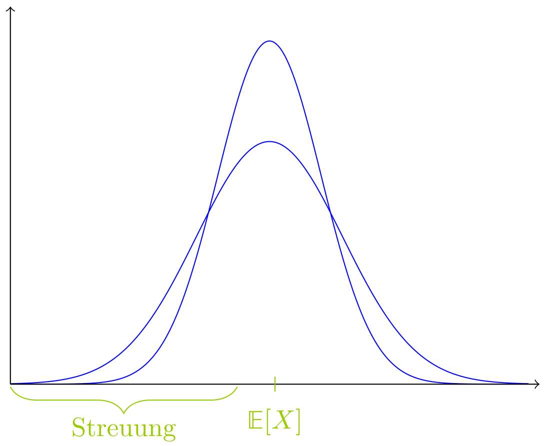
Beweis (Satz V.8). Betrachte $n = 2$, Zufallsvariablen X, Y und Abbildungen f, g. Setze $h(x,y) = f(x)g(y)$, dann folgt für $f, g \geq 0$ mit Lemma V.9

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)g(y)\mathbb{P}(X\in \mathrm{d}x)\mathbb{P}(Y\in \mathrm{d}y)\\ &= \int_{\mathbb{R}^d} f(x)\mathbb{P}(X\in \mathrm{d}x)\int_{\mathbb{R}^d} g(y)\mathbb{P}(Y\in \mathrm{d}y)\\ &= \mathbb{E}[f(X)]\cdot\mathbb{E}[g(Y)].\end{aligned}$$

Für $f(X), g(Y) \in \mathscr{L}^1(\mathbb{P})$ zeigt obige Rechnung

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[|f(X)|]\mathbb{E}[|g(Y)|] < \infty$$

also $f(X)g(Y) \in \mathscr{L}^1(\mathbb{P})$. Die Aussage folgt über die obige Rechnung. Für allgemeines n folgt Satz V.8 durch Iteration mit Folgerung III.19. $\square$



2. Varianz und höhere Momente

Definition V.10 (k-te Momente)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ reelle Zufallsvariable. Dann ist für $k \in \mathbb{N}$

$$\mathbb{E}[X^k] = \int_\Omega X^k(\omega)\mathbb{P}(\omega) = \int_\mathbb{R} x^k \mathbb{P}(X\in \mathrm{d}x)$$

das <u>k-te Moment</u> von X (sofern definiert).

▶ **Bemerkung**

- Erwartungswert $\cong$ erstes Moment
- Das k-te Moment existiert, genau dann wenn

$$\int_\Omega \left|X(\omega)^k\right|\mathbb{P}(\mathrm{d}\omega) < \infty \text{ bzw. } X \in \mathscr{L}^k(\mathbb{P})$$

- MINT: $\mathscr{L}^r(\mathbb{P}) \subseteq \mathscr{L}^s(\mathbb{P})$ für $s \leq r$

Von Interesse ist insbesondere das zweite Moment.

Definition V.11 (Varianz, Standardabweichung)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X, Y \in \mathscr{L}^2(\mathbb{P})$ reelle Zufallsvariablen.

1. Die Varianz von X ist
$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$
2. Die Standardabweichung/Streuung von X ist $\sqrt{\mathbb{V}\mathrm{ar} X}$.
3. Die Kovarianz von X und Y ist
$$\begin{aligned}\mathbb{C}\mathrm{ov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\end{aligned}$$
Hinweis: Wenn die Varianz 0 ist, heißt es nicht dass die Zufallsvariablen unabhängig waren
4. Sind $\mathbb{V}\mathrm{ar} X, \mathbb{V}\mathrm{ar} Y \geq 0$, dann ist die Korrelation von X und Y
$$\mathbb{C}\mathrm{orr}(X, Y) = \frac{\mathbb{C}\mathrm{ov}(X, Y)}{\sqrt{\mathbb{V}\mathrm{ar}(X)\mathbb{V}\mathrm{ar}(Y)}}.$$
5. Gilt $\mathbb{C}\mathrm{orr}(X, Y) = 0$, so heißen X, Y unkorreliert.

▶ **Bemerkung**

- Die Endlichkeit der Ausdrücke in Definition V.11 folgt aus der CAUCHY-SCHWARZ-Ungleichung
$$\mathbb{E}[|XY|] \leq (\mathbb{E}|X|^2)^{\frac{1}{2}} \cdot (\mathbb{E}|Y|^2)^{\frac{1}{2}}$$
- Für die (Ko)varianz gilt
$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] &= \mathbb{E}[XY - X\mathbb{E}[Y] - Y\mathbb{E}[X] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

■ **Beispiel V.12**
Sei $X \sim \mathrm{Bin}(n, p)$, dann gilt
$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}X^2 - \underbrace{(\mathbb{E}X)^2}_{=np}$$

mit

$$\begin{aligned}
\mathbb{E}X^2 &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} k \frac{(n-1)!}{(n-1-(k-1))!(k-1)!} p^{k-1}(1-p)^{n-1-(k-1)} \\
&= np \sum_{l=0}^{n-1} (l+1) \binom{n-1}{l} p^l (1-p)^{n-1-l} \qquad \text{mit } l = k-1 \\
&= np(1 + \sum_{l=0}^{n-1} l \binom{n-1}{l} p^l (1-p)^{n-1-l}) \\
&= np(1 + (n-1)p \cdot 1) \quad \text{mit Binomial Satz} \\
&= np + n(n-1)p^2 \\
&= np + n^2p^2 - np^2 - n^2p^2 \\
&= np + n(n-1)p^2.
\end{aligned}$$

Damit ist die $\mathbb{V}\mathrm{ar}(X) = \cancel{np} + n(n-1)p^2 \cancel{-np} = np(1-p)$.

Satz V.13 (Eigenschaften der (Ko-)varianz)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X, Y, X_1, \dots, X_n \in \mathscr{L}^2(\mathbb{P}), a, b \in \mathbb{R}$.

1. $\mathbb{V}\mathrm{ar}(aX + b) = a^2 \mathbb{V}\mathrm{ar}(X)$
2. Sei $(\mathbb{C}\mathrm{ov}(X,Y))^2 \le \mathbb{V}\mathrm{ar}X \mathbb{V}\mathrm{ar}Y$ und insbesondere $|\mathbb{C}\mathrm{orr}(X,Y)| \le 1$
3. $\mathbb{V}\mathrm{ar}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i) + \sum_{i,j=1//i\neq j}^n \mathbb{C}\mathrm{ov}(X_i, X_j)$ Sind die $X_1, \dots, X_n$ paarweise unkorreliert, so gilt die Formel von Bienaymé:
$$\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)$$
4. $X \perp\!\!\!\perp Y \implies \mathbb{C}\mathrm{orr}(X,Y) = 0$
5. Tschebyscheff-Ungleichung: Für $\varepsilon > 0$
$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) \le \frac{\mathbb{V}\mathrm{ar}X}{\varepsilon^2}$$

Beweis. 1. Da $\mathbb{E}[aX + b] = a\mathbb{E}X + b$, folgt

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(aX + b) &= \mathbb{E}[(a\mathbb{E}X + b - (a\mathbb{E}X + b))^2] \\
&= \mathbb{E}[a^2(X - \mathbb{E}X)^2] = a^2 \mathbb{V}\mathrm{ar}X.
\end{aligned}$$

2. Wegen 1. können wir oBdA annehmen, dass $\mathbb{E}X = 0 = \mathbb{E}Y$. Dann wird 2. zu

$$\mathbb{E}[XY]^2 \le \mathbb{E}X^2 \cdot \mathbb{E}Y^2$$

und dies ist die CAUCHY-SCHWARZ Ungleichung.

3. Wähle wieder oBdA $\mathbb{E}X_i = 0$. Dann

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left[(\sum_{i=1}^n X_i)^2\right] \\ &= \mathbb{E}\left[\sum_{i,j=1}^n X_i X_j\right] \\ &= \sum_{i,j=1}^n \mathbb{E}[X_i X_j] \\ &= \sum_{i,j=1}^n \mathbb{C}\mathrm{ov}(X_i, X_j) \\ &= \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i) + \sum_{\substack{i,j=1 \\ i\neq j}}^n \mathbb{C}\mathrm{ov}(X_i, X_j)\end{aligned}$$

4. Sei

$$\begin{aligned}X \perp\!\!\!\perp Y &\implies \mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y \\ &\implies \mathbb{C}\mathrm{ov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y = 0 \\ &\implies \mathbb{C}\mathrm{orr}(X,Y) = 0\end{aligned}$$

5. Wende die MARKOV-Ungleichung auf $X' = (X - \mathbb{E}X)^2$ an, dann folgt

$$\begin{aligned}\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) &\leq \mathbb{P}((X - \mathbb{E}X)^2 > \varepsilon^2) \\ &\overset{\text{Markov}}{\leq} \frac{1}{\varepsilon^2}\mathbb{E}[(X - \mathbb{E}X)^2] \\ &= \frac{\mathbb{V}\mathrm{ar}X}{\varepsilon^2}.\end{aligned}$$

□

3. Wahrscheinlichkeitserzeugende Funktionen

Definition V.14 (Wahrscheinlichkeitserzeugende Funktion)

1. Ist $\mathbb{P}$ Wahrscheinlichkeitsmaß auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ mit Zähldichte ρ, so heißt

$$\psi_{\mathbb{P}} := \sum_{k\in\mathbb{N}_0} s^k \rho(k) \quad 0 \leq s \leq 1,$$

Wahrscheinlichkeitserzeugende Funktion von $\mathbb{P}$. (probability generating function - pgf)

2. Ist X $\mathbb{N}_0$-wertige Zufallsvariable auf $(\Omega, \mathscr{F}, \mathbb{P})$, so heißt

$$\psi_X = \sum_{k\in\mathbb{N}_0} s^k \mathbb{P}(X = k) \quad 0 \leq s \leq 1,$$

Wahrscheinlichkeitserzeugende Funktion (pgf) von X.

▶ **Bemerkung**

- Da $\sum_{k\in\mathbb{N}_0} \rho(k) = 1$ ist die pgf auf $0 \leq s \leq 1$ wohldefiniert. Zudem ist ψ auf $[0,1)$ unendlich oft differenzierbar.
- Da $\rho(k) \geq 0 \forall k$ ist die pgf stets konvex.
- Durch TAYLORentwicklung von ψ um 0 gilt:

$$\psi_{\mathbb{P}}(s) = \sum_{k\in\mathbb{N}_0} \frac{s^k \psi_{\mathbb{P}}^{(k)}(0)}{k!}.$$

  so dass für alle $k \in \mathbb{N}_0$ folgt

$$\rho(k) = \frac{\psi_{\mathbb{P}}^{(k)}(0)}{k!}$$

  Die Verteilung $\mathbb{P}$ (bzw. $\mathbb{P} \circ X^{-1}$) ist durch $\psi_{\mathbb{P}}$ (bzw. ψ_X) eindeutig bestimmt. Also "erzeugt ψ die Zähldichte".

■ **Beispiel V.15**
Ist $X \sim \text{Bin}(n,p)$, so folgt

$$\begin{aligned}\psi_X(s) &= \sum_{k=0}^{n} s^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (ps)^k (1-p)^{n-k} \\ &= (ps-(1-p))^n = (1+p(s-1))^n. \quad \text{Binomial Satz}\end{aligned}$$

Satz V.16 (Momentenberechnung mit der pgf)
Sei X eine $\mathbb{N}_0$-wertige Zufallsvariable, dann gilt

$$\mathbb{E}[X^n] < \infty \quad n \geq 1 \iff \psi_X^{(n)} = \lim_{s\uparrow 1} \psi_X^{(n)}(s) < \infty.$$

Insbesondere gilt dann

$$\psi_X^{(n)}(1) = \mathbb{E}[X(X-1)\dots(X-n+1)].$$

Beweis. Sei $\rho(k) = \mathbb{P}(X=k)$. Durch n-faches gliedweises Differenzieren der Potenzreihe ψ_X folgt

$$\psi_X^{(n)}(s) = \sum_{k\in\mathbb{N}_0} k(k-1)\cdots(k-n+1)\rho(k)s^{k-n} \quad 0 \leq s < 1.$$

Dann existieren in $[0, \infty)$

$$\begin{aligned}
\lim_{s\uparrow 1} \psi_X &= \lim_{s\uparrow 1} \sum_{k=0}^{\infty} k(k-1)\cdots(k-n+1)\rho(k)s^{k-n} \\
&= \sum_{k=n}^{infty} \rho(k)k(k-1)\cdots(k-n+1) \\
&= \mathbb{E}[X(X-1)\cdots(X-n+1)]
\end{aligned}$$

sowie induktiv

$$\begin{aligned}
\psi^{(n)}(1) &= \lim_{s\uparrow 1} \frac{\psi^{(n-1)} - \psi^{(n-1}(s)}{1-s} \\
&= \lim_{s\uparrow 1} \sum_{k\in\mathbb{N}_0} \rho(k)k(k-1)\cdots(k-(n-1)+1)\frac{1-s^{k-(n-1)}}{1-s} \\
&= \lim_{s\uparrow 1} \sum_{k\in\mathbb{N}_0} \rho(k)k(k-1)\cdots(k-n+2) \sum_{l=0}^{k-(n-1)} s^l \quad \text{Geo. Reihe} \\
&= \sum_{k\in\mathbb{N}_0} \rho(k)k(k-1)\cdots(k-n+2)(k-n+1) \quad \text{Monotonie} \\
&= \mathbb{E}[X(X-1)\cdots(X-n+1)]
\end{aligned}$$

Insbesondere gilt $\mathbb{E}X^n < \infty$ genau dann, wenn $\psi_X^{(n)}(1) < \infty$ bzw. $\lim_{s\uparrow 1} \psi_X^{(n)}(s) < \infty$ □

■ **Beispiel V.17**

Sei $X \sim \text{Bin}(n, p)$, dann gilt Beispiel V.15

$$\psi_X(s) = (1 + p(s-1))^n.$$

Damit

$$\begin{aligned}
\psi'_X(s) &= n(1 + p(s-1))^{n-1}p \\
\psi''_X(s) &= n(n-1)(1 + p(s-1))^{n-2} \cdot p^2
\end{aligned}$$

so dass

$$\mathbb{E}[X] = \psi'_X(1) = np$$

und

$$\begin{aligned}
\mathbb{V}\text{ar}X &= \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \mathbb{E}[X(X-1)] + \mathbb{E}X - (\mathbb{E}X)^2 \\
&= \psi''_X(1) + \psi'_X(1) - (\psi'_X(1))^2 \\
&= n(n-1)p^2 + np - (np)^2 = np - np^2 = np(1-p).
\end{aligned}$$

Satz V.18
Seien X, Y unabhängige Zufallsvariablen, $\mathbb{N}_0$-wertig auf $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum. Dann gilt

$$\begin{aligned}\psi_{X+Y} &= \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X s^Y] \\ &= \mathbb{E}[s^X]\mathbb{E}[s^Y] \\ &= \psi_X(s)\psi_Y(s).\end{aligned}$$

Satz V.19
Sind $\mathbb{P}_1, \mathbb{P}_2$ Wahrscheinlichkeitsmaße auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$, so gilt

$$\psi_{\mathbb{P}_1 \star \mathbb{P}_2} = \psi_{\mathbb{P}_1}(s)\psi_{\mathbb{P}_2}(s) \quad 0 \leq s \leq 1.$$

Kapitel VI

Bedingte Verteilungen und bedingte Erwartungswerte

In Satz III.3: $(\Omega, \mathscr{F}, \mathbb{P}), A, B \in \mathscr{F}$

$$\mathbb{P}(A \mid B) = \begin{cases} \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} & \mathbb{P}(B) > 0 \\ 0/ \text{ beliebig} & \mathbb{P}(B) = 0 \end{cases}$$

In Fall $\mathbb{P}(B) > 0$ ist $\mathbb{P}(\cdot \mid B)$ ein Wahrscheinlichkeitsmaß und wir können das Integral

$$\mathbb{E}[X \mid B] := \int X(\omega)\mathbb{P}(\,\mathrm{d}\omega \mid B)$$

definieren. Wir bezeichnen die als <u>bedingten Erwartungswert</u> von X. Für $X = \mathbb{1}_A$ folgt (für $\mathbb{P}(B) > 0$)

$$\int X(\omega)\mathbb{P}(\,\mathrm{d}\omega \mid B) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{E}[\mathbb{1}_{A\cap B}]}{\mathbb{P}(B)} = \frac{\mathbb{E}[X\mathbb{1}_B]}{\mathbb{P}(B)}$$

und mittels maßtheoretischer Induktion folgt

$$\mathbb{E}[X \mid B] = \frac{\mathbb{E}[X\mathbb{1}_B]}{\mathbb{P}(B)}$$

allgemein $(X \in \mathscr{L}^1, \mathbb{P}(B) > 0)$

<u>Frage:</u> (Wie) können wir bedingte Erwartungswerte definieren, wenn $\mathbb{P}(B) = 0$?

1. Bedingte Verteilungen

Seien X, Y Zufallsvariablen gegeben durch

$$X : (\Omega, \mathscr{F}) \to (\Omega_X, \mathscr{F}_X)$$
$$Y : (\Omega, \mathscr{F}) \to (\Omega_Y, \mathscr{F}_Y)$$

- Falls Ω_Y höchstens abzählbar ist, gilt ($\nearrow$ **???**)

$$\mathbb{P}(X \in A \mid Y = y) = \begin{cases} \frac{\mathbb{P}(X\in A, Y=y)}{\mathbb{P}(Y=y)} & \mathbb{P}(Y = y) > 0 \\ 0/\text{sonst} & \mathbb{P}(Y = y) = 0 \end{cases}$$

Insbesondere folgt mit dem Satz der totalen Wahrscheinlichkeit

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \sum_{y \in B} \mathbb{P}(X \in A \mid Y = y)\mathbb{P}(Y = y) \quad \forall B \in \mathscr{F}_Y \\ &= \int_B \mathbb{P}(X \in A \mid Y = y)\mathbb{P}_Y(\,\mathrm{d}y) \end{aligned} \tag{$\star$}$$

- Idee: Verwende ($\star$) um bedingte Verteilung zu definieren!
  Sei also $\mu_A : \Omega_Y \to \mathbb{R}$ gegeben so dass

  $$\mathbb{P}(X \in A, Y \in B) = \int_B \mu_A(y)\mathbb{P}_Y(\,\mathrm{d}y) \quad \forall B \in \mathscr{F}_Y \tag{$\star\star$}$$

  Da Ω_Y abzählbar, gilt $\{y\} \in \mathscr{F}_Y \quad \forall y \in \Omega_Y$. Also folgt aus ($\star\star$), dass

  $$\begin{aligned}\mathbb{P}(X \in A, Y = y) &= \int_{\{y\}} \mu_A(y)\mathbb{P}_Y(\,\mathrm{d}y) \\ &= \mu_A(y)\mathbb{P}_Y(Y = y)\end{aligned}$$

  Falls $\mathbb{P}(Y = y) \neq 0$ folgt sofort

  $$\mu_A(y) = \frac{\mathbb{P}(X \in A, Y = y)}{\mathbb{P}(Y = y)} = \mathbb{P}(X \in A \mid Y = y)$$

  Andererseits gilt

  $$\mathbb{P}_Y(\{y \in \Omega_Y : \mathbb{P}(Y = y) = 0\}) = \sum_{\substack{y \in \Omega_Y \\ \mathbb{P}(Y=y)=0}} \mathbb{P}_Y(\{y\}) = 0$$

  so dass

  $$\mu_A(y) = \mathbb{P}(X \in A \mid Y = y) \quad \mathbb{P}_Y \text{ f.s (d.h. bis auf } \mathbb{P}_Y\text{-Nullmengen)}$$

  bzw.

  $$\mathbb{P}_Y(\{y \colon \mu_A(y) \neq \mathbb{P}(X \in A \mid Y = y)\}) = 0$$

- Falls Ω_Y überabzählbar ist und $y \in \Omega_Y$ mit $\mathbb{P}(Y = y) = 0$ (z.B. Y hat Dichte). Wir werden sehen, dann existiert $\mu_A : (\Omega_Y, \mathscr{F}_Y) \to (\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ messbar, so dass ($\star\star$) gilt. Insbesondere ist μ_A dann bis auf $\mathbb{P}_Y$-Nullmengen eindeutig bestimmt und wir können definieren:

  $$\mathbb{P}(X \in A \mid Y = y) = \mu_A(y)$$

Wir benötigen:

Satz VI.1 (Radon-Nikodym für endliche Maße)
Seien μ, ν zwei endliche Maße auf $(\Omega, \mathscr{F})$. Dann ist ν absolut stetig bzgl. μ ($\nu \ll \mu$) genau dann wenn ν eine messbare Dichte f bezüglich μ besitzt, d.h. wenn

$$\nu(A) = \int_A f(\omega) \, \mathrm{d}\mu(\omega) \quad \forall A \in \mathscr{F}.$$

Insbesondere ist f μ-f.ü. eindeutig bestimmt.

Beweis. $\nearrow$ MINT Schilling Satz 19.2. □

Folgerung VI.2
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und

$$X : (\Omega, \mathscr{F}) \to (\Omega_X, \mathscr{F}_X)$$
$$Y : (\Omega, \mathscr{F}) \to (\Omega_Y, \mathscr{F}_Y)$$

Zufallsvariablen. Sei $A \in \mathscr{F}_X$ beliebig. Dann existiert $\mu_A : (\Omega_Y, \mathscr{F}_Y) \to ([0,1], \mathscr{B}([0,1]))$ messbar, so dass

$$\mathbb{P}(X \in A, Y \in B) = \int_B \mu_A(y) \mathbb{P}_Y(\,\mathrm{d}y) \quad \forall B \in \mathscr{F}_Y.$$

Wir nennen $\mathbb{P}(X \in A \mid Y = y)$ bedingte Wahrscheinlichkeit.

Beweis. Offensichtlich impliziert $\mathbb{P}(y \in B) = 0$ auch $\mathbb{P}(X \in A, Y \in B) = 0$ so dass

$$\mathbb{P}(X \in A, Y \in \cdot) \ll \mathbb{P}(Y \in \cdot) = \mathbb{P}_Y(\cdot)$$

Nach Satz VI.1 existiert eine messbare Funktion $f : (\Omega_Y, \mathscr{F}_Y) \to (\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$ mit

$$\mathbb{P}(X \in A, Y \in B) = \int_B f(y) \mathbb{P}_Y(\,\mathrm{d}y) \quad \forall B \in \mathscr{F}_Y.$$

Sei $D = \{y : f(y) > 1\}$, dann gilt zudem $\mathbb{P}_Y(D) = 0$, denn

$$\mathbb{P}(Y \in D) \geq \mathbb{P}(X \in A, Y \in D) = \int_D f(y) \mathbb{P}_Y(\,\mathrm{d}y)$$

impliziert

$$0 \geq \mathbb{P}(X \in A, Y \in D) - \mathbb{P}(Y \in D) = \int_D \underbrace{(f(y) - 1)}_{>0 \text{ in } D} \mathbb{P}_Y(\,\mathrm{d}y) \geq 0$$

also gilt $\mathbb{P}(D) = 0$. Setze also

$$\mu_A(y) := \begin{cases} f(y) & y \in D^C \\ 0 & y \in D, \end{cases}$$

dann erfüllt μ_A allen Eigenschaften. □

Für fixiertes $A \in \mathscr{F}_X$ ist die nun definierte bedingte Wahrscheinlichkeit eindeutig bis auf $\mathbb{P}_Y$-Null -mengen. Für fixiertes y (und A variierend) ist $\mathbb{P}(X \in A \mid Y = y)$ aber nicht immer ein Wahrscheinlichkeitsmaß!

■ **Beispiel VI.3**

Betrachte den Wahrscheinlichkeitsraum $([0,1], \mathscr{B}([0,1]), \mathrm{U}([0,1])$ und die Zufallsvariablen X, Y mit $X(\omega) = Y(\omega) = \omega \quad \forall \omega \in [0,1]$. Dann

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(Y \in A \cap B) = \int_B \mathbb{1}_A(\omega) \mathbb{P}_Y(\,\mathrm{d}\omega) \quad \forall B \in \mathscr{F}$$

so dass $\mathbb{1}_A(\omega)$ eine Version von $\mathbb{P}(X \in A \mid Y = y)$. Insbesondere ist $\mathbb{1}_A(y)$ für jedes $y \in [0,1]$ ein Wahrscheinlichkeitsmaß. Setzen wir

$$f(A) = \begin{cases} \sup A & A \neq \varnothing \\ 0 & A = \varnothing \end{cases}$$

so ist auch

$$\mathbb{P}'(X \in A \mid Y = y) = \mathbb{1}_A(y) + \mathbb{1}_{\{f(y)\}}(y)$$

eine Version der bedingten Wahrscheinlichkeit, denn

$$\begin{aligned} \int_B (\mathbb{1}_A(\omega) + \mathbb{1}_{\{f(y)\}}(\omega)) &= \mathbb{P}(Y \in A \cap B) + \underbrace{\cancel{\mathbb{P}(Y \in B \cap f(A))}}_{=0} \\ &= P(X \in A, Y \in B) \end{aligned}$$

Allerdings ist $\mathbb{P}'(X \in \cdot \mid Y = y)$ kein Wahrscheinlichkeitsmaß, denn für beliebiges $y \in [0,1]$ gilt

$$\mathbb{P}'(X \in [0,y] \mid Y = y) = \mathbb{1}_{[0,y]} + \mathbb{1}_{f([0,y])} = 2$$

Wie können solche Maße wie im Beispiel VI.3 ausgeschlossen werden? Dadurch ist folgende Definition motiviert.

Definition VI.4 (reguläre bedingte Verteilung)
Eine bedingte Verteilung $\mathbb{P}(X \in \cdot \mid Y = \cdot)$ heißt regulär, wenn $\mathbb{P}(X \in \cdot \mid Y = y)$ für alle $y \in \Omega_Y$ ein Wahrscheinlichkeitsmaß ist.

Die Existenz regulärer bedingter Verteilungen ist nicht trivial. Wir beschränken uns daher auf den reellen Fall.

Satz VI.5
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X : (\Omega, \mathscr{F}) \to (\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$, $Y : (\Omega, \mathscr{F}) \to (\Omega_Y, \mathscr{F}_Y)$ Zufallsvariablen. Dann existiert eine reguläre bedingte Verteilung

$$\mathbb{P}(X \in \cdot \mid Y = \cdot).$$

Beweis. Nur Beweisskizze (ausführlich wird ergänzt!)
Sei $\tilde{\mathbb{P}}(X \in A \mid Y = \cdot)$ eine beliebige Version der bedingten Verteilung.

- Idee: "Korrigiere" $\tilde{\mathbb{P}}$ auf geeigneten $\mathbb{P}_Y$-Nullmengen um geeigneter reguläre Version zu erhalten.

1. Es gibt eine $\mathbb{P}_Y$-Nullmenge N_1, so dass

$$\tilde{\mathbb{P}}(X \in \mathbb{R}^d \mid Y = y) = 1 \quad \forall y \notin \mathbb{N}_1$$

2. Definiere $\mathscr{G}^d := \left\{ \bigcup_{i=1}^k [a,b_i] \mid a_i, b_i \in \mathscr{G}^d, k \in \mathbb{N} \right\}$. Dann gibt es eine $\mathbb{P}_Y$-Nullmenge N_2, so dass $\tilde{\mathbb{P}}(X \in \cdot \mid Y = y)$ nicht negativ und additiv auf $\mathscr{G}^d$ für $Y \notin \mathbb{N}_2$.
3. Es gibt eine $\mathbb{P}_Y$-Nullmenge N_3, so dass $\tilde{\mathbb{P}}(X \in \cdot \mid Y = y)$ für alle $y \notin N_2 \cup N_3$ σ-additiv auf $\mathscr{G}^d$ ist.
4. Sei $N = N_1 \cup N_2 \cup N_3$ $\mathbb{P}_Y$-Nullmengen. Für $y \in N^C$ existiert eine Erweiterung von $\tilde{\mathbb{P}}(X \in \cdot \mid Y = y)$ zu einem Wahrscheinlichkeitsmaß $\hat{\mathbb{P}}(X \in \cdot \mid Y = y)$ auf $\mathscr{B}(\mathbb{R}^d)$. Definiere

$$\mathbb{P}(X \in \cdot \mid Y = y) = \begin{cases} \hat{\mathbb{P}}(X \in \cdot \mid Y = y) & y \notin N \\ \mathbb{P}_0 \text{ beliebiges Wahrscheinlichkeitsmaß} & y \in N \end{cases}$$

dann ist $\mathbb{P}(Y \in \cdot \mid Y = y)$ ein Wahrscheinlichkeitsmaß.
5. $\mathbb{P}(X \in \cdot \mid Y = y)$ ist eine Version der bedingten Verteilung.

($\nearrow$ befindet sich in einer PDF-Datei auf Opal, eventuell hier hinzufügen später!) $\square$

Satz VI.6
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X, Y\colon (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariablen mit gemeinsamer Dichte $\rho_{X,Y}(x,y)$. Dann besitzen X und Y die Randdichten ("Marginaldichten")

$$\rho_X(x) = \int_{\mathbb{R}} \rho_{X,Y}(x,y)\,\mathrm{d}y \text{ und } \rho_Y(y) = \int_{\mathbb{R}} \rho_{X,Y}(x,y)\,\mathrm{d}x$$

Zudem gilt für alle $A \in \mathscr{B}(\mathbb{R}), y \in \mathbb{R}$:

$$\mathbb{P}(X \in A \mid Y = y) = \begin{cases} \int_A \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}\,\mathrm{d}x & \rho_Y(y) \neq 0 \\ 0 & \text{sonst} \end{cases}$$

Insbesondere besitzt $\mathbb{P}(X \in \cdot \mid Y = y)$ für alle $y \in \mathbb{R}$ mit $\rho_Y(y) > 0$ die Dichte (bedingte Dichte)

$$\rho_{X|Y}(x,y) = \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}$$

und ist in diesem Fall ein Wahrscheinlichkeitsmaß.

Beweis. Randdichten:

$$\begin{aligned} \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) &= \int_{\mathbb{R}} \int_A \rho_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_A \underbrace{\int_{\mathbb{R}} \rho(x,y)\,\mathrm{d}y}_{\rho_X(x)}\,\mathrm{d}x \end{aligned}$$

Zudem gilt für $A, B \in \mathscr{B}(\mathbb{R})$

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \int_B \int_A \rho_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_B \int_A \rho_{X,Y}(x,y)\mathbb{1}_{\rho_Y(y)>0}\,\mathrm{d}x\,\mathrm{d}y \\
&+ \int_B \int_A \rho_{X,Y}(x,y)\mathbb{1}_{\rho_Y(y)=0}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_B \int_A \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}\mathbb{1}_{\rho_Y(y)>0}\,\mathrm{d}x\rho_Y(y)\,\mathrm{d}y + \\
&+ \underbrace{\int_{B\cap\rho_Y(y)=0} \underbrace{\int_A \rho_{X,Y}(x,y)\,\mathrm{d}x}_{\leq\int_{\mathbb{R}}\rho_{X,Y}(x,y)\,\mathrm{d}x=\rho_Y(y)} \quad \mathrm{d}y}_{=0} \\
&= \int_B \int_A \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}\mathbb{1}_{\rho_Y(y)>0}\,\mathrm{d}x\mathbb{P}_Y(\,\mathrm{d}y),
\end{aligned}$$

so dass

$$\mathbb{P}(X \in A \mid Y = y) = \begin{cases} \int_A \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}\,\mathrm{d}x & \rho_Y(y) > 0 \\ 0 & \text{sonst} \end{cases}$$

nach Definition der bedingten Verteilung Definition VI.4. Für y mit $\rho_Y(y) > 0$ folgt

$$\mathbb{P}(X \in \mathbb{R} \mid Y = y) = \int_{\mathbb{R}} \frac{\rho_{X,Y}(x,y)}{\rho_Y(y)}\,\mathrm{d}x = \frac{1}{\rho_Y(y)}\int_{\mathbb{R}} \rho_{X,Y}(x,y)\,\mathrm{d}x = \frac{\rho_Y(y)}{\rho_Y(y)} = 1. \qquad \square$$

■ **Beispiel VI.7**

Betrachte $f(x,y) = cxe^{-x(y+1)} \quad x, y > 0 \quad c \in \mathbb{R}$ bestimmt. Damit f die Dichte zweier Zufallsvariablen X, Y ist, muss

$$\int_{\mathbb{R}_+} \int_{\mathbb{R}_+} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \overset{!}{=} 1$$

gelten. Dazu sei

$$\begin{aligned}
\int_0^\infty \int_0^\infty cxe^{-x(y+1)}\,\mathrm{d}y\,\mathrm{d}x &= c\int_0^\infty xe^{-x}\int_0^\infty e^{-xy}\,\mathrm{d}y\,\mathrm{d}x \\
&= c\int_0^\infty xe^{-x}\left[-\frac{1}{x}e^{-xy}\right]_{y=0}^\infty \\
&= c\underbrace{\int_0^\infty e^{-x}\,\mathrm{d}x}_{=1} = c \overset{!}{=} 1
\end{aligned}$$

Wähle $x = 1$. Als Randdichten von X und Y folgen

$$\begin{aligned}
f_X(x) &= \int_0^\infty f(x,y)\,\mathrm{d}y = \int_0^\infty xe^{-x(y+1)}\,\mathrm{d}y \\
&\overset{\text{s.o.}}{=} e^{-x} \quad x > 0
\end{aligned}$$

(d.h. $X \sim \mathrm{Exp}(1)$) und

$$\begin{aligned} f_Y(y) &= \int_0^\infty f(x,y)\,\mathrm{d}x = \int_0^\infty x e^{-x(y+1)}\,\mathrm{d}x \\ &\overset{\text{P.I.}}{=} \left[-x\frac{1}{y+1}e^{-x(y+1)}\right]_{x=0}^\infty + \frac{1}{y+1}\int_0^\infty e^{-x(y+1)}\,\mathrm{d}x \\ &= 0 + \frac{1}{y+1}\left[\frac{-1}{y+1}e^{-x(y+1)}\right]_{x=0}^\infty \\ &= \frac{1}{(y+1)^2} \qquad y>0 \end{aligned}$$

(Dies ist die Dichte einer Pareto-Verteilung.) Insbesondere sind X und Y nicht unabhängig, da sonst $f(x,y) = f_X(x)f_Y(y)$ gelten müsste. Die bedingte Dichte von X gegeben $Y = y_0, y_0 > 0$ berechnet sich als

$$f_{X\mid Y=y_0}(x) = \frac{f(x,y_0)}{f_Y(y_0)} = x(y_0+1)^2e^{-x(y_0+1)}$$

und dies ist die Dichte einer Gamma$(y_0+1,2)$-Verteilung, d.h.

$$X \mid Y = y_0 \sim \mathrm{Gamma}(y_0+1,2).$$

2. Bedingte Erwartungswerte

Satz VI.8
$(\Omega,\mathscr{F},\mathbb{P})$ Wahrscheinlichkeitsraum und $X\colon(\Omega,\mathscr{F}) \to (\mathbb{R},\mathscr{B}(\mathbb{R}))$ und $Y\colon(\Omega,\mathscr{F}) \to (\Omega_Y,\mathscr{F}_Y)$ Zufallsvariable, so dass $\mathbb{E}[X]$ existiert ($X \in \mathscr{L}^1(\mathbb{P})$ oder $X \geq 0$). Dann existiert eine messbare Funktion $g\colon(\Omega_Y,\mathscr{F}_Y) \to (\bar{\mathbb{R}},\mathscr{B}(\bar{\mathbb{R}}))$, so dass gilt

$$\mathbb{E}[X\mathbb{1}_{y\in B}] = \int_{y\in B} X\,\mathrm{d}\mathbb{P} = \int_B g(y)\mathbb{P}_Y(\,\mathrm{d}y) \quad \forall B \in \mathscr{F}_Y$$

Dieses g ist $\mathbb{P}_Y$-f.s. (fast sicher) eindeutig bestimmt. Wir nennen $\mathbb{E}[X \mid Y = y] = g(y)$ <u>bedingten Erwartungswert</u>/<u>bedingte Erwartung</u> von X gegeben $Y = y$.

Beweis.

1. Sei $X \geq 0$. Definiere ein Maß μ auf $\mathscr{F}_Y$ durch

$$\mu(B) = \mathbb{E}[X\mathbb{1}_{y\in B}]$$

dann ist $\mu \ll \mathbb{P}_Y$. Aus dem Satz von Radon-Nikodym (Satz VI.1) folgt die Existenz einer $\mathbb{P}_Y$-f.s. bestimmten Dichte g.

2. Sei $X \in \mathscr{L}^1(\mathbb{P})$. Es gilt $\mathbb{E}[X^-] < \infty$ oder $\mathbb{E}[X^+] < \infty$. Sei oBdA $\mathbb{E}[X^-] < \infty$, dann folgt auch $\mathbb{E}[X^- \mid Y = y] < \infty$ $\mathbb{P}_Y$ f.s.. Setze

$$\mathbb{E}[X \mid Y = y] := \begin{cases} \mathbb{E}[X^+ \mid Y = y] - \mathbb{E}[X^- \mid Y = y] & \mathbb{E}[X^- \mid Y = y] < \infty \\ 0 & \text{sonst.} \end{cases}$$

Dann folgt

$$\begin{aligned}
\mathbb{E}[X\mathbb{1}_{y\in B}] &= \mathbb{E}[X^+\mathbb{1}_{y\in B}] - \mathbb{E}[X^-\mathbb{1}_{y\in B}] \\
&= \int_B \mathbb{E}[X^+ \mid Y = y]\mathbb{P}_Y(\,\mathrm{d}y) - \int_B \mathbb{E}[X^- \mid Y = y]\mathbb{P}_Y(\,\mathrm{d}y) \\
&= \int_B \mathbb{E}[X \mid Y = y]\mathbb{P}_Y(\,\mathrm{d}y) \quad \forall B \in \mathscr{F}_Y.
\end{aligned}$$

Sind g_1, g_2 zwei Versionen der bedingten Erwartung, dann gilt wegen $\mathbb{E}[X^-] < \infty$ auch $\int_{\Omega_Y} g_1^-(y)\mathbb{P}_Y(\,\mathrm{d}y) < \infty$ für $i = 1, 2$ und dann folgt

$$\begin{aligned}
\int_B g_1^+(y)\mathbb{P}_Y(\,\mathrm{d}y) &= \int_B g_1^-(y)\mathbb{P}_Y(\,\mathrm{d}y) \\
&= \int_B g_1(y)\mathbb{P}_Y(\,\mathrm{d}y) \\
&= \int_B g_2(y)\mathbb{P}_Y(\,\mathrm{d}y) \\
&= \int_B g_2^+(y)\mathbb{P}_Y(\,\mathrm{d}y) - \int_B g_2^-(y)\mathbb{P}_Y(\,\mathrm{d}y) \\
&\implies \int_B (g_1^+(y) + g_2^+(y))\mathbb{P}_Y(\,\mathrm{d}y) \\
&= \int_B (g_2^+(y) + g_1^-(y))\mathbb{P}_Y(\,\mathrm{d}y) \quad \forall B \in \mathscr{F}_Y
\end{aligned}$$

also

$$g_1^+ + g_2^+ = g_2^+ + g_1^- \, \mathbb{P}_Y\text{-f.s.}$$

impliziert

$$g_1^+ - g_1^- = g_1 = g_2 = g_2^+ - g_2^- \;\; \mathbb{P}_Y\text{-fast-sicher.}$$

□

▶ **Bemerkung**

Bedingte Erwartung und bedingte Verteilung hängen zusammen:

- Sind X, Y so dass $\mathbb{E}X$ existiert und eine reguläre bedingte Verteilung $\mathbb{P}_{X|Y=y}(B) = \mathbb{P}(X \in B \mid Y = y)$ existiert, dann folgt

$$\mathbb{E}[X \mid Y = y] = \int_{\mathbb{R}} x\mathbb{P}_{X|Y=y}(\,\mathrm{d}x).$$

Für Treppenfunktionen $X = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}, \alpha_i > 0, A_i$ disjunkt, folgt dies aus

$$\begin{aligned}
\mathbb{E}[X \mathbb{1}_{y \in B}] &= \sum_{i=1}^{n} \alpha_i \mathbb{E}[\mathbb{1}_{A_i} \mathbb{1}_{y \in B}] \\
&= \sum_{i=1}^{n} \alpha_i \mathbb{P}(X = \alpha_i, y \in B) \\
&= \sum \alpha_i \int_B \mathbb{P}(X = \alpha_i \mid Y = y) \mathbb{P}_Y(\,\mathrm{d}y) \\
&= \int_B \sum \alpha_i \mathbb{P}(X = \alpha_i \mid Y = y) \mathbb{P}_Y(\,\mathrm{d}y) \\
&= \int_B \int_{\mathbb{R}} x \mathbb{P}_{X|Y=y}(\,\mathrm{d}x) \mathbb{P}(\,\mathrm{d}y)
\end{aligned}$$

Für allgemeines X folgt die Aussage mittels maßtheoretischer Induktion.

Durch Einsetzen von $X = \mathbb{1}_A$ in Gleichung in (**??**) folgt

$$\begin{aligned}
\int_B \mathbb{E}[\mathbb{1}_A \mid Y = y] \mathbb{P}_Y(\,\mathrm{d}y) = \int_{Y \in B} \mathbb{1}_A \,\mathrm{d}\mathbb{P} \\
= \mathbb{E}[\mathbb{1}_A \mathbb{1}_{y \in B}] \\
= \mathbb{P}(A \cap \{Y \in B\})
\end{aligned}$$

und durch Vergleich mit Definition der bedingten Verteilung (Definition VI.4)

$$\mathbb{E}[\mathbb{1}_A \mid Y = y] = \mathbb{P}(A \mid Y = y) \quad \mathbb{P}\text{-f.s.}.$$

■ **Beispiel VI.9**

Betrachte Beispiel VI.7, d.h. gegeben seien zwei Zufallsvariablen X, Y mit Dichte

$$f(x, y) = x e^{-x(y+1)} \quad x, y > 0$$

Die Erwartungswerte von X und Y folgen aus den Randdichten:

$$\mathbb{E}[X] = 1 \qquad \text{da } X \sim \mathrm{Exp}(1)$$

und

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^\infty y f_y(y) \,\mathrm{d}y = \int_0^\infty y \frac{1}{(y+1)^2} \\
&= \int_0^\infty (z-1) \frac{1}{z^2} \,\mathrm{d}z \\
&= \int_1^\infty \frac{1}{z} \,\mathrm{d}z - \int_1^\infty \frac{1}{z^2} \,\mathrm{d}z = \infty
\end{aligned}$$

Der bedingte Erwartungswert von X gegeben $Y = y_0 > 0$ ist

$$\begin{aligned}\mathbb{E}[X \mid Y = y_0] &= \int_0^\infty x f_{X|Y=y_0}(x)\,\mathrm{d}x \\ &= \int_0^\infty x^2 (y_0+1)^2 e^{-x(y_0+1)}\,\mathrm{d}x \\ &= \frac{2}{y_0+1}\end{aligned}$$

da $X \mid Y = y_0 \sim \text{Gamma}(y_0+1, 2)$.

2.1. Der bedingte Erwartungswert als Zufallsvariable

Bisher: Bedingung der Form: $Y = y \implies$ Sowohl bedingte Verteilung

$$\mu_A : (\Omega_Y, \mathscr{F}_Y) \to ([0,1], \mathscr{B}([0,1]))$$

als auch bedingte Erwartung

$$g : (\Omega_Y, \mathscr{F}_Y) \to (\bar{\mathbb{R}}, \mathscr{B}(\bar{\mathbb{R}}))$$

sind messbar.

Wir können also auch die Zufallsvariablen

$$\begin{aligned}\mu_A(Y) &=: \mathbb{P}(X \in A \mid Y) \\ g(Y) &=: \mathbb{E}[X \mid Y]\end{aligned}$$

betrachten.

Definition VI.10 (bedingte Erwartung und Wahrscheinlichkeit)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ und $Y : (\Omega, \mathscr{F}) \to (\Omega_Y, \mathscr{F}_Y)$ Zufallsvariablen, so dass $\mathbb{E}[X]$ existiert. Die bedingte Erwartung von X gegeben Y ist die Zufallsvariable $g(Y)$ mit

$$\begin{aligned}\int_{\{Y \in B\}} X(\omega)\mathbb{P}(\,\mathrm{d}\omega) &= \int_{\{Y \in B\}} g(Y)(\omega)\mathbb{P}(\,\mathrm{d}\omega) \\ &= \int_B g(y)\mathbb{P}_Y(\,\mathrm{d}y) \quad \forall B \in \mathscr{F}_Y.\end{aligned}$$

Schreibe $g(Y) =: \mathbb{E}[X \mid Y] =: \mathbb{E}[X \mid \sigma(Y)]$.

$$\mathbb{P}(A \mid \sigma(Y)) := \mathbb{P}(A \mid Y) = \mathbb{E}[\mathbb{1}_A \mid Y]$$

ist die bedingte Wahrscheinlichkeit von A gegeben Y.

2.2. Bedingen auf beliebige σ-Algebren

Definition VI.11
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariablen, so dass $\mathbb{E}[X]$ existiert und $\mathscr{G} \subset \mathscr{F}$ σ-Algebra. Die bedingte Erwartung von X gegeben $\mathscr{G}$ ist die $\mathscr{G}$-messbare Zufallsvariable $X_{\mathscr{G}}$ mit

$$\int_G X(\omega)\mathbb{P}(\,\mathrm{d}\omega) = \int_G X_{\mathscr{G}}(\omega)\mathbb{P}(\,\mathrm{d}\omega) \quad \forall G \in \mathscr{G} \tag{$\star$}$$

Schreibe: $X_{\mathscr{G}} = \mathbb{E}[X \mid \mathscr{G}]$. $\mathbb{P}(A \mid \mathscr{G}) := \mathbb{E}[\mathbb{1}_A \mid \mathscr{G}]$ heißt bedingte Wahrscheinlichkeit von A gegeben $\mathscr{G}$.

▶ **Bemerkung**

- Existenz und Eindeutigkeit (bis auf $\mathscr{G}$-Nullmengen) der bedingten Erwartung folgen wieder aus Radon-Nikodym (Satz VI.1).
- Die Gleichung ($\star$) können wir umschreiben:

$$\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G] \quad \forall G \in \mathscr{G}$$

- $\mathbb{E}[X \mid \mathscr{G}]$ ist nur bis auf $\mathscr{G}$-Nullmengen bestimmt. Spreche daher von "Versionen" der bedingten Erwartung.

Die bedingte Erwartung übernimmt Eigenschaften der Erwartung:

Lemma VI.12 (Rechenregeln bedingter Erwartungswert 1)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $\mathscr{G} \subseteq \mathscr{F}$ σ-Algebra, $X, Y : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariablen, so dass $\mathbb{E}[X], \mathbb{E}[Y]$ existieren.

1. Positivität:
$$X \geq 0 \quad \mathbb{P}\text{-f.s.} \quad \Longrightarrow \mathbb{E}[X \mid \mathscr{G}] \geq 0 \quad \mathbb{P}\text{-f.s.}$$
2. Konservativität:
$$X \equiv c \in \mathbb{R} \implies \mathbb{E}[X \mid \mathscr{G}] = c \quad \mathbb{P}\text{-f.s.}$$
3. Linearität: Für $a, b \in \mathbb{R}$ und $X, Y \in \mathscr{L}^1(\mathbb{P})$
$$\mathbb{E}[aX + bY \mid \mathscr{G}] = a\mathbb{E}[X \mid \mathscr{G}] + b\mathbb{E}[Y \mid \mathscr{G}] \quad \mathbb{P}\text{-f.s.}$$
4. Monotonie:
$$X \leq Y \implies \mathbb{E}[X \mid \mathscr{G}] \leq \mathbb{E}[Y \mid \mathscr{G}] \quad \mathbb{P}\text{-f.s.}$$

Beweis.

1. Folgt aus Definition VI.11.
2. Betrachte

$$\int_G X(\omega)\mathbb{P}(\,\mathrm{d}\omega) = \int_G x\mathbb{P}(\,\mathrm{d}\omega)$$

$\implies$ c ist Version der bedingten Erwartung.

3. Betrachte

$$\begin{aligned}\int_G (aX+bY)\,\mathrm{d}\mathbb{P} &= a\int_G X\,\mathrm{d}\mathbb{P} + b\int_G Y\mathbb{P} \text{ nutze Linearität des Integrals}\\ &\overset{(\star)}{=} a\int_G \mathbb{E}[X \mid \mathscr{G}]\,\mathrm{d}\mathbb{P} + b\int_G \mathbb{E}[Y \mid \mathscr{G}]\,\mathrm{d}\mathbb{P}\\ &= \int_G (a\mathbb{E}[X \mid \mathscr{G}] + b\mathbb{E}[Y \mid \mathscr{G}])\,\mathrm{d}\mathbb{P}\end{aligned}$$

$\implies$ $a\mathbb{E}[X \mid \mathscr{G}] + b\mathbb{E}[Y \mid \mathscr{G}]$ ist Version von $\mathbb{E}[aX+bY \mid \mathscr{G}]$

4. $X \leq Y \implies Y - X \geq 0$ und mit 1. und 3. $\implies$ Behauptung. □

Satz VI.13 (Konvergenzsätze der bedingten Erwartung)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum $\mathscr{G} \subseteq \mathscr{F}$ σ-Algebra, $X, X_1, X_2, \dots : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariablen

1. Beppo-Levi bedingt:

$$X_n \geq 0 \quad X_n \in \mathscr{L}^1(\mathbb{P}) \quad X_n \uparrow X \text{ und } \sup_{n\in\mathbb{N}} \mathbb{E}[X_n] < \infty \quad \mathbb{P}\text{-f.s.}$$

   Dann gilt: $\mathbb{E}[X \mid \mathscr{G}] = \sup_{n\in\mathbb{N}} \mathbb{E}[X_n \mid \mathscr{G}] = \lim_{n\to\infty} \mathbb{E}[X_n \mid \mathscr{G}] \quad \mathbb{P}$-f.s.

2. Fatou bedingt: $X_n \geq 0, X_n \in \mathscr{L}^1(\mathbb{P})$ und $\liminf_{n\to\infty} \mathbb{E}[X_n] < \infty$
   Dann gilt

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \mid \mathscr{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n \mid \mathscr{G}] \quad \mathbb{P}\text{-f.s.}$$

3. Dominierte-Konvergenz bedingt: $\lim_{n\to\infty} X_n = X \quad \mathbb{P}$-f.s. und $|X_n| \leq Y \quad \mathbb{P}$-f.s. mit $Y \in \mathscr{L}^1(\mathbb{P})$, dann gilt $X \in \mathscr{L}^1(\mathbb{P})$ und

$$\lim_{n\to\infty} \mathbb{E}[X_n \mid \mathscr{G}] = \mathbb{E}[X \mid \mathscr{G}] \quad \mathbb{P}\text{-f.s.}$$

Beweis.

1. Aus Beppo-Levi (klassisch) folgt:

$$\mathbb{E}[X] = \sup_{n\in\mathbb{N}} \mathbb{E}[X_n] < \infty$$

   also $X \in \mathscr{L}^1(\mathbb{P})$ und damit existiert $\mathbb{E}[X \mid \mathscr{G}]$. $\mathbb{E}[X_n \mid \mathscr{G}]$ ist monoton wachsend (Lemma VI.12). Also

$$\begin{aligned}\int_G \sup_{n\in\mathbb{N}} \mathbb{E}[X_n \mid \mathscr{G}]\,\mathrm{d}\mathbb{P} &= \sup_{n\in\mathbb{N}} \int_G \mathbb{E}[X_n \mid \mathscr{G}]\,\mathrm{d}\mathbb{P}\\ &\overset{\text{Definition VI.11}}{=} \sup_{n\in\mathbb{N}} \int_G X_n\,\mathrm{d}\mathbb{P}\\ &= \int_G \sup_{n\in\mathbb{N}} X_n\,\mathrm{d}\mathbb{P}\end{aligned}$$

2. ↗ HA 8.2a

3. ↗ HA 8.2b □

Satz VI.14 (Rechenregeln bedingter Erwartungswert 2)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $\mathscr{G} \subseteq \mathscr{F}$ σ-Algebra und $X, Y : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariablen, so dass $\mathbb{E}[X]$ existiert. Es gelten:

1. Triviale σ-Algebra:

$$\mathbb{E}[X \mid \{\varnothing, \Omega\}] = \mathbb{E}[X]$$

2. einfache Turmeigenschaft:

$$\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]] = \mathbb{E}[X]$$

3. allgemeine Turmeigenschaft: Sei $\mathscr{H} \subseteq \mathscr{G} \subseteq \mathscr{F}$ σ-Algebra, dann

$$\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}] = \mathbb{E}[X \mid \mathscr{H}] \quad \mathbb{P}\text{-f.s.}$$

4. Messbares Herausziehen: Ist Y $\mathscr{G}$-messbar und gilt $\mathbb{E}[XY] < \infty$, dann

$$\mathbb{E}[XY \mid \mathscr{G}] = Y\mathbb{E}[X \mid \mathscr{G}]$$

Beweis. 1. Offenbar ist $\{\varnothing, \Omega\}$ σ-Algebra. Es gilt:

$$\int_G X \, d\mathbb{P} = \begin{cases} 0 & G = \varnothing \\ \mathbb{E}[X] & G = \Omega \end{cases}$$

und

$$\int_G \mathbb{E}[X] \, d\mathbb{P} = \begin{cases} 0 & G = \varnothing \\ E[X] & G = \Omega \end{cases}$$

Also ist $\mathbb{E}[X]$ eine Version von $\mathbb{E}[X \mid \{\varnothing, \Omega\}]$.

2. Sei

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]] &\overset{\text{Def. EW}}{=} \int_\Omega \mathbb{E}[X \mid \mathscr{G}] \, d\mathbb{P} \\ &\overset{VI.11, \Omega \in \mathscr{G}}{=} \int_\Omega X \, d\mathbb{P} = \mathbb{E}[X] \end{aligned}$$

3. Für $H \in \mathscr{H}$ gilt

$$\begin{aligned} \int_H \mathbb{E}[X \mid \mathscr{G}] \, d\mathbb{P} &\overset{VI.11, H \in \mathscr{G}}{=} \int_H X \, d\mathbb{P} \\ &\overset{VI.11}{=} \int_H \mathbb{E}[X \mid \mathscr{H}] \, d\mathbb{P} \end{aligned}$$

also ist $\mathbb{E}[X \mid \mathscr{H}]$ eine Version von $\mathbb{E}[\mathbb{E}[X \mid \mathscr{G}] \mid \mathscr{H}]$.

4. Sei zunächst $Y = \mathbb{1}_B$ mit $B \in \mathscr{G}$. Dann folgt für $G \in \mathscr{G}$

$$\begin{aligned}\int_G Y\mathbb{E}[X \mid \mathscr{G}]\,\mathrm{d}\mathbb{P} &= \int_{G\cap B} \mathbb{E}[X \mid \mathscr{G}]\,\mathrm{d}\mathbb{P} \\ &\overset{VI.11, G\cap B\in\mathscr{G}}{=} \int_{G\cap B} X\,\mathrm{d}\mathbb{P} \\ &= \int_G XY\,\mathrm{d}\mathbb{P} \\ &\overset{VI.11}{=} \int_G \mathbb{E}[XY \mid \mathscr{G}]\,\mathrm{d}\mathbb{P}\end{aligned}$$

und $Y\mathbb{E}[X \mid \mathscr{G}]$ ist Version von $\mathbb{E}[XY \mid \mathscr{G}]$. Die allgemeine Aussage folgt mit maßtheoretischer Induktion. □

■ **Beispiel VI.15**

Ein KFZ-Versicherungsunternehmen modelliert die Schäden, die aus einem Portfolio von Verträgen stammen. Alle Zufallsvariablen $X_1, X_2, X_3, \ldots$ sind unabhängig und identisch verteilt. (u.i.v) Zudem sei N eine Zufallsvariable mit Werten in $\mathbb{N}_0$, welche die Anzahl der Schäden in einem Jahr modelliert. N und $\{X_1, X_2, \ldots\}$ seien unabhängig. Der Gesamtschaden im Jahr des Portfolios ist dann

$$S = \sum_{i=1}^{N} X_i$$

Der erwartete Gesamtschaden ist

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}[\sum_{i=1}^{N} X_i] \\ &= \mathbb{E}[\mathbb{E}[\sum_{i=1}^{N} X_i \mid N]] \quad \text{(Turmeigs.)} \\ &= \sum_{n=0}^{\infty} \mathbb{E}[\sum_{i=1}^{N} \mid N = n]\mathbb{P}(N = n) \quad \text{(Def. EW.)} \\ &= \sum_{n=0}^{\infty} \mathbb{E}[\sum_{i=1}^{n} X_i]\mathbb{P}(N = n) \quad (X_i \perp\!\!\!\perp N) \\ &= \sum_{n=0}^{\infty}\sum_{i=1}^{n} \mathbb{E}[X_i]\mathbb{P}(N = n) \quad \text{(Linearität} \\ &= \sum_{n=0}^{\infty}\sum_{i=1}^{n} \mathbb{E}[X_1]\mathbb{P}(N = n) \quad (X_i) \text{ identisch verteilt)} \\ &= \mathbb{E}[X_1]\sum_{n=0}^{\infty} n\mathbb{P}(N = n) \\ &= \mathbb{E}[X_1] \cdot \mathbb{E}[N] \quad \text{(Wald-Gleichung)}\end{aligned}$$

Bedingte Varianz

Definition VI.16 (Bedingte Varianz)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $\mathscr{G} \subseteq \mathscr{F}$ σ-Algebra. $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit $X \in \mathscr{L}^2(\mathbb{P})$. Dann ist

$$\mathbb{V}\mathrm{ar}(X \mid \mathscr{G}) = \mathbb{E}\big[(X - \mathbb{E}[X \mid \mathscr{G}])^2 \mid \mathscr{G}\big]$$

die bedingte Varianz von X gegeben $\mathscr{G}$.

Lemma VI.17 (Eigenschaft der bedingten Varianz)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $\mathscr{G} \subseteq \mathscr{F}$ σ-Algebra, $X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariable in $\mathscr{L}^2(\mathbb{P})$. Es gelten:

1. Positivität: $\mathbb{V}\mathrm{ar}(X \mid \mathscr{G}) \geq 0$ $\mathbb{P}$-f.s.
2. Für $a, b \in \mathbb{R}$:
$$\mathbb{V}\mathrm{ar}(aX + b \mid \mathscr{G}) = a^2 \mathbb{V}\mathrm{ar}(X \mid \mathscr{G})$$
3. Verschiebungssatz:
$$\mathbb{V}\mathrm{ar}(X \mid \mathscr{G}) = \mathbb{E}[X^2 \mid \mathscr{G}] - (\mathbb{E}[X \mid \mathscr{G}])^2$$

Beweis. 1. Definition und Positivitätseigenschaft der bedingten Erwartung.
2. Linearität der bedingten Erwartung und Rechenregeln wie im klassischen Fall.
3. Betrachte

$$\begin{aligned}
&\mathbb{E}[(X - \mathbb{E}[X \mid \mathscr{G}])^2 \mid \mathscr{G}] = \\
&\quad = \mathbb{E}[X^2 \mid \mathscr{G}] - 2\mathbb{E}[X \underbrace{\mathbb{E}[X \mid \mathscr{G}]}_{\mathscr{G}\text{mb}} \mid \mathscr{G}] - \mathbb{E}[(\mathbb{E}[X \mid \mathscr{G}])^2 \mid \mathscr{G}] \\
&\quad \overset{\text{mb herausziehen}}{=} \mathbb{E}[X^2 \mid \mathscr{G}] - 2\mathbb{E}[X \mid \mathscr{G}]\mathbb{E}[X \mid \mathscr{G}] + (\mathbb{E}[X \mid \mathscr{G}])^2 \quad \underbrace{\mathbb{E}[1 \mid \mathscr{G}]}_{1,\ \text{Konservation}} \\
&\quad = \mathbb{E}[X^2 \mid \mathscr{G}] - (\mathbb{E}[X \mid \mathscr{G}])^2.
\end{aligned}$$

□

Lemma VI.18 (Varianzzerlegung)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $\mathscr{G} \subseteq \mathscr{F}, X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ Zufallsvariable in $\mathscr{L}^2(\mathbb{P})$. Dann:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}[\mathbb{V}\mathrm{ar}(X \mid \mathscr{G})] + \mathbb{V}\mathrm{ar}(\mathbb{E}[X \mid \mathscr{G}])$$

Beweis. Sei

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(\mathbb{E}[X \mid \mathscr{G}]) &= \mathbb{E}[(\mathbb{E}[X \mid \mathscr{G}])^2] - \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]]^2 \\
&\overset{\text{Turmegs.}}{=} \mathbb{E}[(\mathbb{E}[X \mid \mathscr{G}])^2] - (\mathbb{E}[X])^2
\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\mathbb{V}\mathrm{ar}(X \mid \mathscr{G})] &\overset{VI.17}{=} \mathbb{E}[\mathbb{E}[X^2 \mid \mathscr{G}] - \mathbb{E}[X \mid \mathscr{G}]^2] \\ &= \mathbb{E}[\mathbb{E}[X^2 \mid \mathscr{G}]] - \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]^2] \\ &\overset{\text{Turm}}{=} \mathbb{E}[X^2] - \mathbb{E}[\mathbb{E}[X \mid \mathscr{G}]^2]\end{aligned}$$

Durch Addition folgt die Behauptung. □

■ **Beispiel VI.19**

Betrachte $S = \sum_{i=1}^N X_i$ aus Beispiel VI.15 mit $N, X_i \in \mathscr{L}^2(\mathbb{P})$. Aus der Varianzzerlegung folgt

$$\mathbb{V}\mathrm{ar}(S) = \mathbb{E}[\mathbb{V}\mathrm{ar}(S \mid N)] + \mathbb{V}\mathrm{ar}(\mathbb{E}[S \mid N])$$

wobei

$$\begin{aligned}\mathbb{E}[\mathbb{V}\mathrm{ar}(S \mid N)] &= \sum_{n=0}^{\infty} \mathbb{V}\mathrm{ar}(S \mid N = n)\mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathbb{V}\mathrm{ar}(\sum_{i=1}^{\infty} X_i)\mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty}\sum_{i=1}^{\infty} \mathbb{V}\mathrm{ar}(X_i)\mathbb{P}(N = n) \quad \text{Bienaymé } X_i \text{ unabh.} \\ &= \mathbb{V}\mathrm{ar}(X_1)\sum_{n=0}^{\infty} n\mathbb{P}(N = n) \\ &= \mathbb{V}\mathrm{ar}(X_1)\mathbb{E}[N]\end{aligned}$$

und

$$\begin{aligned}\mathbb{V}\mathrm{ar}(\mathbb{E}[S \mid N]) &= \mathbb{E}[(\mathbb{E}[S \mid N] - \underbrace{\mathbb{E}[\mathbb{E}[S \mid N]]}_{\mathbb{E}[S]=\mathbb{E}[X_1]\mathbb{E}[N], VI.15})^2] \quad \text{Def. Varianz} \\ &= \sum_{n=0}^{\infty} (\mathbb{E}[S \mid N = n] - \mathbb{E}[X_1]\mathbb{E}[N])^2\mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} (n\mathbb{E}[X_1] - \mathbb{E}[X_1]\mathbb{E}[N])^2\mathbb{P}(N = n) \\ &= \mathbb{E}[X_1]^2 \sum_{n=0}^{\infty} (n - \mathbb{E}[N])^2\mathbb{P}(N = n) \\ &= \mathbb{E}[X_1]^2 \cdot \underbrace{\sum_{n=0}^{\infty} n^2\mathbb{P}(N = n)}_{\mathbb{E}[N^2]} - 2\mathbb{E}[N] \cdot \underbrace{\sum_{n=0}^{\infty} n\mathbb{P}(N = n))}_{\mathbb{E}[N]} \\ &+ \mathbb{E}[N]^2 \underbrace{\sum_{n=0}^{\infty} \mathbb{P}(N = n)}_{=1} \\ &= \mathbb{E}[X_1]^2(\mathbb{E}[N^2] - \mathbb{E}[N]^2) = \mathbb{E}[X_1]^2\mathbb{V}\mathrm{ar}(N) \\ &\implies \mathbb{V}\mathrm{ar}(S) = \mathbb{V}\mathrm{ar}(X_1)\mathbb{E}[N] + (\mathbb{E}[X_1])^2\mathbb{V}\mathrm{ar}(N)\end{aligned}$$

Kapitel VII

Die Normalverteilung

Definition VII.1
Sei $\mu \in \mathbb{R}, \sigma^2 > 0$. Das Wahrscheinlichkeitsmaß $\mathcal{N}(\mu, \sigma^2)$ auf $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mit Dichtefunktion

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

heißt Normalverteilung mit Parametern μ und σ^2. Im Fall $\mu = 0, \sigma^2 = 1$, so heißt $\mathcal{N}(0,1)$ Standardnormalverteilung.

▶ **Bemerkung**

1. g_{μ,σ^2} ist Wahrscheinlichkeitsdichte, da $\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \,\mathrm{d}x = \sqrt{2\pi}$.
2. μ und σ^2 sind gerade Erwartungswert und Varianz der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$. Sei $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} \,\mathrm{d}x \quad y = \frac{x-\mu}{\sqrt{\sigma^2}} \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (\sqrt{\sigma^2} y + \mu) e^{-\frac{y^2}{2}} \,\mathrm{d}y \\
&= \sqrt{\frac{\sigma^2}{2\pi}} \underbrace{\int_{\mathbb{R}} y e^{-\frac{y^2}{2}} \,\mathrm{d}y}_{=0,\text{ Symmetrie}} + \frac{1}{\sqrt{2\pi}} \underbrace{\int_{\mathbb{R}} \mu e^{-\frac{y^2}{2}} \,\mathrm{d}y}_{\mu\sqrt{2\pi}} = \mu \\
\mathbb{V}\mathrm{ar}(X) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \,\mathrm{d}x \quad y = \frac{x-\mu}{\sqrt{\sigma^2}} \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} y^2 e^{-\frac{y^2}{2}} \,\mathrm{d}y \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{\mathbb{R}} (-y)(-y e^{-\frac{y^2}{2}}) \,\mathrm{d}y \\
&\overset{\text{P.I.}}{=} \frac{\sigma^2}{\sqrt{2\pi}} \left(\underbrace{\left[-y e^{-\frac{y^2}{2}} \right]_{-\infty}^{\infty}}_{=0} + \underbrace{\int_{\mathbb{R}} e^{-\frac{y^2}{2}} \,\mathrm{d}y}_{=\sqrt{2\pi}} \right) \\
&= \sigma^2
\end{aligned}$$

Die Popularität der Normalverteilung erklärt sich insbesondere aus dem zentralen Grenzwertsatz. Hier eine einfache Form davon

Satz VII.2 (de Moivre-Laplace, lokaler Grenzwertsatz)
Seien $p \in (0,1), B_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$ Zähldichte $\mathrm{Bin}(n,p), g(x)$ Dichte der Standardnormalverteilung. Dann gilt für $c > 0$ beliebig mit

$$x_n(k) = \frac{k - np}{\sqrt{np(1-p)}}$$

$$\lim_{n\to\infty} \max_{k:|x_n(k)|<c} \left| \frac{\sqrt{np(1-p)} \cdot B_{n,p}(k)}{g(x_n(k))} - 1 \right| = 0$$

Beweis. $\nearrow$ pdf im Opal, bzw. später als Korollar des zentralen Grenzwertsatzes. Wird eventuell noch ergänzt, wenn Zeit ist. □

Lemma VII.3 (Stirling-Formel)
Es gilt

$$n! \overset{n\to\infty}{\sim} \sqrt{2\pi n} n^n e^{-n}$$

Beweis. $\nearrow$ pdf im Opal, wird auch eventuell mit ergänzt. □

Kapitel VIII

Momenterzeugende & charakteristische Funktionen

Ziel: "Übersetze" Verteilungen in Funktionen. Insbesondere einfache Faltungsoperation ($\nearrow$ 5.3).

Definition VIII.1

1. $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum $X : \Omega \to \mathbb{R}$ Zufallsvariable. Dann heißt

$$m_X(u) := \mathbb{E}[e^{uX}] \quad u \in \mathbb{R}, \text{ so dass } m_X(u) < \infty$$

momenterzeugende Funktion von X (mgf = moment generating function)

2. Ist $\mathbb{P}$ Verteilung auf $\mathbb{R}$, so heißt

$$m_{\mathbb{P}}(u) = \int_{\mathbb{R}} e^{uX} \mathbb{P}(\mathrm{d}x) \quad u \in \mathbb{R}, \text{ so dass } m_{\mathbb{P}}(u) < \infty$$

momenterzeugende Funktion von $\mathbb{P}$.

■ **Beispiel VIII.2**

Sei $X \sim \text{Gamma}(\lambda, r)$.

$$\begin{aligned}
m_X(u) &= \mathbb{E}[e^{uX}] \\
&= \int_0^\infty e^{ux} \lambda e^{-\lambda x} \frac{(\lambda x)^{r-1}}{\Gamma(r)} \,\mathrm{d}x \\
&= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty e^{-(\lambda-u)x} x^{r-1} \,\mathrm{d}x \quad y = (\lambda - u)x \\
&= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty e^{-y} \frac{y^{r-1}}{(\lambda-u)^{r-1}} \frac{\mathrm{d}y}{(\lambda-u)} \\
&= \left(\frac{\lambda}{\lambda-u}\right)^r \quad u < \lambda
\end{aligned}$$

Lemma VIII.3

Ist X $\mathbb{N}_0$-wertig, so gilt für alle $u \in \mathbb{R}$ mit $m_X(u) < \infty$

$$m_X(u) = \mathbb{E}[e^{uX}] = \psi_X(e^u)$$

Beweis. Klar, da folgt aus Definition VIII.1. □

Satz VIII.4
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen mit mgfs m_X, m_Y. Es gelten:

1. $m_X(0) = 1$
2. $a, b \in \mathbb{R}$

$$m_{aX+b}(u) = e^{bu} m_X(au) \quad \text{für } u \text{ so dass } m_X(au) < \infty$$

3. $X \perp\!\!\!\perp Y \implies m_{X+Y}(u) = m_X(u) m_Y(u)\ \forall u$, so dass $m_X(u), m_Y(u) < \infty$

Beweis. 1. Klar!

2. Sei $m_{aX+b}(u) = \mathbb{E}[e^{aXu+bu}] = e^{bu}\mathbb{E}[e^{auX}] = e^{bu} m_X(au)$.
3.

$$\begin{aligned} m_{X+Y}(u) &= \mathbb{E}[e^{Xu+Yu}] \overset{X \perp\!\!\!\perp Y}{=} \mathbb{E}[e^{uX}]\mathbb{E}[e^{uY}] \\ &= m_X(u) m_Y(u) \end{aligned}$$ □

Der Bezeichnung "momenterzeugend" erklärt sich mit der folgenden Proposition:

Satz VIII.5
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}$ Zufallsvariable mit momenterzeugender Funktion m_X, so dass ein $\varepsilon > 0$ existiert mit $m_X(u) < \infty$ auf $[0, \varepsilon)$. Dann gilt

$$\mathbb{E}[X^n] = \frac{\mathrm{d}^n}{\mathrm{d}u^n} m_X(0) \quad \forall n \in \mathbb{N}$$

Beweis. Für alle $u \in \mathbb{R}$ mit $m_X(u) < \infty$ folgt

$$\begin{aligned} m_X(u) = \mathbb{E}[e^{uX}] &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{(uX)^k}{k!}\right] \\ &\overset{\text{Lebesgue}}{=} \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k] u^k}{k!} \end{aligned}$$

n-fachen Differenzieren folgt

$$\frac{\mathrm{d}^n}{\mathrm{d}u^n} m_X(0) = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} k(k-1)\cdots(k-n+1) u^{k-n}$$

so dass

$$\frac{\mathrm{d}^n}{\mathrm{d}u^n} m_X(0) = \frac{\mathbb{E}[X^n]}{n!} n(n-1)\cdots(n-n+1) = \mathbb{E}[X^n].$$ □

Die mgf charakterisiert eine Verteilung eindeutig:

Satz VIII.6
Seien $(\Omega, \mathscr{F}, \mathbb{R}), (\Omega', \mathscr{F}', \mathbb{P}')$ Wahrscheinlichkeitsräume, $X : \Omega \to \mathbb{R}, Y : \Omega' \to \mathbb{R}$ Zufallsvariablen mit mgfs m_X, m_Y. Wenn $m_X(u), m_Y(u)$ in einer Umgebung um Null definiert sind und im Definitionsbereich gilt $m_X(u) = m_Y(u)$, so haben X und Y die selben Verteilungen ($X \overset{\mathrm{d}}{=} Y$, d = distribution)

Beweis. Sind X, Y $\mathbb{N}_0$-wertig, so folgt dies aus Lemma VIII.3 und dementsprechenden Resultat für pgfs. Der allgemeine Fall folgt aus dem Resultat zu charakteristischen Funktionen (Satz VIII.12). □

■ **Beispiel VIII.7 ((vgl. Lemma IV.3))**
Seien $X \sim \mathrm{Gamma}(\lambda, r), Y \sim \mathrm{Gamma}(\lambda, s)$ unabhängig, dann gilt nach Beispiel VIII.2

$$m_X(u) = \left(\frac{\lambda}{\lambda - u}\right)^r \quad u < \lambda \text{ und } m_Y(u) = \left(\frac{\lambda}{\lambda - u}\right)^s \quad u < \lambda$$

und nach Satz VIII.4

$$m_{X+Y}(u) = m_X(u) m_Y(u) = \left(\frac{\lambda}{\lambda - u}\right)^r \left(\frac{\lambda}{\lambda - u}\right)^s = \left(\frac{\lambda}{\lambda - u}\right)^{r+s}$$

Dies ist die mgf einer $\mathrm{Gamma}(\lambda, r+s)$ Verteilung. Nach Satz VIII.6 folgt $X+Y \sim \mathrm{Gamma}(\lambda, r+s)$.

Anmerkung
- pgf: $\psi_X(u) = \mathbb{E}[u^X]$
- mgf: $m_X(u) = \mathbb{E}[e^{uX}]$

Definition VIII.8
1. Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}^d$ Zufallsvariable. Dann ist
$$\varphi_X(u) := \mathbb{E}[e^{\mathrm{i}\langle u, X\rangle}] \quad u \in \mathbb{R}^d$$
charakteristische Funktion von X.
2. Ist $\mathbb{P}$ Verteilung in $\mathbb{R}^d$, dann ist
$$\varphi_{\mathbb{P}}(u) = \int_{\mathbb{R}^d} e^{\mathrm{i}\langle u, X\rangle} \mathbb{P}(\mathrm{d}x) \quad u \in \mathbb{R}^d$$
charakteristische Funktion von $\mathbb{P}$.

▶ **Bemerkung**
- Da $\left|e^{\mathrm{i}\langle u, X\rangle}\right| = 1$, ist die charakteristische Funktion für alle $u \in \mathbb{R}^d$ definiert.
- Die charakteristische Funktion ist die inverse FOURIERtransformation des Maßes $\mathbb{P}$ bzw. $\mathbb{P}_X$. Die mgf ist hingegen mit der LAPLACEtransformation verwandt.
- Existiert die mgf in einer Umgebung der Null, dann ist sie dort holomorph (komplex diffbar) und kann daher in die komplexe Ebene fortgesetzt werden (siehe ↗ Funktionentheorie). Dies

führt auf die charakteristische Funktion.

- Ist X eine $\mathbb{N}_0$-wertige Funktion, so gilt

$$\varphi_X(u) = \psi_X(e^{\mathrm{i}u})$$

■ **Beispiel VIII.9**

Sei $X \sim \text{Gamma}(\lambda, r)$. Dann folgt mittels Rechnung wie in Beispiel VIII.2 und dem Integralsatz von Cauchy:

$$\varphi_X(u) = \left(\frac{\lambda}{\lambda - \mathrm{i}u}\right)^r \quad u \in \mathbb{R}$$

Satz VIII.10 (Rechenregeln)

Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X, Y : \Omega \to \mathbb{R}^d$ mit charakteristischer Funktion φ_X, φ_Y. Dann

1. $\varphi_X(0) = 1$
2. Seien $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$. Dann gilt
$$\varphi_{AX+b}(u) = e^{\mathrm{i}\langle u,b\rangle}\varphi_X(A^T u)$$
3. Wenn $X \perp\!\!\!\perp Y$ dann folgt $\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u) \quad u \in \mathbb{R}^d$

Beweis. Analog zu Satz VIII.4. □

Satz VIII.11

Sei $X \sim \mathscr{N}(\mu, \sigma^2)$, dann

$$\varphi_X(u) = e^{\mathrm{i}\mu u - \frac{\sigma^2 u^2}{2}} \quad u \in \mathbb{R}$$

und insbesondere

$$\varphi_{\mathscr{N}(0,1)}(u) = e^{-\frac{u^2}{2}} \quad u \in \mathbb{R}$$

Beweis. Betrachte die Standardnormalverteilung:

$$\varphi(u) := \varphi_{\mathscr{N}(0,1)}(u) = \int_{\mathbb{R}} e^{\mathrm{i}ux} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x$$

Mit dem Differenzierbarkeitslemma für Parameterintergrale (↗ Schilling MINT, Satz 12.2)

$$\begin{aligned}
\varphi'(u) &= \int_{\mathbb{R}} \left(\frac{\mathrm{d}}{\mathrm{d}u} e^{\mathrm{i}ux}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{i}x e^{\mathrm{i}ux} e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} (-\mathrm{i}) e^{\mathrm{i}ux} \left(\frac{\mathrm{d}}{\mathrm{d}x} e^{-\frac{x^2}{2}}\right) \mathrm{d}x \\
&\overset{\text{P.I.}}{=} \underbrace{\frac{-\mathrm{i}}{\sqrt{2\pi}} \left[e^{\mathrm{i}ux} e^{-\frac{x^2}{2}}\right]_{x=-\infty}^{\infty}}_{=0} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{i} \left(\frac{\mathrm{d}}{\mathrm{d}x} e^{\mathrm{i}ux}\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{i} \cdot \mathrm{i}u e^{\mathrm{i}u} e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= -u \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\mathrm{i}ux} e^{-\frac{x^2}{2}} \,\mathrm{d}x \\
&= -u\varphi(u)
\end{aligned}$$

Die DGL $\varphi'(u) = -u\varphi(u)$ besitzt die Lösung

$$\varphi(u) = \varphi(0) e^{-\frac{x^2}{2}}$$

mit $\varphi(0) = 1$ nach Satz VIII.10, also folgt $\varphi(u) = e^{-\frac{u^2}{2}}$. Für $X \sim \mathscr{N}(\mu, \sigma^2)$ gilt $Z := \frac{X-\mu}{\sqrt{\sigma^2}} \sim \mathscr{N}(0,1)$ (↗ HA 9.1) und nach Satz VIII.10 folgt

$$\varphi_X(u) = \varphi_{\sigma Z+\mu}(u) = e^{\mathrm{i}\mu u} \varphi_Z(\sigma u) = e^{\mathrm{i}\mu u} e^{-\frac{\sigma^2 u^2}{2}}. \qquad \square$$

Die Charakteristische Funktion charakterisiert eine Verteilung eindeutig:

Satz VIII.12
Seien $(\Omega, \mathscr{F}, \mathbb{P}), (\Omega', \mathscr{F}', \mathbb{P}')$ Wahrscheinlichkeitsräume und $X : \Omega \to \mathbb{R}^d, Y : \Omega' \to \mathbb{R}^d$ Zufallsvariablen mit charakteristischen Funktionen φ_X, φ_Y. Dann:

$$X \overset{\mathrm{d}}{=} Y \iff \varphi_X(u) = \varphi_Y(u) \quad \forall u \in \mathbb{R}^d$$

Für den Beweis benötigen wir:

Lemma VIII.13
Seien $(\Omega, \mathscr{F}, \mathbb{P}), (\Omega', \mathscr{F}', \mathbb{P}')$ Wahrscheinlichkeitsräume, $X : \Omega \to \mathbb{R}^d$, $Y : \Omega' \to \mathbb{R}^d$ Zufallsvariablen. Dann gilt $X \overset{\mathrm{d}}{=} Y$ genau dann, wenn

$$\mathbb{E}[f(X)] = \mathbb{E}'[f(Y)] \quad \forall f \in C_c(\mathbb{R}^d)$$

(Wobei $\mathbb{E}'[f(Y)]$ Erwartungswert bzgl $\mathbb{P}'$ und C_c die Menge der stetigen Funktionen mit kompakten Träger sind)

Beweis (Lemma VIII.13). 1. $\Longrightarrow$: klar.

2. $\Longleftarrow$: Es genügt zu zeigen

$$\mathbb{E}[\mathbb{1}_K(X)] = \mathbb{P}(X \in K) = \mathbb{P}'(Y \in K) = \mathbb{E}'[\mathbb{1}_K(Y)] \tag{$*$}$$

für alle $K \subset \mathbb{R}^d$ kompakt, denn die kompakten Mengen sind ein $\cap$-stabiler Erzeuger von $\mathscr{B}(\mathbb{R}^d)$ und es gibt eine aufsteigende Folge $K_n \uparrow \mathbb{R}^d$. Die Indikatoren $\mathbb{1}_K$ können wir durch C_c-Funktionen approximieren. Sei

$$\begin{aligned}\operatorname{dist}(x, A) &:= \inf_{y\in A} |x-y| \quad A \subset \mathbb{R}^d \\ f_n(x) &:= \frac{x, U_n^c}{\operatorname{dist}(x, U_n^c) + \operatorname{dist}(x, K)} \quad U_n = \left\{ y \in \mathbb{R}^d \colon \operatorname{dist}(y, k) < \frac{1}{n} \right\}\end{aligned}$$

Dann ist $f_n \in C_c$ mit $f_n \downarrow \mathbb{1}_K$. Mit monotoner Konvergenz folgt ($*$) aus der Vorraussetzung. □

Beweis (Satz VIII.12).

($\Rightarrow$) klar

($\Leftarrow$) Sei $\varphi_X(u) = \varphi_Y(u)$, $u \in \mathbb{R}^d$. Wir konstruieren d unabhängige, identisch verteilte (u.i.v) Zufallsvariablen $N_1, \dots, N_d \sim \mathscr{N}(0,1)$ unabhängig von X. Dann ist auch $N := (N_1, \dots, N_d)$ unabhängig von X ($\nearrow$ Folgerung III.19). Zudem gilt

$$\begin{aligned}\mathbb{P}(\sqrt{\sigma^2} N \in \mathrm{d}y) &= \prod_{i=1}^d \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y_i^2}{2\sigma^2}} \,\mathrm{d}y_i \\ &= (2\pi\sigma^2)^{-\frac{d}{2}} e^{-\frac{|y|^2}{2\sigma^2}} \,\mathrm{d}y \quad \text{mit } |y| = \sqrt{\sum y_i^2}.\end{aligned}$$

Es gilt für alle $f \in C_c(\mathbb{R}^d)$:

$$\begin{aligned}\mathbb{E}[f(X + \sqrt{\sigma^2} N)] &\overset{X \perp\!\!\!\perp N}{=} \int_{\mathbb{R}^d} \mathbb{E}[f(X+y)] (2\pi\sigma^2)^{-\frac{d}{2}} e^{-\frac{|y|^2}{2\sigma^2}} \,\mathrm{d}y \\ &= (2\pi\sigma^2)^{-\frac{d}{2}} \int_{\mathbb{R}^d} f(z) \mathbb{E}[e^{-\frac{|X-z|^2}{2\sigma^2}}] \,\mathrm{d}z,\end{aligned}$$

wobei für $X \in \mathbb{R}^d$

$$\begin{aligned}e^{-\frac{|X-z|^2}{2\sigma^2}} &= \prod_{i=1}^d \underbrace{e^{-\frac{|X_i - z_i|^2}{\sigma^2}}}_{\varphi_{\mathscr{N}(0,\sigma^2)}(X_i - z_i)} \\ &= \prod_{i=1}^d \int_{\mathbb{R}} e^{\mathrm{i}(X_i - z_i) y_i} \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{\sigma^2 y_i^2}{2}} \,\mathrm{d}y_i \\ &= \frac{\sigma^d}{(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{\mathrm{i}\langle X - z, y\rangle} e^{-\frac{\sigma^2 |y|^2}{2}} \,\mathrm{d}y,\end{aligned}$$

so dass

$$\begin{aligned}\mathbb{E}[f(X + \sqrt{\sigma^2} N)] &= (2\pi)^{-d} \int_{\mathbb{R}^d} f(z) \mathbb{E}[\int_{\mathbb{R}} e^{\mathrm{i}\langle X - z, y\rangle} e^{-\frac{\sigma^2|y|^2}{2}} \,\mathrm{d}y] \,\mathrm{d}z \\ &= (2\pi)^{-d} \int_{\mathbb{R}^d} f(z) \int_{\mathbb{R}^d} \underbrace{\mathbb{E}[e^{\mathrm{i}\langle X, Y\rangle}]}_{\varphi_X(y)} e^{-\mathrm{i}\langle z, y\rangle} e^{-\frac{\sigma^2|y|^2}{2}} \,\mathrm{d}y \,\mathrm{d}z \quad \text{dom. Konv.}\end{aligned}$$

Dieselbe Rechnung auf $(\Omega', \mathscr{F}', \mathbb{P}')$ mit einem weiteren Vektor N' u.i.v $\mathscr{N}(0,1)$ Zufallsvariablen zeigt dann wegen $\varphi_X(u) = \varphi_Y(u)$.

$$\mathbb{E}[f(X + \sqrt{\sigma^2} N)] = \mathbb{E}'[f(Y + \sqrt{\sigma^2} N')]$$

Mit dominierter Konvergenz folgt für $\sigma^2 \to 0$

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y)] \quad \forall f \in C_c(\mathbb{R}^d)$$

Mit Lemma VIII.13 folgt die Behauptung. $\square$

Folgende Proposition is analog zu Satz VIII.5.

Satz VIII.14
$(\mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X : \Omega \to \mathbb{R}^d$ Zufallsvariable mit charakteristischer Funktion φ_X. Wenn $\mathbb{E}[\|X\|^n] < \infty$ für ein $n \in \mathbb{N}$, dann existiert

$$\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial^{\alpha_1} u_1 \cdots \partial^{\alpha_d}} \varphi_X(u)$$
$$\forall \alpha \in \mathbb{N}_0^d \text{ mit } |\alpha| = \sum_{i=1}^d \alpha_i \leq n$$

Zudem ist $\partial_u^\alpha \varphi_X$ stetig und es gilt

$$\mathbb{E}[X^\alpha] = i^{|-\alpha|} \partial_u^\alpha \varphi_X(0)$$

Beweis. Da $\|x^\alpha\| \leq \|x\|^{|\alpha|}$ für $x \in \mathbb{R}^d, \alpha \in \mathbb{N}_0^d$ folgt aus der Annahme $\mathbb{E}[\|X^\alpha\|] < \infty$. Damit

$$\begin{aligned}
\partial_u^\alpha \varphi_X(u) &= \int_{\mathbb{R}^d} e^{\mathrm{i}\langle u,x\rangle} \mathbb{P}(X \in \mathrm{d}x) \\
&\overset{(\star)}{=} \partial_u^\alpha \int_{\mathbb{R}^d} \partial_u^\alpha e^{\mathrm{i}\langle u,x\rangle} \mathbb{P}(X \in \mathrm{d}x) \\
&= \int_{\mathbb{R}^d} (\mathrm{i}x)^\alpha e^{\mathrm{i}\langle u,x\rangle} \mathbb{P}(X \in \mathrm{d}x) \\
&= \mathrm{i}^{|\alpha|} \int_{\mathbb{R}^d} e^{\mathrm{i}\langle u,x\rangle} \mathbb{P}(X \in \mathrm{d}x)
\end{aligned}$$

so dass die Ableitung existiert und für $u = 0$ die Momentenformel folgt. In $(\star)$ haben wir $|\alpha|$-mal das Differenzierbarkeitslemma ($\nearrow$ MINT Satz 12.1) verwendet, wobei $\partial_u^\alpha e^{\mathrm{i}\langle u,x\rangle}$ intbar ist, da

$$\left\|\partial_u^\alpha e^{\mathrm{i}\langle u,x\rangle}\right\| = \left\|(\mathrm{i}x)^\alpha e^{\mathrm{i}\langle u,x\rangle}\right\| = \|(\mathrm{i}x)^\alpha\| = \|x^\alpha\|. \qquad \square$$

Kapitel IX

Konvergenzbegriffe und Gesetze der großen Zahlen

Ziel: Für $X_1, X_2, \dots$ Zufallsvariablen (u.i.v) betrachten das Langzeitmittel $Y := \frac{1}{n}\sum_{i=1}^n X_i$ für $n \to \infty$.

Wissen bereits: Für $X_i \sim \text{Bernoulli}(p), p \in (0,1)$ unabhängig

- Für $np \to \lambda, n \to \infty$:

$$\mathbb{P}(nY_n = k) \xrightarrow{n\to\infty} e^{-\lambda}\frac{\lambda^k}{k!} = \text{Poisson}(\{k\})G$$

  (Poisson-Approx. Satz II.7).

- Für $c > 0$ gilt mit $x(k) := \frac{k-np}{\sqrt{np(1-p)}}$

$$\lim_{n\to\infty} \max_{k:|x_n(k)|\le c} \left| \frac{\sqrt{np(1-p)}\mathbb{P}(nY_n = k)}{g(x_n(k))} - 1 \right| = 0$$

  (De Moivre-Laplace Satz VII.2).

1. Schwaches Gesetz der großen Zahlen

(WLLN- Weak law of large numbers)
Sei $X_1, X_2, \dots$ u.i.v. Bernoulli($\frac{1}{2}$) Zufallsvariablen (z.B. Münzwurf). Wir erwarten, etwa bei der Hälfte der Zufallsvariablen eine Null/Eins (Kopf/Zahl) zu sehen, also

$$Y_n = \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{n\to\infty} \frac{1}{2} = \mathbb{E}[X_n] \quad \text{im geeigneten Sinne.}$$

Tatsächlich folgt mit der Stirling-Formel (Lemma VII.3)

$$\begin{aligned}
\mathbb{P}(Y_{2n} = \frac{1}{2}) &= \mathbb{P}(\text{Bin}(2n,p) = \frac{2n}{2}) \\
&= \binom{2n}{n} 2^{-2n} \\
&\overset{n\to\infty}{\sim} \frac{\sqrt{2\pi 2n}(2n)^{2n}e^{-2n}}{(\sqrt{2\pi n}n^n e^{-n})^2} 2^{-2n} \\
&= \frac{1}{\sqrt{\pi n}} \\
&\xrightarrow{n\to\infty} 0
\end{aligned}$$

Wir benötigen also einen geeigneten Grenzwertbegriff!

Definition IX.1 (stochastische Konvergenz)
Seien $Y, Y_1, Y_2, \ldots$ reelle Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$. Falls für alle $\varepsilon > 0$ gilt

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - Y| \le \varepsilon) = 1$$

so konvergiert $(Y_n)_{n\in\mathbb{N}}$ stochastisch oder in Wahrscheinlichkeit gegen Y.
Schreibweise: $Y_n \xrightarrow[\mathbb{P}]{n\to\infty} Y$ oder $\mathbb{P} - \lim_{n\to\infty} Y_n = Y$. (oder auch $\operatorname{plim}_{n\to\infty} Y_n = Y$).

▶ **Bemerkung**

Für Zufallsvariablen in $\mathbb{R}^d$ gilt eine entsprechende Definition mit euklidischer Norm anstelle des Betrags.

Lemma IX.2
Seien $Y, Z, Y_1, Y_2, \ldots$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$. Gelten

$$Y_n \xrightarrow[\mathbb{P}]{n\to\infty} Y \text{ und } Y_n \xrightarrow[\mathbb{P}]{n\to\infty} Z$$

so folgt

$$Y = Z \quad \mathbb{P} - \text{f.s.}$$

Beweis. Δ-Ungleichung: $|Y - Z| \le |Y - Y_n| + |Y_n - Z|$. Damit

$$\{|Y - Z| > 2\varepsilon\} \subseteq \{|Y - Y_n| > \varepsilon\} \cup \{|Y_n - Z| > \varepsilon\}$$

so dass

$$\mathbb{P}(|Y - Z| > 2\varepsilon) \le \mathbb{P}(|Y - Y_n| > \varepsilon) + \mathbb{P}(|Y_n - Z| > \varepsilon) \xrightarrow{n\to\infty} 0$$

Also

$$\begin{aligned}\mathbb{P}(Y \ne Z) = \mathbb{P}(|Y - Z| > 0) &= \mathbb{P}\Big(\bigcup_{k=1}^{\infty} \big\{|Y - Z| > k^{-1}\big\}\Big)\\ &= \sum_{k=1}^{\infty} \mathbb{P}(|Y - Z| > k^{-1}) = 0 \implies Y = Z \quad \mathbb{P} - \text{f.s.}\end{aligned}$$ □

Lemma IX.3
Seien $Y_1, Y_2, \dots, Z_1, Z_2, \dots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ und $(a_n)_{n\in\mathbb{N}}$ reellwertige Folge. Es gelten:

1. $Y_n \xrightarrow[\mathbb{P}]{n\to\infty} 0$ und $Z_n \xrightarrow[\mathbb{P}]{n\to\infty} 0$, dann $Y_n + Z_n \xrightarrow[\mathbb{P}]{n\to\infty} 0$.
2. $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$ und $(a_n)_{n\in\mathbb{N}}$ beschränkt, dann

$$a_n Y_n \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Beweis. Für $\varepsilon > 0$ beliebig gilt

$$\mathbb{P}(|Y_n - Z_n| > \varepsilon) \leq \mathbb{P}(|Y_n| > \frac{\varepsilon}{2}) + \mathbb{P}(|Z_n| > \frac{\varepsilon}{2}) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

und andererseits, wenn $|a_n| \leq A$

$$\mathbb{P}(|a_n Y_n| > \varepsilon) \leq \mathbb{P}(|Y_n| > \frac{\varepsilon}{A}) \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$ □

Satz IX.4 (WLLN, $\mathscr{L}^2$-Version)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X_1, X_2, \dots$ paarweise unkorrelierte, reelle Zufallsvariablen auf Ω in $\mathscr{L}^2(\mathbb{P})$, so dass

$$v := \sup_{i\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_i) < \infty$$

Dann gilt für alle $\varepsilon > 0$.

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

also insbesondere

$$1/n \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i]) \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

Falls $\mathbb{E}[X_i] = \mathbb{E}[X_1] \quad \forall i$, so gilt

$$1/n \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1]$$

Beweis. Sei $Y_n := 1/n \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])$, dann ist $Y_n \in \mathscr{L}^2(\mathbb{P})$ mit $\mathbb{E}[Y_n] = 0$ und

$$\mathbb{V}\mathrm{ar}(Y_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(X_i) \leq \frac{V}{n} \quad \text{Bienamé}$$

und die Aussage folgt aus der Ungleichung von Tschebyscheff (Satz V.13). □

Satz IX.5 (WLLN, $\mathscr{L}^1$-Version)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum, $X_1, X_2, \ldots$ paarweise unabhängige, identisch verteilte reelle Zufallsvariablen auf Ω in $\mathscr{L}^1(\mathbb{P})$. Dann gilt

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1].$$

Beweis. Wir verwenden ein Abschneideargument:
Definiere

$$X_i^\flat := X_i \mathbb{1}_{\{|X_i|\le i^{1/4}\}} \text{ und } X_i^\sharp := X_i \mathbb{1}_{\{|X_i|> i^{1/4}\}}$$

sowie

$$Y_n^\flat := \frac{1}{n}\sum_{i=1}^{n}(X_i^\flat - \mathbb{E}[X_i^\flat]) \text{ und } Y_n^\sharp := \frac{1}{n}\sum_{i=1}^{n}(X_i^\sharp - \mathbb{E}[X_i^\sharp])$$

Wir zeigen $Y_n^\flat \xrightarrow[n\to\infty]{\mathbb{P}} 0$ und $Y_n^\sharp \xrightarrow[n\to\infty]{\mathbb{P}} 0$. Mit Lemma IX.3 folgt dann die Behauptung.
Zu $Y_n^\flat$: Nach Folgerung III.19 sind auch $X_i^\flat$ paarweise unabhängig. Es folgt

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(Y_n^\flat) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}(X_i^\flat) \quad (\text{Bienaymé}) \\
&\le \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}[\underbrace{(X_i^\flat)^2}_{\le i^{1/2}}] \\
&\le \frac{1}{n^2} n n^{1/2} = n^{-1/2}
\end{aligned}$$

Mit Tschebyscheff (Lemma V.6) folgt $Y_n^\flat \xrightarrow[n\to\infty]{\mathbb{P}} 0$.
Zu $Y_n^\sharp$: Es gilt

$$\begin{aligned}
\mathbb{E}[X_i^\sharp] = \mathbb{E}[X_i \mathbb{1}_{\{|X_i|\le i^{1/4}\}}] &\overset{\text{i.v.}}{=} \mathbb{E}[X_i \mathbb{1}_{\{|X_i|\le i^{1/4}\}}] \\
&= \mathbb{E}[X_1] - \mathbb{E}[X_1 \mathbb{1}_{\{|X_i|\le i^{1/4}\}}] \xrightarrow[n\to\infty]{} 0.
\end{aligned}$$

wegen monotoner Konvergenz. Also

$$\mathbb{E}[|Y_n^\sharp|] \le \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^\sharp] \xrightarrow[n\to\infty]{} 0.$$

Mit Markov-Ungleichung (Lemma V.6) folgt für alle $\varepsilon > 0$

$$\mathbb{P}(|Y_n^\sharp| \ge \varepsilon) \le \frac{\mathbb{E}[|Y_n^\sharp|]}{\varepsilon} \xrightarrow[n\to\infty]{} 0$$

also $Y_n^\sharp \xrightarrow[n\to\infty]{\mathbb{P}} 0$. □

2. Das starke Gesetz der großen Zahlen

(SLLN - Strong law of large numbers)

Definition IX.6 (<u>$\mathbb{P}$ f.s. Konvergenz</u>)
Seien $Y, Y_1, Y_2, \dots$ reellen Zufallsvariablen auf Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$. Falls

$$\mathbb{P}(\Big\{\omega\colon \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\Big\}) = 1$$

so konvergiert $(Y_n)_{n\in\mathbb{N}}$ <u>$\mathbb{P}$-fast sicher</u> gegen Y.
Schreibweise: $Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Y$ oder $\lim_{n\to\infty} Y_n = Y$ f.s.

▶ **Bemerkung**

- Diese Konvergenzart wurde bereits verwendet, z.B. in MINT, Konvergenzsatz von Lebesgue.
- Auch bei der f.s. Konvergenz ist der Grenzwert f.s. eindeutig bestimmt:
  $Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Y$ und $Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Z \implies \mathbb{P}(Y = Z) = \mathbb{P}(\{\lim Y_n = Y\} \cap \{\lim Y_n = Z\}) = 1$
- Erweiterung auf Zufallsvariablen in $\mathbb{R}^d$ ist offensichtlich.

Lemma IX.7
$Y, Y_1, Y_2 \dots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$. Dann

$$Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Y \implies Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y$$

Beweis. Für $\varepsilon > 0$ gilt

$$\mathbb{P}(|Y_n - Y| > \varepsilon) = \mathbb{E}[\underbrace{\mathbb{1}_{\{|Y_n - Y| > \varepsilon\}}}_{=:Z_n}]$$

Die Zufallsvariablen Z_n sind gleichmäßig durch $Z \equiv 1 \in \mathscr{L}^1(\mathbb{P})$ beschränkt und $Z_n \xrightarrow[n\to\infty]{\text{f.s.}} 0$. Nach dem Konvergenzsatz von Lebesgue folgt

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - Y| > \varepsilon) = \mathbb{E}[\lim Z_n] = 0.$$ □

▶ **Bemerkung**

Die Umkehrung gilt i. A. nicht: Definiere eine Folge von Zufallsvariablen auf $([0,1], \mathscr{B}([0,1]), \mathrm{U}([0,1])$ durch

$$Y_k := \mathbb{1}_{[m2^{-n},(m+1)2^{-n}]} \qquad \text{falls } k = 2^n + m \text{ mit } 0 \le m < 2^n$$

Dann gilt für alle $0 < \varepsilon < 1$

$$\mathbb{P}(|Y_k| > \varepsilon) = 2^{-n} \xrightarrow[n\to\infty]{} 0$$

Also

$$Y_n \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Allerdings $\mathbb{P}(\{\omega\colon \lim Y_n(\omega) = 0\}) = 0$, also $Y_n \not\xrightarrow[n\to\infty]{\text{f.s.}} 0$

▶ **Erinnerung**

$$\begin{aligned}\omega \in \limsup_{n\to\infty} A_n &= \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n \\ \iff \omega &\text{ ist in } \infty \text{ vielen der } A_n \text{ enthalten.}\end{aligned}$$

Satz IX.8 (Borel-Cantelli-Lemma)
Sei $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $(A_n)_{n\in\mathbb{N}}$ Folge in $\mathscr{F}$.

1. Falls $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, so folgt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$.
2. Falls $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ und die A_n paarweise unabhängig sind, so folgt $\mathbb{P}(\limsup_{n\to\infty} A_n) = 1$

▶ **Bemerkung**
Das BC-Lemma ist ein Null-Eins-Gesetz. (Gibt noch viel mehr davon! z.B. in Random Graphs, siehe Talk Logic and Random Graphs.)

Beweis. Es gilt

$$\begin{aligned}\omega \limsup_{n\to\infty} A_n &\iff \omega \text{ ist in } \infty \text{ vielen } A_n \text{ enthalten} \\ &\iff \sum_{n=1}^{\infty} \mathbb{1}_{A_n}(\omega) = 0 \end{aligned} \tag{$\star$}$$

1. $$\mathbb{E}[\sum_{n=1}^{\infty} \mathbb{1}_{A_n}] \overset{\text{BL}}{=} \sum_{n=1} \mathbb{E}[\mathbb{1}_{A_n}] = \sum_{n=1} \mathbb{P}(A_n) < \infty$$

   Die zeigt $\sum_{n=1} \mathbb{1}_{A_n} < \infty$ $\mathbb{P}$-f.s. und mit ($\star$) $\mathbb{P}(\limsup A_n) = 0$
2. Setze $S_n = \sum_{i=1}^{n} \mathbb{1}_{A_i}$ und $S = \lim_{n\to\infty} S_n = \sum_{i=1}^{\infty} \mathbb{1}_{A_i}$. Wegen paarweiser Unabhängigkeit gilt über Biennaymé

   $$\begin{aligned}\mathbb{V}\mathrm{ar}(S_n) = \mathbb{V}\mathrm{ar}(\sum_{i=1}^{n} \mathbb{1}_{A_i}) &= \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}(\mathbb{1}_{A_i}) \\ &\leq \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_{A_i}^2] = \sum_{i=1}^{n} \mathbb{E}[\mathbb{1}_{A_i}] = \sum_{i=1}^{n} \mathbb{P}(A_i) \\ &=: m_n \xrightarrow[n\to\infty]{} \infty\end{aligned}$$

Da $S_n \leq S$ folgt $\{S \leq 1/2 m_n\} \subseteq \{S_n \leq 1/2 m_n\}$ so dass

$$\begin{aligned}\mathbb{P}(S \leq 1/2 m_n) \leq \mathbb{P}(S_n \leq 1/2 m_n) &= \mathbb{P}(S_n - m_n \leq -1/2 m_n) \quad m_n = \mathbb{E}[S_n] \\ &\leq \mathbb{P}(|S_n - m_n| \geq 1/2 m_n)\end{aligned}$$

und mit Tschebyscheff:

$$\begin{aligned}\mathbb{P}(S \leq 1/2 m_n) &\leq \mathbb{P}(|S_n - m_n| \geq 1/2 m_n) \\ &\leq \frac{\mathbb{V}\mathrm{ar}(S_n)}{m_n^2/4} \\ &\leq \frac{4}{m_n} \xrightarrow[n\to\infty]{} 0\end{aligned}$$

Damit folgt mit monotoner Konvergenz

$$\mathbb{P}(S < \infty) = \mathbb{P}(\bigcup_{n=1}^{\infty} \{S \leq 1/2 m_n\}) = \lim_{n\to\infty} \mathbb{P}(S \leq 1/2 m_n) = 0$$

und daraus folgt dann $\mathbb{P}(\limsup A_n) = 1$ mit ($\star$). □

Satz IX.9 (SLLN, $\mathscr{L}^2$-Version)
$(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum und $X_1, X_2, \dots$ paarweise unkorrelierte (reelle) Zufallsvariablen auf Ω in $\mathscr{L}^2(\mathbb{P})$, so dass $v := \sup_{i\in\mathbb{N}} \mathbb{V}\mathrm{ar}(X_i) < \infty$. Dann gilt

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \xrightarrow[n\to\infty]{\text{f.s.}} 0.$$

Beweis. Nehme oBdA an, dass $\mathbb{E}[X_i] = 0 \quad \forall i$, sonst betrachte

$$X_i' = X_i = \mathbb{E}[X_i]$$

Setze $Y_n = 1/n \sum_{i=1}^n X_i$.
Wir zeigen zunächst $Y_n \xrightarrow[n\to\infty]{\text{f.s.}} 0$, bevor wir zu Y_n übergehen. Aus Satz IX.4 (WLLN, $\mathscr{L}^2$) ist bekannt:

$$\forall \varepsilon > 0 \quad \mathbb{P}(|Y_{n^2}| \geq \varepsilon) < \infty.$$

Insbesondere folgt

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_{n^2}| \geq \varepsilon) < \infty.$$

Aus Borel-Cantelli folgt

$$\mathbb{P}(\limsup_{n\to\infty} |Y_{n^2}| > \varepsilon) \leq \mathbb{P}(\limsup_{n\to\infty} \{|Y_{n^2}| \geq \varepsilon\}) = 0$$

und für $\varepsilon \to 0$ folgt

$$\begin{aligned}\mathbb{P}(\lim_{n\to\infty} Y_{n^2} \neq 0) &= \mathbb{P}(\limsup_{n\to\infty} |Y_{n^2}| > 0) \\ &= \lim_{n\to\infty} \mathbb{P}(\limsup_{n\to\infty} |Y_{n^2}| > \varepsilon) = 0\end{aligned}$$

also gilt $Y_{n^2} \xrightarrow{\text{f.s.}} 0$. Wähle nun für jedes $m \in \mathbb{N}$ ein $n = n(m)$ so dass

$$n^2 \leq m < (n+1)^2$$

Wir "vergleichen" Y_m mit Y_{n^2}:

$$\begin{aligned}
\mathbb{P}(\left|mY_m - n^2 Y_{n^2}\right| > \varepsilon n^2) &= \mathbb{P}(\left|\sum_{i=n^2+1}^{m} X_i\right| > \varepsilon n^2) \\
&\leq \frac{1}{\varepsilon^2 n^2} \mathbb{V}\text{ar}(\sum_{i=n^2+1}^{m} X_i) \quad \text{Tschebyscheff} \\
&\leq \frac{v(m-n^2)}{\varepsilon^2 n^4} \quad \text{Bienaymé}
\end{aligned}$$

Es folgt

$$\begin{aligned}
\sum_{m=1}^{\infty} \mathbb{P}(\left|mY_m - n^2 Y_{n^2}\right| > \varepsilon n^2) &\leq \frac{v}{\varepsilon} \sum_{m} \frac{(m-n^2)}{n^4} \\
&= \frac{v}{\varepsilon} \sum_{m} \frac{(m-n(m))}{n(m)^4} \\
&= \frac{v}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{m=n^2}^{(n+1)^2-1} \frac{(m-n^2)}{n^4} \\
&= \frac{v}{\varepsilon^2} \sum_{n=1}^{\infty} \sum_{k=0}^{2n} \frac{k}{n^4} \\
&= \frac{v}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{(2n)(2n+1)}{2n^4} < \infty
\end{aligned}$$

und mit Borel-Cantelli:

$$\mathbb{P}(\lim_{m\to\infty} \left|\frac{mY_m}{n^2} - Y_{n^2}\right| = 0) = 1$$

Zusammen mit $Y_{n^2} \xrightarrow{\text{f.s.}} 0$ folgt

$$\mathbb{P}(\lim_{m\to\infty} \left|\frac{mY_m}{n^2}\right| = 0) = 1$$

und da

$$|Y_m| \leq \frac{m}{n^2} |Y_m|$$

impliziert das

$$\mathbb{P}(\lim_{m\to\infty} |Y_m| = 0) = 1. \qquad \square$$

▶ **Bemerkung**

Das SLLN gilt auch unter dem Bedingungen von Satz IX.5 (WLLN, $\mathscr{L}^1$). Der Beweis basiert auf einem Teilfolgen- sowie Abschneideargument. ($\nearrow$ z.B. Klenke, Schilling WT, ...)

"Schnelle" stochastische Konvergenz impliziert f.s. Konvergenz.

Lemma IX.10
$Y, Y_1, Y_2 \dots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ so dass $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y$ und so das für eine Nullfolge von $\varepsilon_n \downarrow 0$ gilt

$$\sum_{n=1}^{\infty} \mathbb{P}(|Y_n - Y| > \varepsilon_n) < \infty$$

Dann folgt

$$Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Y$$

Beweis. Nach Borel-Cantelli folgt

$$\mathbb{P}(|Y_n - Y| > \varepsilon_n \text{ für unendliche viele } n) = 0$$

Dies ist äquivalent zu: $\mathbb{P}(|Y_n - Y| > \varepsilon_n$ für höchstens endliche viele $n) = 1 \iff \exists \Omega' \colon \mathbb{P}(\Omega') = 1$, so dass

$$\forall \omega \in \Omega' \exists N(\omega') \quad \forall n \geq N(\omega') \colon \left|Y_n(\omega') - Y(\omega')\right| \leq \varepsilon_n$$

impliziert das

$$\exists \Omega' \subseteq \Omega, \mathbb{P}(\Omega') = 1, \text{ so dass } \left|Y_n(\omega') - Y(\omega')\right| \xrightarrow[n\to\infty]{} 0 \implies Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Y.$$ □

Folgerung IX.11
$Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$, so dass $Y_n \xrightarrow[n\to\infty]{\text{f.s.}} Y$. Dann existiert eine Teilfolge $(Y_{n_k})_{k\in\mathbb{N}} \subseteq (Y_n)_{n\in\mathbb{N}}$, so dass $Y_{n_k} \xrightarrow[k\to\infty]{\text{f.s.}} Y$.

Beweis. Nach Vorrausetzung gilt

$$\forall k \in \mathbb{N} \quad \forall \varepsilon > 0 \exists N(k,\varepsilon) \colon \forall n \geq N(k,\varepsilon) \colon \mathbb{P}(|Y_n - Y| > \varepsilon) \geq 2^{-k}$$

Wähle $\varepsilon_k = 2^{-k}$ und $n_k = N(k, 2^{-k})$, dann ist

$$\sum_{k=1}^{\infty} \mathbb{P}(|Y_{n_k} - Y| > \varepsilon_k) \leq \sum_{i=1}^{\infty} 2^{-k} < \infty$$

und die Behauptung folgt aus Lemma IX.10. □

3. Der Satz von Glivenko-Cantelli

Gegeben seien n Realisierungen $x_1 = X_1(\omega), x_2 = X_2(\omega), \dots x_n = X_n(\omega)$ von u.i.v. Zufallsvariablen $X_1, \dots, X_n$ in $\mathbb{R}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$. Die Verteilung F von X_i ist unbekannt.

Frage: Wie können wir $\mathbb{P}_{x_i}$ bzw. die zugehörige Verteilungsfunktion F bestimmen/ approximieren?

Definition IX.12 (empirische Verteilungsfunktion, Zufallsproben)
Seien $X_1, X_2, \ldots$ u.i.v. Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ in $\mathbb{R}$, dann definiert

$$\hat{F}_n(x, \omega) := \frac{\#\{i \colon 1 \leq i \leq n \mid X_i(\omega) \leq x\}}{n}$$

die empirische Verteilungsfunktion der Zufallsproben $(X_1, \ldots, X_n)$.

▶ **Bemerkung**

- Für alle $\omega \in \Omega$ ist $\hat{F}_n(x) : \mathbb{R} \to [0, 1]$ eine Verteilungsfunktion (↗ Übung)
- $\left\{\hat{F}_n, x \in \mathbb{R}\right\}$ ist eine Familie von Zufallsvariablen und damit ein stochasticher Prozess.

Satz IX.13 (Eigenschaften der empirischen Verteilungsfunktion)
Seien $X_1, X_2, \ldots$ u.i.v. Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ in $\mathbb{R}$ mit Verteilungsfunktion F. Dann gelten:

1. $n\hat{F}_n(x) \sim \text{Bin}(n, F(x)) \quad \forall x \in \mathbb{R}$
2. $\mathbb{E}[\hat{F}_n(x)] = F(x) \quad \forall x \in \mathbb{R}$
3. $\mathbb{V}\text{ar}(\hat{F}_n(x)) = \frac{F(x)(1-F(x))}{n} \quad \forall x \in \mathbb{R}$
4. $\hat{F}_n(x) \xrightarrow[n\to\infty]{\text{f.s.}} F(x) \quad x \in \mathbb{R}$

Beweis. 1. Es gilt

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}$$

mit

$$\mathbb{1}_{\{X_i \leq x\}} \sim \text{Bernoulli}(F(x))$$

Mit Unabhängigkeit der X_i folgt die Behauptung.

2. (2&3) folgt aus 1) $\forall x$

$$\mathbb{E}[n\hat{F}_n(x)] = nF(x)$$
$$\mathbb{V}\text{ar}(n\hat{F}_n(x))(1 - F(x))$$

3. Da die Zufallsvariablen $Y_i = \mathbb{1}_{\{X_i \leq x\}}$ für jedes x u.i.v. und in $\mathscr{L}^2$ sind, gilt nach dem SLLN

$$\begin{aligned} \hat{F}_(x) &= \frac{1}{n} Y_i \xrightarrow[n\to\infty]{\text{f.s.}} \mathbb{E}[Y_i] \\ &= \mathbb{E}[\mathbb{1}_{\{X_i \leq x\}}] = \mathbb{P}(X_i \leq x) = F(x). \end{aligned}$$

□

Satz IX.14 (Gilvenko-Cantelli)
$X_1, X_2, \ldots$ u.i.v. Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ in $\mathbb{R}$ mit Verteilungsfunktion F. Dann gilt

$$\mathbb{P}(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| = 0) = 1$$
$$\sup_{x\in\mathbb{R}} \left|\hat{F}_n(x) - F(x)\right| \xrightarrow[n\to\infty]{\text{f.s.}} 0.$$

Beweis. Nehme zunächst an, dass F stetig ist. Für jedes $m \in \mathbb{N}$ existieren dann

$$-\infty = z_0 < z_1 < \cdots < z_{m-1} < z_m = \infty$$

so dass

$$F(z_0) = 0 \quad F(z_1) = \frac{1}{m} \quad \ldots \quad F(z_k) = \frac{k}{m} \quad \ldots \quad F(z_n) = 1$$

Es folgt für jedes $z \in [z_n, z_{k+1})$

$$\hat{F}_n(z) - F(z) \leq \hat{F}_n(z_{k+1}) - F(z_k) = \hat{F}(z_{k+1}) - F(z_{k+1}) + \frac{1}{m}$$
$$\hat{F}_n(z) - F(z) \geq \hat{F}_n(z_n) - F(z_{k+1}) = \hat{F}_n(z_k) - F(z_k) - \frac{1}{m} \tag{$\star$}$$

Setze für $k \in \{0, 1, \ldots, m\}, m \in \mathbb{N}$. Dann

$$A_{m,k} := \left\{\omega \in \Omega \colon \lim_{n\to\infty} \hat{F}_n(z_k, \omega) = F(z_n)\right\}$$

dann gilt nach Satz IX.13 $\mathbb{P}(A_{m,k}) = 1 \quad \forall m, k$. Für $A_m := \bigcup_{k=0}^{m} A_{m,k}$ folgt

$$\mathbb{P}(A_m) = 1 - \mathbb{P}(A_m^C) = 1 - \mathbb{P}(\bigcup A_m^C) \geq 1 - \sum \mathbb{P}(A_{m,k}^C) = 1$$

Für jedes $\omega \in A_m$ gilbt es $N(\omega) \in \mathbb{N}$, so dass $\forall n \geq N(\omega)$ und $\forall k \in \{0, \ldots, m\}$

$$\left|\hat{F}_n(z_k, \omega) - F(z_k)\right| < \frac{1}{m}$$

Zusammen mit ($\star$) folgt $\forall n \geq N(\omega)$ und alle $\omega \in A_m$

$$\sup_{x\in\mathbb{R}} \left|\hat{F}_n(z) - F(z)\right| < \frac{2}{m} \tag{$\star\star$}$$

Für $A = \bigcap_{m\in\mathbb{N}} A_m$ folgt (wie bei A_m) auch $\mathbb{P}(A) = 1$ und aus ($\star\star$), dass für jedes $\omega \in A$ und jedes $\varepsilon > 0$ ein $N(\omega, \varepsilon) \in \mathbb{N}$ existiert, so dass $\forall n > N(\omega, \varepsilon)$ gilt

$$\sup_{z\in\mathbb{R}} \left|\hat{F}_n(z, \omega) - F(z)\right| < \varepsilon$$

Für $\varepsilon \to 0$ folgt die Behauptung.
Ist F nicht notwendigerweise stetig, so wähle für $m \in \mathbb{N}$

$$-\infty = z_0 < z_1 < \cdots < z_{m-1} < z_m = \infty$$

so dass

$$F(z_{k+1}^-) - F(z_k) = \lim_{y\uparrow\infty} F(y) - F(z_k) \leq \frac{1}{m}$$

Dann folgt analog zu ($\star$) für $z \in [z_n, z_{k+1})$

$$\hat{F}_n(z) - F(z) \leq \hat{F}(z_{k+1}^-) - F(z_{k+1}^-) + \frac{1}{m}$$
$$\hat{F}_n(z) - F(z) \geq \hat{F}_n(z_n) - F(z_k) - \frac{1}{m}$$

und damit

$$A'_{m,k} = \left\{\omega \in \Omega \colon \lim_{n\to\infty} \hat{F}_n(z_k^-, \omega) = F(z_k^-)\right\}$$

folgt der verbleibende Beweis analog zum stetigen Fall. □

4. $\mathscr{L}^p$-Konvergenz

Definition IX.15 ($\mathscr{L}^p$-Konvergenz)
Seien $Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}\, \mathbb{P})$. Falls $(Y, Y_n \in \mathscr{L}^p(\mathbb{P})), n \in \mathbb{N}$ für ein $p \in [1, \infty]$ und

$$\lim_{n\to\infty} \|Y - Y_n\|_p := \lim_{n\to\infty} (\mathbb{E}[Y_n - Y]^p)^{1/p} = 0$$

so konvergiert $(Y_n)_{n\in\mathbb{N}}$ in $\mathscr{L}^p$/ im p-ten Mittel gegen Y. Wir schreiben:

$$Y_n \xrightarrow[n\to\infty]{\mathscr{L}^p} Y \text{ oder } \mathscr{L}^p - \lim_{n\to\infty} Y_n = Y$$

▶ **Bemerkung**

- Der Grenzwert einer $\mathscr{L}^p$-konvergenten Folge ist f.s. eindeutig: Minkowski-Ungleichung (↗ Schilling MINT, Korollar 14.5) liefert für $1 \leq p \leq \infty$

$$\|Y - Z\|_p = \|Y - Y_n + Y_n - Z\|_p \leq \underbrace{\|Y - Y_n\|_p}_{\to 0} + \underbrace{\|Y_n - Z\|_p}_{\to 0}$$
$$\implies \|Y - Z\|_p = \implies Y = Z \quad \mathbb{P}\text{-f.s.}$$

- $\mathscr{L}^p$-Konvergenz in $\mathbb{R}^d$ lässt sich analog definieren.

Lemma IX.16
$Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$. Es gelten für $1 \leq p \leq \infty$

1. $Y_n \xrightarrow[n\to\infty]{\mathscr{L}^p} Y \implies Y_n \xrightarrow[n\to\infty]{\mathscr{L}^1} Y$
2. $Y_n \xrightarrow[n\to\infty]{\mathscr{L}^1} Y \implies Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y$

Beweis. 1. Für $1 \leq p \leq \infty$ setze $q = (1 - 1/p)^{-1}$ mit $q = 1$ für $p = \infty$ so dass $1/p + 1/q = 1$ und mit

Hölder

$$\begin{aligned}
\|Y_n - Z\|_1 &= \mathbb{E}[Y_n - Y] = \int |Y_n - Y| \cdot 1 \,\mathrm{d}\mathbb{P} \\
&= \left(\int |Y_n - Y|^p \right)^{1/p} \cdot \underbrace{\left(\int 1^q \,\mathrm{d}\mathbb{P} \right)^{1/q}}_{=1} \\
&= \left(\mathbb{E}[|Y_n - Y|^p]\right)^{1/p} \\
&= |Y_n - Y|_p \xrightarrow[n\to\infty]{0}
\end{aligned}$$

2. Mit der Markov-Ungleichung folgt $\forall \varepsilon > 0$

$$\mathbb{P}(|Y_n - Y| > \varepsilon) \leq \frac{\mathbb{E}[|Y_n - Y|]}{\varepsilon} \xrightarrow[n\to\infty]{} 0. \qquad \square$$

$$\begin{tikzcd}
\text{f.s. Konv.} \arrow[d, Rightarrow, "\uparrow \text{TF}"] & \mathscr{L}^p\text{-Konv.} \arrow[l, dotted, "\text{TF}"'] \arrow[d, Rightarrow] \\
\text{Stoch. Konv.} \arrow[d, Rightarrow] & \mathscr{L}^1\text{-Konv.} \arrow[l, Rightarrow] \\
\text{Verteilungskonv.} &
\end{tikzcd}$$

Kapitel X

Verteilungskonvergenz und der zentrale Grenzwertsatz

1. Die Verteilungskonvergenz

Definition X.1 (schwache Konvergenz)
$Y, Y_1, Y_2, \ldots$ reelle Zufallsvariablen ("jede ZV darf ihren eigenen WR mitbringen"). Falls für alle $f \in C_b(\mathbb{R})$ (stetig und beschränkt) gilt, dass

$$\lim_{n\to\infty} \mathbb{E}[f(Y_n)] = \mathbb{E}[f(Y)]$$

so konvergiert $(Y_n)_{n\in\mathbb{N}}$ schwach / in Verteilung gegen Y.
Schreibe: $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ oder $Y_n \Rightarrow Y, n \to \infty$ oder $\mathbb{P}_{Y_n} \Rightarrow \mathbb{P}_Y$.

▶ **Bemerkung**

- Formal sollten wir eigentlich schreiben

$$\lim \mathbb{E}_n[f(Y_n)] = \mathbb{E}[f(Y)]$$

  wobei $\mathbb{E}_n[f(Y_n)]$ bzgl. $\mathbb{P}_n, Y_n$ auf $(\Omega_n, \mathscr{F}_n, \mathbb{P}_n)$. Dies wird aber in der Regeln vernachlässigt.
- Für Zufallsvariablen in $\mathbb{R}^d$ lässt sich schwache Konvergenz mittels $f \in C_b(\mathbb{R}^d)$ (d.h. $f : \mathbb{R}^d \to \mathbb{R}$) definieren.

Der Grenzwert einer schwach konvergenten Folge ist eindeutig in Verteilung.

Lemma X.2
$Y, Z, Y_1, Y_2, \ldots$ reelle Zufallsvariablen, so dass

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y \text{ und } Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Z$$

Dann gilt: $Y \overset{\mathrm{d}}{=} Z$ bzw. $P_Y = \mathbb{P}_Z$.

Beweis. Betrachte ein fixes kompaktes Intervall $[a, b] \subset \mathbb{R}$. Da die kompakten Intervalle ein $\cap$-stabiler Erzeuger von $\mathscr{B}(\mathbb{R})$ sind, genügt es zu zeigen

$$\int \mathbb{1}_{[a,b]} \,\mathrm{d}\mathbb{P}_Y = \mathbb{P}_Y([a, b]) = \mathbb{P}_Z([a, b]) = \int \mathbb{1}_{[a,b]} \,\mathrm{d}\mathbb{P}_Z.$$

Dazu konstruiere eine Folge $(f_k)_{k\in\mathbb{N}}$ in $C_b(\mathbb{R})$, so dass $f_k \downarrow f = \mathbb{1}_{[a,b]}$ ($\nearrow$ Beweis zu Lemma VIII.13). Dann folgt mit monotoner Konvergenz

$$\int \mathbb{1}_{[a,b]} \, d\mathbb{P}_Y = \lim_{k\to\infty} \int f_k \, d\mathbb{P}_Y = \lim_{k\to\infty} \lim_{n\to\infty} \int f_k \, d\mathbb{P}_{Y_n}$$

und analog

$$\int \mathbb{1}_{[a,b]} \, d\mathbb{P}_Z = \lim_{k\to\infty} \lim_{n\to\infty} \int f_k \, d\mathbb{P}_{Y_n}$$

Das liefert die Behauptung. □

Satz X.3 (Portmanteau)
$Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen. Die folgenden Aussagen sind äquivalent:

1. $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$
2. $\lim_{n\to\infty} \mathbb{E}[f(Y_n)] = \mathbb{E}[f(Y)] \quad \forall f \in C_b^g(\mathbb{R})$ (glm stetig und beschränkt)
3. $\limsup_{n\to\infty} \mathbb{P}(Y_n \in F) \le \mathbb{P}(Y \in F) \quad \forall F \subset \mathbb{R}$ abgeschlossen
4. $\liminf_{n\to\infty} \mathbb{P}(Y_n \in O) \ge \mathbb{P}(Y \in O) \quad \forall O \subset \mathbb{R}$ offen
5. $\lim_{n\to\infty} \mathbb{P}(Y_n \in C) = \mathbb{P}(Y \in C) \quad C \in \mathscr{B}(\mathbb{R})$ mit $\mathbb{P}_Y(\partial C) = 0$ (Rand von C)

Beweis. 1. $\implies$ 2.: ist klar

2. $\implies$ 3.: Sei F abgeschlossen und definiere für $k \in \mathbb{N}$

$$f_k(x) = (1 - k \operatorname{dist}(x, F))^+ \text{ mit } \operatorname{dist}(x, F) = \inf_{y\in F} |x - y|$$

Dann ist f_k beschränkt und glm. stetig, denn

$$\begin{aligned} |f_k(y) - f_k(x)| &\le k|\operatorname{dist}(y, F) - \operatorname{dist}(x, F)| \\ &\le k|y - x| \quad \forall x, y \in \mathbb{R} \end{aligned}$$

da

$$\begin{aligned} \operatorname{dist}(x, F) = \inf_z |x - z| &\le \inf_z (|x - z| + |y - z|) \\ &= |x - y| + \operatorname{dist} y, F. \end{aligned}$$

Zudem gilt $f_k \le \mathbb{1}_F$ und $f_k \downarrow \mathbb{1}_F$, so dass

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}(Y_n \in F) &= \limsup_{n\to\infty} \mathbb{E}[\mathbb{1}_F(Y_n)] \\ &\le \lim_{n\to\infty} \mathbb{E}[f_k(Y_n)] \\ &\overset{2.}{=} \mathbb{E}[f_k(Y)] \end{aligned}$$

Mit monotoner Konvergenz folgt

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{P}(Y_n \in F) &\le \inf_{k\in\mathbb{N}} \mathbb{E}[f_k(Y)] \\ &= \mathbb{E}[\mathbb{1}_F(Y)] = \mathbb{P}(Y \in F) \end{aligned}$$

3. $\implies$ 4.: Für jedes $O \subset \mathbb{R}$ offen ist O^C abgeschlossen, so dass

$$\begin{aligned}\liminf_{n\to\infty} \mathbb{P}(Y_n \in O) &= \liminf_{n\to\infty}(1 - \mathbb{P}(Y_n \in O^C)) \\ &= 1 - \limsup_{n\to\infty} \mathbb{P}(Y_n \in O^C) \\ &\overset{3.}{\geq} 1 - \mathbb{P}(Y \in O^C) \\ &= \mathbb{P}(Y \in O).\end{aligned}$$

4. $\implies$ 3.: Analog und vertausche $\limsup$ mit $\liminf$.

4. und 3. $\implies$ 5.: Sei $C \in \mathscr{B}(\mathbb{R})$ und $\mathring{C}$ dass offene Innere von C, $\bar{C} = \mathring{C} \cup \partial C$ der Abschluss.

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(Y_n \in C) &\leq \limsup_{n\to\infty} \mathbb{P}(Y_n \in \bar{C}) \\ &\overset{3.}{\leq} \mathbb{P}(Y \in \bar{C}) \\ &= \mathbb{P}(Y \in \mathring{C}) \quad (\text{ da } \mathbb{P}_Y(\partial C) = 0) \\ &\overset{4.}{\leq} \liminf_{n\to\infty} \mathbb{P}(Y_n \in \mathring{C}) \\ &\leq \limsup_{n\to\infty} \mathbb{P}(Y_n \in C).\end{aligned}$$

5. $\implies$ 1.: Sei $f \in C_b(\mathbb{R})$ positiv. (wenn nicht positiv: in positiven und negativen Anteil zerlegen und dann mit Linearität arbeiten). Da $\partial\{f \geq t\} = \{f = t\}$ gilt, folgt $\mathbb{P}_Y(\partial\{f \geq t\}) > 0$ für höchstens abzählbar viele t und das impliziert

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}[f(Y_n)] &= \lim_{n\to\infty} \int_{\mathbb{R}} f \, \mathrm{d}\mathbb{P}_{Y_n} \\ &= \lim_{n\to\infty} \int_0^\infty \mathbb{P}_{Y_n}(f \geq t) \, \mathrm{d}t \quad \text{folgt mit Schilling MINT Satz 16.7} \\ &= \lim_{n\to\infty} \int_0^\infty \mathbb{P}(f(Y_n) \geq t) \, \mathrm{d}t \\ &= \int_0^\infty \lim_{n\to\infty} \mathbb{P}(f(Y_n) \geq t) \, \mathrm{d}t \quad \text{dom. Konvergenz} \\ &= \int_0^\infty \mathbb{P}(f(Y) \geq t) \, \mathrm{d}t \quad \text{nutze 5.} \\ &= \mathbb{E}[f(Y)] \quad \text{Satz 16.7.}\end{aligned}$$

Für allgemeines f folgt die Aussage mittels Linearität. □

Lemma X.4

$Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$. Es gilt

$$Y_n \xrightarrow[n\to\infty]{\mathbb{P}} \implies Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$$

Beweis. Seien $f \in C_b^g(\mathbb{R})$ und $\varepsilon > 0$ fixiert. Betrachte

$$|\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y)]| \leq \mathbb{E}[|f(Y_n) - f(Y)|] = \int_{\mathbb{R}} |f(Y_n) - f(Y)| \, \mathrm{d}\mathbb{P}$$

Da $f \in C_0^g$ existiert $M \in \mathbb{R}$ mit $|f| \leq M$ und

$$\exists \delta = \delta(\varepsilon) \in (0,1) \text{ so dass } \forall |x-y| \leq \delta \colon |f(x) - f(y)| < \varepsilon \tag{$\star$}$$

Schreibe nun

$$\int |f(Y_n) - f(Y)| \,\mathrm{d}\mathbb{P} = \int_{\{|Y_n - Y| \leq \delta\}} \ldots \mathrm{d}\,\mathbb{P} + \int_{\{|Y_n - Y| > \delta\}} \ldots \mathrm{d}\mathbb{P} = E_1 + E_2$$

mit

$$\begin{aligned} E_1 &\overset{(\star)}{\leq} \varepsilon \mathbb{P}(|Y_n - Y| \leq \delta) \\ E_2 &\leq 2M\mathbb{P}(|Y_n - Y| > \delta) \\ E_1 + E_2 &\leq \varepsilon \mathbb{P}(|Y_n - Y| \leq \delta) + 2M\mathbb{P}(|Y_n - Y| > \delta) \\ &\leq \varepsilon + 2M \underbrace{P(|Y_n - Y| > \delta)}_{\xrightarrow[n\to\infty]{} 0} \to \varepsilon \end{aligned}$$

Für $\varepsilon \to 0$ folgt die Behauptung. □

Lemma X.5
$Y, Y_1, Y_2, \ldots$ reelle Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$. Für $c \in \mathbb{R}$ konstant gilt

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y \equiv c \Longleftrightarrow Y_n \xrightarrow[n\to\infty]{\mathbb{P}} Y \equiv c$$

Beweis.

- $\Leftarrow$: Lemma X.4
- $\Rightarrow$: Für $\varepsilon > 0$ fixiert, wähle $f \in C_b(\mathbb{R})$ so dass

$$f(0) = 0 \quad f(x) \geq \mathbb{1}_{[-\varepsilon,\varepsilon]^C}(x) \quad \forall x \in \mathbb{R}$$

  Dann

$$\begin{aligned} \mathbb{P}(|Y_n - Y| > \varepsilon) &= \int_{[-\varepsilon,\varepsilon]^C} (Y_n - Y) \,\mathrm{d}\mathbb{P} \\ &\leq \int f(Y_n - Y) \,\mathrm{d}\mathbb{P} \\ &= \int f(Y_n - c) \,\mathrm{d}\mathbb{P} \\ &\xrightarrow[n\to\infty]{} \int f(Y - c) \,\mathrm{d}\mathbb{P} = \int f(0) \,\mathrm{d}\mathbb{P} = 0. \end{aligned}$$

□

Satz X.6
$Y, Y_1, Y_2, \dots$ reelle Zufallsvariablen mit Verteilungsfunktionen $F_Y, F_{Y_1}, F_{Y_2}, \dots$.
Dann gilt $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$ genau dann wenn

$$\lim_{n\to\infty} F_{Y_n} = F_Y(x) \quad \forall \text{ Stetigkeitsstellen } x \text{ von } F_Y.$$

Ist F_Y stetig, so gilt in diesem Fall sogar gleichmäßige Konvergenz

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_{Y_n}(x) - F_Y(x)| = 0.$$

Beweis. Es gelte $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$, d.h. $\mathbb{E}[f(Y_n)] \to \mathbb{E}[f(Y)]$ $\forall f \in C_b(\mathbb{R})$. Sei $x \in \mathbb{R}$ fixiert, $\varepsilon > 0$ und wähle $f \in C_b(\mathbb{R})$ mit

$$\mathbb{1}_{(-\infty,x]} \leq f \leq \mathbb{1}_{(-\infty,x+\varepsilon]}$$

Dann

$$\begin{aligned}
\limsup_{n\to\infty} F_{Y_n}(x) &= \limsup_{n\to\infty} \mathbb{P}(Y_n \leq x) \\
&= \limsup_{n\to\infty} \mathbb{E}[\mathbb{1}_{(-\infty,x]}(Y_n)] \\
&\leq \limsup_{n\to\infty} \mathbb{E}[f(Y_n)] \\
&= \lim \mathbb{E}[f(Y_n)] = \mathbb{E}[f(Y)] \\
&\leq \mathbb{E}[\mathbb{1}_{(-\infty,x+\varepsilon]}(Y)] = F_Y(x+\varepsilon)
\end{aligned}$$

Für $\varepsilon \to 0$ folgt, da F_Y rechtsstetig

$$\limsup_{n\to\infty} F_{Y_n}(x) \leq F_Y(x)$$

Ist x Stetigkeitsstelle von F_Y, so gilt auch $F_Y(x-\varepsilon) \to F_Y(x), \varepsilon \to 0$ und eine analoge Rechnung zeigt

$$\liminf_{n\to\infty} F_{Y_n}(x) \geq F_Y(x)$$

also

$$\lim F_{Y_n}(x) = F_Y(x)$$

Umgekehrt: Es gelte $\lim_{n\to\infty} F_{Y_n}(x) = F_Y(x)$ $\forall$ Stetigkeitsstellen x von F_Y. Fixiere $f \in C_b^g(\mathbb{R})$ (Portmanteau!), $\varepsilon > 0$ und wähle Stetigkeitsstellen $x_1 < x_2 < \dots x_n$ von F_Y, so dass

$$F_Y(x_1) < \varepsilon, F_Y(x_k) > 1 - \varepsilon \text{ und}$$
$$|f(y) - f(x_i)| < \varepsilon \quad \forall x_{i-1} \leq y \leq x_i$$

(möglich, da f gleichmäßig stetig und F_Y wegen Monotonie nur abzählbar viele Unstetigkeitsstellen besitzt.)

Das impliziert

$$\begin{aligned}
\mathbb{E}[f(Y_n)] &= \sum_{i=2}^{k} \mathbb{E}[f(y_n)\mathbb{1}_{\{x_{i-1}<Y_n\le x_i\}}] + \mathbb{E}[f(Y_n)\mathbb{1}_{\{Y_n\le x_1\}\cup\{Y_n>x_k\}}] \\
&\le \sum_{i=2}^{k}(f(x_i)+\varepsilon)(F_{Y_n}(x_i)-F_{Y_n}(x_{i-1})) + 2\varepsilon\|f\|_\infty \\
&\xrightarrow[n\to\infty]{} \sum_{i=2}^{k}(f(x_i)+\varepsilon)(F_Y(x_i)-F_Y(x_{i-1}) + 2\varepsilon\|f\|_\infty \\
&\le \mathbb{E}[f(Y)] + 2\varepsilon(1+2\|f\|_\infty)
\end{aligned}$$

Für $\varepsilon \to 0$ folgt

$$\limsup_{n\to\infty} \mathbb{E}[f(Y_n)] \le \mathbb{E}[f(Y)]$$

Der $\liminf$ folgt analog

$$\lim \mathbb{E}[f(Y_n)] = \mathbb{E}[f(Y)].$$

Zur gleichmäßigen Konvergenz: Sei F_Y stetig und $\varepsilon = k^{-1}, k \in \mathbb{N}$. Dann existieren $z_i \in \mathbb{R}$ mit

$$F_Y(z_i) = \frac{i}{k} \quad 0 < i < k$$

Da auch F_{Y_n} monoton wächst, gilt

$$\sup_{x\in\mathbb{R}} |F_{Y_n}(x) - F_Y(x)| \le \varepsilon + \underbrace{\max_{0\le i\le k} |F_{Y_n}(z_i) - F_Y(z_i)|}_{\xrightarrow[n\to\infty]{} 0}. \qquad \square$$

Satz X.7 (Stetigkeitssatz)
$Y, Y_1, Y_2, \ldots$ Zufallsvariablen in $\mathbb{R}^d$ mit charakteristischen Funktionen $\varphi_Y, \varphi_{Y_1}, \varphi_{Y_2}, \ldots$. Dann gilt

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y \iff \lim_{n\to\infty} \varphi_{Y_n}(u) = \varphi_Y(u) \quad \forall u \in \mathbb{R}^d$$

Für den Beweis benötigen wir:

Lemma X.8
$Y, Y_1, Y_2, \ldots$ Zufallsvariablen in $\mathbb{R}^d$ und sei $N = (N_1, \ldots, N_d)^T$ Zufallsvektor so dass $N_1, N_2, \ldots, N_d$ u.i.v. mit $N_i \sim \mathcal{N}(0,1)$, unabhängig von $Y, Y_1, Y_2, \ldots$ Falls:

$$Y_n + \sigma N \xrightarrow[n\to\infty]{\mathrm{d}} Y + \sigma N \quad \sigma > 0$$

dann folgt

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$$

Beweis. Sei $f \in C_b^g(\mathbb{R})$ und $\varepsilon > 0$ beliebig.

$$\begin{aligned}|\mathbb{E}[f(Y_n) - f(Y)]| \leq{} & |\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y_n + \sigma N)]| \\ & + \underbrace{|\mathbb{E}[f(Y_n + \sigma N)] - \mathbb{E}[f(Y + \sigma N)]|}_{\xrightarrow[n\to\infty]{} 0} \\ & + |\mathbb{E}[f(Y + \sigma N)] - \mathbb{E}[f(Y)]|\end{aligned}$$

Da $f \in C_b^g(\mathbb{R})$ existiert $\delta > 0$, so dass

$$|f(x) - f(y)| \leq \varepsilon \text{ für } |x - y| \leq \delta$$

Damit

$$\begin{aligned}|\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y_n + \sigma N)]| &= \left| \mathbb{E}[\underbrace{f(Y_n) - f(Y_n + \sigma N)}_{\leq \varepsilon \text{ mit } \|\sigma N\| < \delta}] \right| \\ &\leq \varepsilon + 2\|f\|_\infty \mathbb{P}(\|N\| > \frac{\delta}{\sigma}) \leq 2\varepsilon \quad \text{ für } \sigma \leq \sigma_0(\delta, \varepsilon), \forall n \in \mathbb{N}\end{aligned}$$

Dann folgt für n hinreichend groß $|\mathbb{E}[f(Y_n)] - \mathbb{E}[f(Y)]| \leq 5\varepsilon$. Daraus folgt die Behauptung. □

Beweis (Satz X.7). Es gilt $Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$, dann folgt

$$\begin{aligned}\varphi_{Y_n}(u) &= \mathbb{E}[e^{\mathrm{i}\langle u, Y_n\rangle}] = \mathbb{E}[\cos(\langle u, Y_n\rangle)] + \mathrm{i}\mathbb{E}[\sin(\langle u, Y_n\rangle)] \\ &\xrightarrow[n\to\infty]{} \mathbb{E}[\cos(\langle u, Y\rangle)] + \mathrm{i}\mathbb{E}[\sin(\langle u, Y\rangle)] \\ &= \varphi_Y(u).\end{aligned}$$

Gelte umgekehrt $\varphi_{Y_n}(u) \to \varphi_Y(u) \forall u \in \mathbb{R}^d$. Nach Lemma X.8 genügt zu zeigen, dass

$$Y_n + \sigma N \xrightarrow{\mathrm{d}} Y + \sigma N \quad \sigma > 0$$

mit N wie in Lemma X.8. Sei $f \in C_b(\mathbb{R})$, dann gilt (↗ Beweis zu Lemma VIII.12).

$$\mathbb{E}[f(Y_n + \sigma N)] = \int_{\mathbb{R}^d} f(z) \rho_{Y_n + \sigma N}(z) \, \mathrm{d}z$$

mit

$$\rho_{Y_n + \sigma N}(z) = (2\pi)^d \int_{\mathbb{R}^d} \varphi_{Y_n}(-y) e^{\mathrm{i}\langle z, y\rangle} e^{-\sigma^2 y^2/2} \, \mathrm{d}y$$

so dass

$$\begin{aligned}|\mathbb{E}[f(Y_n + \sigma N)] - \mathbb{E}[f(Y + \sigma N)]| &= \left| \int_{\mathbb{R}^d} f(z)(\rho_{Y_n + \sigma N}(z) - \rho_{Y + \sigma N}(z) \, \mathrm{d}z \right| \\ &\leq \|f\|_\infty \int_{\mathbb{R}^d} |\rho_{Y_n + \sigma N}(z) - \rho_{Y + \sigma N}(z)| \, \mathrm{d}z \\ &= \|f\|_\infty \int_{\mathbb{R}^d} (\rho_{Y_n + \sigma N}(z) - \rho_{Y + \sigma N}(z))^+ + (\rho_{Y_n + \sigma N}(z) - \rho_{Y + \sigma N}(z))^- \, \mathrm{d}z \\ &\overset{(\star)}{=} 2\|f\|_\infty \int_{\mathbb{R}^d} (\rho_{Y_n + \sigma N} - \rho_{Y + \sigma N})^+ \, \mathrm{d}z \\ &\xrightarrow[n\to\infty]{} 0\end{aligned}$$

Majorante: $\rho_{Y_n+\sigma N}(z)$ und $e^{-\sigma^2 y^2/2}$ nach dominierter Konvergenz, da $\varphi_{Y_n}(u) \to \varphi_Y(u)$.
Zu Gleichung I.$\star$:

$$\begin{aligned}
&\int_{\mathbb{R}^d} (\rho_{Y_n+\sigma N}(z) - \rho_{Y+\sigma N}(z))^+ \,\mathrm{d}z - \int_{\mathbb{R}^d} (\rho_{Y_n+\sigma N}(z) - \rho_{Y+\sigma N}(z))^- \,\mathrm{d}z \\
&= \int_{\mathbb{R}^d} (\rho_{Y_n+\sigma N}(z) - \rho_{Y+\sigma N}(z)) \,\mathrm{d}z \\
&= 1 - 1 = 0 \qquad (\star)
\end{aligned}$$

□

Mit dem Stetigkeitssatz lassen sich bekannte Resultate reproduzieren.

Lemma X.9 (Poissonapproximation)
Sei $X_n \sim \mathrm{Bin}(n, p_n)$, so dass $np_n \xrightarrow[n\to\infty]{}$ dann gilt

$$X_n \xrightarrow[n\to\infty]{\mathrm{d}} X \sim \mathrm{Poisson}(\lambda).$$

Beweis. Nach Beispiel V.15

$$\begin{aligned}
\varphi_{X_n}(u) &= \psi_{X_n}(e^{\mathrm{i}u}) = (1 + p_n(e^{\mathrm{i}u} - 1))^n \\
&= (1 + \frac{np_n(e^{\mathrm{i}u} - 1)}{n})^n \\
&\xrightarrow[n\to\infty]{} e^{\lambda(e^{\mathrm{i}u}-1)} = \psi_X(u)
\end{aligned}$$

und mit dem Stetigkeitssatz folgt die Behauptung. □

Satz X.10 (WLLN, u.i.v-Version in $\mathbb{R}^d$)
Seien $X_1, X_2, \ldots$ u.i.v $\mathbb{R}^d$-wertige Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ mit $X_i \in \mathscr{L}^1(\mathbb{P})$. Dann gilt

$$1/n \sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1].$$

Beweis. Für $u \in \mathbb{R}^d$ gilt

$$\begin{aligned}
\varphi_{1/n \sum X_i}(u) &= \varphi_{\sum X_i}(u/n) \\
&\overset{\text{unabh.}}{=} \prod_{i=1}^n \varphi_{X_i}(u/n) \overset{\text{ident. verteilt}}{=} (\varphi_{X_1}(u/n))^n \\
&\overset{\text{Taylor}}{=} (1 + 1/n\big\langle u, \varphi'_{X_1}(0)\big\rangle + o(\|u/n\|))^n \\
&= (1 + 1/n \mathrm{i}\langle u, \mathbb{E}[X_1]\rangle + o(\|u/n\|))^n \quad VIII.14 \\
&\xrightarrow[n\to\infty]{} \exp(\mathrm{i}\langle u, \mathbb{E}[X_1]\rangle)
\end{aligned}$$

und dies ist die charakteristische Funktion des Dirac-Maßes in $\mathbb{E}[X_1]$. Mit Stetigkeitssatz

$$\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow[n\to\infty]{\mathrm{d}} \mathbb{E}[X_1].$$

Und Lemma X.5 gibt

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_1].$$ □

Folgerung X.11 (Cramér-Wold device)
$Y, Y_1, Y_2, \ldots$ Zufallsvariablen in $\mathbb{R}^d$. Dann gilt

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y \iff \langle u, Y_n\rangle \xrightarrow[n\to\infty]{\mathrm{d}} \langle u, Y\rangle \quad \forall u \in \mathbb{R}^d.$$

Beweis. Es gilt

$$\varphi_{\langle u, Y_n\rangle}(t) = \mathbb{E}[e^{\mathrm{i}\langle u, Y_n\rangle}] = \varphi_{Y_n}(tu)$$

und $\varphi_{\langle u,Y\rangle}(t) = \varphi_{Y(tu)}$

$$\begin{aligned} Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y &\iff \varphi_{Y_n}(v) \to \varphi_Y(v) \quad \forall v \in \mathbb{R}^d \quad \text{Stetigkeitslemma} \\ &\overset{\text{s.o.}}{\iff} \varphi_{\langle u,Y_n\rangle}(t) \to \varphi_{\langle u,Y\rangle}(t) \quad \forall t \in \mathbb{R}, u \in \mathbb{R}^d \\ &\iff \langle u, Y_n\rangle \xrightarrow[n\to\infty]{\mathrm{d}} \langle u, Y\rangle \quad \forall u \in \mathbb{R}^d \quad \text{Stetigkeitslemma} \end{aligned}$$ □

Satz X.12 (Lemma von Slutsky)
Seien $Y, Y_1, Y_2, \ldots, Z, Z_1, Z_2, \ldots$ $\mathbb{R}^d$-wertige Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$, so dass

$$Y_n \xrightarrow[n\to\infty]{\mathrm{d}} Y \text{ und } Y_n - Z_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

Dann gilt

$$Z_n \xrightarrow[n\to\infty]{\mathrm{d}} Y$$

Beweis. Es gilt

$$\begin{aligned} \varphi_{Z_n}(u) &= \mathbb{E}[e^{\mathrm{i}\langle u, Z_n\rangle}] = \mathbb{E}[e^{\mathrm{i}\langle u, Z_n - Y_n\rangle} e^{\mathrm{i}\langle u, Y_n\rangle}] \\ &= \mathbb{E}[(e^{\mathrm{i}\langle u, Z_n - Y_n\rangle} - 1)e^{\mathrm{i}\langle u, Y_n\rangle}] + \underbrace{\mathbb{E}[e^{\mathrm{i}\langle u, Y_n\rangle}]}_{\varphi_{Y_n}(u) \xrightarrow[n\to\infty]{} \varphi_Y(u) \quad \forall u \in \mathbb{R}^d} \end{aligned}$$

Es genügt zu zeigen, dass

$$E := \mathbb{E}[(e^{\mathrm{i}\langle u, Z_n - Y_n\rangle} - 1)e^{\mathrm{i}\langle u, Y_n\rangle}] \xrightarrow[n\to\infty]{} 0$$

Dazu

$$\begin{aligned} |E| &\le \mathbb{E}\big[\big|(e^{\mathrm{i}\langle u, Z_n - Y_n\rangle} - 1)e^{\mathrm{i}\langle u, Y_n\rangle}\big|\big] \\ &= \mathbb{E}\big[\big|e^{\mathrm{i}\langle u, Z_n - Y_n\rangle} - 1\big|\big] \end{aligned}$$

Die Funktion $z \mapsto e^{\mathrm{i}\langle u,z\rangle}$, $u \in \mathbb{R}^d$, ist lokal LIPSCHITZ-stetig, dann für $u, y, z \in \mathbb{R}^d$

$$\begin{aligned}
\left|e^{\mathrm{i}\langle u,z\rangle} - e^{\mathrm{i}\langle u,y\rangle}\right| &= \left|e^{\mathrm{i}\langle u,z-y\rangle - 1}\right| \\
&= \left|\int_0^{\mathrm{i}\langle u,z-y\rangle} e^{\zeta}\, \mathrm{d}\zeta\right| \\
&\leq \sup_{|\zeta|\leq|\langle u,z-y\rangle|} \left|e^{\mathrm{i}\zeta}\right| \cdot |\mathrm{i}\langle u, z-y\rangle| = |\mathrm{i}\langle u, z-y\rangle| \\
&\leq |u| \cdot |z-y|
\end{aligned}$$

Damit

$$\begin{aligned}
|E| &\leq \mathbb{E}[\left|e^{\mathrm{i}\langle u,Z_n-Y_n\rangle}\right| \mathbb{1}_{\{|Z_n-Y_n|\leq\delta\}}] \\
&+ \mathbb{E}[\left|e^{\mathrm{i}\langle u,Z_n-Y_n\rangle-1} \mathbb{1}_{\{|Z_n-Y_n|\leq\delta\}}\right|] \\
&\leq \delta|u| + 2\, \underbrace{\mathbb{P}(|Z_n - Y_n| \leq \delta)}_{\xrightarrow[n\to\infty]{} 0} \\
&\xrightarrow[n\to\infty]{} \delta \cdot |u| \\
&\xrightarrow[\delta\to 0]{} 0.
\end{aligned}$$

□

2. Der Zentrale Grenzwertsatz

(CLT, central limit theorem)

Satz X.13 (CLT, u.i.v -Version)
$X_1, X_2, \dots$ reelle u.i.v Zufallsvariablen mit $\mathbb{E}[X_1] =: \mu$ und $0 < \mathbb{V}\mathrm{ar}(X_1) =: \sigma^2 < \infty$. Dann gilt

$$\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{\sigma^2 n}} \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0,1).$$

Beweis. Verwende den Stetigkeitssatz:

$$\begin{aligned}
\varphi_{\sum(X_k-\mu)/\sqrt{\sigma^2 n}}(u) &= \mathbb{E}[e^{\mathrm{i}u(\sum(X_k-\mu)/\sqrt{\sigma^2 n})}] \\
&= \prod_{k=1}^n \mathbb{E}[e^{\mathrm{i}u(\sum(X_k-\mu)/\sqrt{\sigma^2 n})}] \quad X_k \text{ unabh.} \\
&= \left(\mathbb{E}[e^{\mathrm{i}u(\sum(X_k-\mu)/\sqrt{\sigma^2 n}}]\right)^n \quad X_k \text{ identisch verteilt} \\
&= \left(\varphi_{X_1-\mu}(u/\sqrt{\sigma^2 n})\right)^n \\
&= \left(1 + \mathrm{i}\frac{u}{\sigma\sqrt{\mu}} \underbrace{\mathbb{E}[X_1-\mu]}_{=0} + \frac{\mathrm{i}^2}{2} \underbrace{\mathbb{E}[(X_1-\mu)^2]}_{\sigma^2} + o(\left\|\frac{u}{\sqrt{\sigma^2 n}}\right\|^2)\right) \quad \text{TAYLOR} \\
&= (1 - \frac{1}{2}\frac{u^2}{n} + o(\left\|\frac{u}{\sqrt{\sigma^2 n}}\right\|^2))^n \\
&\xrightarrow[n\to\infty]{} \exp(-1/2u^2) = \varphi_{\mathcal{N}(0,1)}(u).
\end{aligned}$$

□

Folgerung X.14 (CLT, De Moivre-Laplace)
Sei $S_n = \mathrm{Bin}(n,p), p \in (0,1)$, dann

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow[n\to\infty]{\mathrm{d}} Z \sim \mathcal{N}(0,1).$$

Beweis. Es gilt: $S_n \overset{\mathrm{d}}{=} X_1 + \cdots + X_n$, für $X_i \sim \mathrm{Bernoulli}(p)$ unabhängig, mit $\mathbb{E}[X_1] = p, \mathbb{V}\mathrm{ar}(X_1) = p(1-p)$. □

▶ **Bemerkung**

Auch der lokale Grenzwertsatz von De Moivre-Laplace (Satz VII.2) impliziert Folgerung X.14, siehe z.B. ↗ Dehling & Haupt.

Die Bedingungen des CLT lassen sich abschwächen. Zum Beispiel kann "identisch verteilt" durch die Lindeberg-Bedingung ersetzt werden:

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} \frac{1}{\sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)} \sum_{k=1}^n \mathbb{E}[(X_k \mathbb{E}[X_k])^2 \mathbb{1}_{\{|X_k| > \varepsilon \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)\}}] = 0$$

und dann gilt

$$\frac{\sum_{k=1}^n (X_k - \mathbb{E}[X_k])}{\sqrt{\sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)}} \xrightarrow[n\to\infty]{\mathrm{d}} Z \sim \mathcal{N}(0,1)$$

Der folgende CLT geht noch etwas weiter und betrachtet die Zufallsvariablen in einem Dreiecksschema:

$$\begin{array}{llllll}
X_{1,1} & X_{1,2} & \dots & X_{1,k(1)} & & \\
X_{2,1} & X_{2,2} & & & X_{2,k(2)} & \\
\vdots & & & & & \text{mit } k(n) \to \infty \text{ für } n \to \infty \\
X_{k,1} & X_{k,2} & & & & X_{k,k(k)} \\
\vdots & & & & &
\end{array}$$

Satz X.15 (CLT, Lindeberg-Feller)
Seien $\{X_{n,k}, k = 1, \ldots, k(n), n \in \mathbb{N}\}$ reelle Zufallsvariablen in einem Dreiecksschema mit $\mathbb{E}[X_{n,k}] = 0$ und $0 < \mathbb{V}\mathrm{ar}(X_{n,k}) = \sigma_{n,k}^2 < \infty$, so dass $\forall n \in \mathbb{N}$ die Zufallsvariablen $X_{n,1}, \ldots, X_{n,k(n)}$ unabhängig sind. Zudem gelten

$$\sum_{k=1}^{k(n)} \sigma_{n,k}^2 \xrightarrow[n\to\infty]{} \sigma^2 \in (0, \infty)$$

und die LINDEBERG-Bedingung für Schemata:

$$\forall \varepsilon > 0\colon \lim_{n\to\infty} \sum_{k=1}^{k(n)} \mathbb{E}[X_{n,k}^2 \mathbb{1}_{|X_{n,k}|>\varepsilon}] = 0$$

Dann gilt

$$\sum_{k=1}^{k(n)} X_{n,k} \xrightarrow[n\to\infty]{\mathrm{d}} Z \sim \mathscr{N}(0, \sigma^2)$$

Beweis. Wir beginnen mit einer TAYLORentwicklung der charakteristischen Funktion der $X_{n,k}$:

$$\varphi_{X_{n,k}} = 1 + ... + \mathrm{i}\,\underbrace{\mathbb{E}[X_{n,k}]}_{=0} \cdot u - \frac{1}{2} \underbrace{\mathbb{E}[X_{n,k}^2]}_{\sigma_{n,k}^2} u^2 + R_{n,k}(u)$$

wobei wir $R_{n,k}$ folgendermaßen abschätzen. Für $x \in \mathbb{R}$ gilt

$$\begin{aligned} \left| e^{\mathrm{i}x} - 1 - \mathrm{i}x - 1/2(\mathrm{i}x)^2 \right| &= \left| \int_0^x \int_0^t (1 - e^{\mathrm{i}y}) \,\mathrm{d}y \,\mathrm{d}t \right| \\ &\leq \left| \int_0^{|x|} \int_0^{|t|} (1 - e^{\mathrm{i}y}) \,\mathrm{d}y \,\mathrm{d}t \right| \end{aligned}$$

mit $\left|1 - e^{\mathrm{i}y}\right| = \left|\int_0^y e^{\mathrm{i}\,z\,\mathrm{d}z}\right| \leq |y| \wedge 2$ so dass $\left|e^{\mathrm{i}x} - 1 - \mathrm{i}x + 1/2x^2\right| \leq {}^{|x|^3}\!/_6 \wedge x^2$. Für $x = uX_{n,k}$ folgt im Erwartungswert

$$\begin{aligned} |R_{n,k}(u)| &\leq u^2 \mathbb{E}[X_{n,k}^2 \wedge \frac{|u| \cdot \left|X_{n,k}^3\right|}{6}] \\ &= u^2 \left(\int_{\left|X_{n,k}>\varepsilon\right|} (X_{n,k}^2 \wedge \frac{|u| \cdot \left|X_{n,k}^2\right|}{6}) \,\mathrm{d}\mathbb{P} + \int_{\left|X_{n,k}\leq\varepsilon\right|} (X_{n,k}^2 \wedge \frac{|u| \cdot \left|X_{n,k}^3\right|}{6}) \,\mathrm{d}\mathbb{P} \right) \\ &\leq= u^2 \left(\int_{\left|X_{n,k}>\varepsilon\right|} (X_{n,k}^2) \,\mathrm{d}\mathbb{P} + \int_{\left|X_{n,k}\leq\varepsilon\right|} \frac{|u| \cdot \left|X_{n,k}^3\right|}{6} \,\mathrm{d}\mathbb{P} \right) \\ &\leq u^2 \mathbb{E}[X_{n,k}^2 \mathbb{1}_{\left|X_{n,k}>\varepsilon\right|} + \varepsilon/6 |u|^3 \mathbb{E}[X_{n,k}^2]] \end{aligned}$$

Seien $\{Y_{n,k}, k = 1, \ldots, k(n), n \in \mathbb{N}\}$ Zufallsvariablen mit $Y_{n,k} \sim \mathscr{N}(0, \sigma_{n,k}^2)$, so dass $\forall n$ die Zufallsvariablen $Y_{n,1}, \ldots, Y_{n,k(n)}$ unabhängig sind. Dann gilt (analog zu oben)

$$\varphi_{Y_{n,k}}(u) = 1 - 1/2\sigma_{n,k}^2 u^2 + \tilde{R}_{n,k}(u)$$

mit

$$\begin{aligned}\left|\tilde{R}_{n,k}(u)\right| &\leq u^2\mathbb{E}[Y_{n,k}^2 \wedge \frac{|u||Y_{n,k}|^3}{6}] \leq |u|^3\frac{\mathbb{E}[\left|Y_{n,k}^3\right|]}{6} \\ &\leq |u|^3 C\sigma_{n,k}^3 \quad \text{Übung und für eine Konstante } C \\ &\leq |u|^3 C\sigma_{n,k}^2 \max_{1\leq k\leq k(n)} \sigma_{n,k}\end{aligned}$$

Zusammen folgt

$$\begin{aligned}\left|\varphi_{X_{n,k}}(u) - \varphi_{Y_{n,k}}(u)\right| &\leq |R_{n,k}(u)| + \left|\tilde{R}_{n,k}(u)\right| \\ &\leq u^2 \underbrace{\mathbb{E}[X_{n,k}^2 \mathbb{1}_{|X_{n,k}>\varepsilon|}]}_{\xrightarrow[n\to\infty]{}0 \text{ Lindeberg}} + \underbrace{\frac{\varepsilon}{6}\sigma_{n,k}^2}_{\to 0, \varepsilon\to 0} + |u|^3 C \underbrace{\max_{1\leq k\leq k(n)} \sigma_{n,k}}_{\xrightarrow[n\to\infty]{}0}\end{aligned}$$

(vgl. Satz 11.5) und mit der Unabhängigkeitsannahme:

$$\begin{aligned}\left|\varphi_{X_{n,1}+\ldots+X_{n,k(n)}}(u) - \varphi_{Y_{n,1}+\ldots+Y_{n,k(n)}}(u)\right| &= \left|\prod_{k=1}^{k(n)} \varphi_{X_{n,k}}(u) - \prod_{k=1}^{k(n)} \varphi_{Y_{n,k}}(u)\right| \\ &\overset{(\star)}{=} \left|\sum_{i=1}^{k(n)} (\prod_{j=1}^{n-1} \varphi_{X_{n,j}}(u)) \cdot (\varphi_{X_{n,k}}(u) - \varphi_{Y_{n,k}}(u))(\prod_{j=n+1}^{k(n)} \varphi_{Y_{n,j}}(u)\right| \\ &\leq \sum_{k=1}^{k(n)} \left|\varphi_{X_{n,k}}(u) - \varphi_{Y_{n,k}}(u)\right| \xrightarrow[n\to\infty, \varepsilon\to 0]{} 0\end{aligned}$$

mit ($\star$), da

$$\begin{aligned}\sum_{k=1}^{n}(\prod_{j=1}^{k-1} a_j(a_k - b_k) \prod_{j=k-1}^{n} b_j) &= (a_1 - b_1)\prod_{j=2}^{n} b_j + a_1(a_2 - b_2)\prod_{j=3}^{n} b_j + \ldots + \prod_{j=1}^{n-1} a_j(a_n - b_n) \\ &= \prod_{k=1}^{n} a_n - \prod_{k=!}^{n} b_n \qquad (\star)\end{aligned}$$

Da

$$\begin{aligned}\varphi_{Y_{n+1}+\ldots+Y_{n,k(n)}} &= \prod_{k=1}^{k(n)} \varphi_{Y_{n,k(n)}} \\ &= \exp(-u^2(\sigma_{n,1}^2 + \ldots + \sigma_{n,k(n)}^2)1/2) \\ &\xrightarrow{n\to\infty} e^{-u^2\sigma^2/2} \quad u \in \mathbb{R}\end{aligned}$$

folgt

$$\varphi_{X_{n,1}+\ldots+X_{n,k(n)}}(u) \xrightarrow{n\to\infty} e^{-u^2\sigma^2/2} \quad u \in R$$

und mit Stetigkeitsatz (put ref here!) folgt die Behauptung. □

Kapitel XI

Diskrete Martingale

Definition XI.1
Eine Folge $(X_n)_{n\in\mathbb{N}_0}$ von reellen Zufallsvariablen auf $(\Omega, \mathscr{F}\,\mathbb{P})$ heißt Martingal, falls

1. $\mathbb{E}[|X_n|] < \infty$ also $X_n \in \mathscr{L}^1(\mathbb{P}) \forall n \in \mathbb{N}_0$
2. $\mathbb{E}[X_{n-1} \mid X_0, X_1, \ldots, X_n] = X_0, \mathbb{P}$-f.s. $\forall n \in \mathbb{N}$

Die Folge $(X_n)_{n\in\mathbb{N}}$ heißt Super-/Submartingal, falls 1. und folgendes gilt

1. $\mathbb{E}[X_{n-1} \mid X_0, X_1, \ldots, X_n] \leq / \geq X_0, \mathbb{P}$-f.s. $\forall n \in \mathbb{N}$

■ **Beispiel XI.2**
Sei $(Y_n)_{n\in\mathbb{N}}$ Folge von u.i.v. Zufallsvariablen auf $(\Omega, \mathscr{F}, \mathbb{P})$ in $\mathscr{L}^1(\mathbb{P})$, reellwertig mit $\mathbb{E}[Y_1] = 0$. Dann ist $(X_n)_{n\in\mathbb{N}}$ mit

$$X_0 = 0 \text{ und } X_n = \sum_{k=1}^{n} Y_k \quad n \geq 1$$

ein Martingal, denn

1. $\mathbb{E}[|X_n|] \leq \sum_{k=1}^{n} \mathbb{E}[|Y_k|] < \infty$
2.
$$\begin{aligned}\mathbb{E}[X_{n+1} \mid X_0, \ldots, X_n] &= \mathbb{E}[X_n + Y_{n+1} \mid X_0, ..., X_n] = \mathbb{E}[X_n \mid X_0, ...X_n] + \mathbb{E}[Y_{n+1} \mid X_0, ...X_n] \\ &= X_n + \mathbb{E}[Y_{n+1}] = X_n \quad \mathbb{P}\text{-f.s. } \forall n \in \mathbb{N}\end{aligned}$$

   Für $\mathbb{E}[Y_1] \geq / \leq 0$ erhält man dementsprechend ein Sub-/Supermartingal.

■ **Beispiel XI.3**
Sei $(X_n)_{n\in\mathbb{N}}$ ein (Super-/Sub-)Martingal auf $(\Omega, \mathscr{F}, \mathbb{P})$ und $(C_n)_{n\in\mathbb{N}}$ eine Folge von beschränkten Zufallsvariablen in $[0, \infty)$, sodass C_n $\sigma(X_0, \ldots, X_{n-1})$-messbar ist. Dann ist $(Y_n)_{n\in\mathbb{N}_0}$ mit

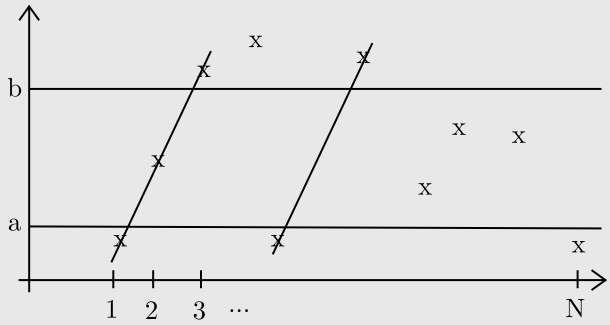
$$Y_0 := 0 \quad Y_n = \sum_{i=1}^{n} C_i(X_i - X_{i-1}) \quad \text{für } n \leq 1$$

ein (Super-/Sub-)Martingal ($\nearrow$ Übung).
Interpretation:

- $(X_i - X_{i-1})$ entspricht Gewinn in Runde i pro Einsatzeinheit (X Martingal $\rightsquigarrow$ faires Spiel mit Supermartingal $\rightsquigarrow$ nachteilig und Submartingal $\rightsquigarrow$ vorteilig)
- C_i entspricht Einsatz in Runde i

- Y_n entspricht Gewinn nach n Runden

Lemma XI.4 (Doob's Upcrossing Lemma)
Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Supermartingal auf $(\Omega, \mathscr{F}, \mathbb{P})$ und für $a, b \in \mathbb{R}, N \in \mathbb{N}$. Sei $U_N[a,b] = \#$ Upcrossings von $[a,b]$ durch X bis zur Zeit N:



d.h.

$$U_N[a,b](\omega) = \{\max k \in \mathbb{N}_0 : \exists 0 \leq S_1 < t_1 < S_2 < t_2 < ... < s_k < t_k \leq N$$
$$: X_{s_i} < a, X_{t_i} > b, i \in \{1, ..., k\}\}$$

Dann gilt

$$\mathbb{E}[U_N[a,b]] \leq \frac{\mathbb{E}[(X_N - a)^-]}{b-a}$$

Beweis. Interpretiere $(X_i - X_{i-1})$ als Gewinn in Spielrunde i pro Einsatzeinheit. Wähle als Spielstrategie

$$C_1 := \mathbb{1}_{\{X_0 < a\}}$$
$$C_n := \mathbb{1}_{\{C_{n-1}=1\}} \mathbb{1}_{\{X_{n-1} \leq b\}} + \mathbb{1}_{\{C_{n-1}=0\}} \mathbb{1}_{\{X_{n-1} \leq a\}}$$

Dann ist $(C_n)_{n\in\mathbb{N}}$ beschränkt, nicht-negativ und C_n ist $\sigma(X_0, \ldots, X_{n-1})$ messbar. Nach Beispiel XI.3 ist $(Y_n)_{n\in\mathbb{N}_0}$ mit

$$Y_0 = 0$$
$$Y_n = \sum_{i=1}^{n} C_i(X_i - X_{i-1})$$

ein Superimartingal. Es folgt

$$\mathbb{E}[Y_n] = \mathbb{E}[\mathbb{E}[Y_N \mid Y_0, ..., Y_{N-1}]] \leq \mathbb{E}[Y_{N-1}] \leq \cdots \leq \mathbb{E}[Y_0] = 0$$

Zudem gilt $\forall \omega \in \Omega$

$$Y_N(\omega) \geq (b-a)U_N[a,b] - (X_N(\omega) - a)^-$$
$$\implies (b-a)\mathbb{E}[U_N[a,b]] \leq \mathbb{E}[Y_N] + \mathbb{E}[(X_N - a)^-]$$
$$\leq \mathbb{E}[(X_N - a)^-]$$

□

Satz XI.5 (CLT für Martingale)
Sei $(X_n)_{n\in\mathbb{N}_0}$ ein Martingal mit $X_n \in \mathscr{L}^2(\mathbb{P}) \forall n \in \mathbb{N}_0$, so dass $X_0 = 0$ und

$$\mathbb{E}[\Delta_n^2 \mid \mathscr{F}_{n-1}] = \sigma^2$$

deterministisch ist, wobei

$$\Delta_n := X_n - X_{n-1} \text{ und } \mathscr{F}_n := \sigma(X_0, \dots, X_n)$$

Gilt zudem die Lindberg- Bedingung für Martingale

$$\forall \varepsilon > 0\colon \lim_{n\to\infty} 1/s_n^2 \sum_{k=1}^{n} \mathbb{E}[\Delta_k^2 \mathbb{1}_{|\Delta_k| > \varepsilon s_n} \mid \mathscr{F}_{k-1}] = 0$$

mit $s_n^2 = \sum_{k=1}^n \sigma_k^2$. Dann folgt

$$\frac{x_n}{s_n} \xrightarrow[\mathrm{d}]{n\to\mathbb{N}} Z \sim \mathscr{N}(0,1)$$

Die Folge $(\mathscr{F}_n)_{n\in\mathbb{N}}$ ist eine "Filtration".

Beweis. Ähnlich zum Beweis des CLT nach Lindberg-Feller:

$$\mathbb{E}[e^{\mathrm{i}u\Delta_k} \mid \mathscr{F}_{k-1}] = 1 + \mathrm{i}u \underbrace{\mathbb{E}[\Delta_k \mid \mathscr{F}_{k-1}]}_{\mathbb{E}[X_k - X_{k-1} \mid \mathscr{F}_{k-1}]} - u^2 \underbrace{\mathbb{E}[\Delta_k^2 \mid \mathscr{F}_{k-1}]}_{\sigma_k^2} + R_k(u)$$

wobei

$$\begin{aligned}\mathbb{E}[\Delta_k \mid \mathscr{F}_{k-1}] &= \mathbb{E}[X_k - X_{k-1} \mid \mathscr{F}_{k-1}] \\ &= X_{k-1} - X_{k-1} \quad \text{Martingale-Eigenschaft und mb. Herausziehen} \\ &= 0\end{aligned}$$

mit

$$|R_k(u)| \leq u^2 \mathbb{E}[\Delta_k^2 \mathbb{1}_{|\Delta_k| > \varepsilon s_n} \mid \mathscr{F}_{k-1}] + \frac{\varepsilon}{6}|u|^3 s_n \sigma_k^2$$

Seien $(Y_k \sim \mathscr{N}(0, \sigma^2))$ unabhängige Zufallsvariablen, unabhängig von $(X_n)_{n\in\mathbb{N}_0}$ mit $Y_k \sim \left\|0, \sigma_k^2\right\|$, dann folgt (wieder analog!)

$$\varphi_{Y_k}(u) = 1 - \frac{1}{2}\sigma_k^2 u^2 + \tilde{R}_k(u)$$

mit

$$\left|\tilde{R}_k(u)\right| \leq u^2 \mathbb{E}[Y_k^2 \mathbb{1}_{|Y_k| > \varepsilon s_n}] + \frac{\varepsilon}{6}|u|^3 + s_n \sigma_k^2$$

sodass

$$\left|\mathbb{E}[e^{iu\Delta_k} \mid \mathscr{F}_{k-1}] - \mathbb{E}[e^{iuY_k}]\right| \leq u^2 \mathbb{E}[\Delta_k^2 \mathbb{1}_{|\Delta_k| < \varepsilon s_n} \mid \mathscr{F}_{k-1}] + u^2 \mathbb{E}[Y_k^2 \mathbb{1}_{|Y_k| > \varepsilon s_n}] + \frac{\varepsilon}{3}|u|^3 s_n \sigma_k^2 \qquad (*)$$

zudem

$$z_n := \sum_{k=1}^{n} \sim \mathcal{N}(0,\sigma_1^2)\cdot\dots\cdot\mathcal{N}(0,\sigma_n^2) = \mathcal{N}\left(0, \sum_{k=1}^{n}\sigma_k^2\right) = \mathcal{N}(0,s_n^2)$$

sodass

$$\frac{z_n}{s_n} \sim \mathcal{N}(0,1)$$

und nach Satz VIII.VIII.11

$$\varphi_{z_n/s_n}(u) = \exp\left(-\frac{u^2}{2}\right)$$

Es folgt

$$\begin{aligned}
|\varphi_{x_n}(u) - \varphi(z_n)(u)| &= \left|\mathbb{E}[e^{iux_n}] - \mathbb{E}[e^{iuz_n}]\right| \\
&= \left|\mathbb{E}[\mathbb{E}[e^{iux_n} - e^{iuz_n} \mid \mathscr{F}_{n-1}]]\right| \\
&= \left|\mathbb{E}[\mathbb{E}[e^{iux_{n-1}}e^{iu\Delta_n} - e^{iuz_{n-1}}e^{iuY_n} \mid \mathscr{F}_{n-1}]]\right| \\
&= \left|\mathbb{E}[e^{iux_{n-1}}\mathbb{E}[e^{iu\Delta_n} \mid \mathscr{F}_{n-1}] - \mathbb{E}[e^{iuz_{n-1}}]\mathbb{E}[e^{iuY_n}]]\right| \\
&\leq \left|\mathbb{E}[e^{iux_{n-1}}(\mathbb{E}[e^{iu\Delta_n} \mid \mathscr{F}_{n-1}] - \mathbb{E}[e^{iuY_n}])]\right| \\
&+ \left|\mathbb{E}[(e^{iux_{n-1}} - \mathbb{E}[e^{iuz_{n-1}}])\mathbb{E}[e^{iuY_n}]]\right| \\
&\leq \mathbb{E}[\left|\mathbb{E}[e^{iu\Delta_n} \mid \mathscr{F}_{n-1}] - \mathbb{E}[e^{iuY_n}]\right|] + \underbrace{\left|\mathbb{E}[(e^{iux_{n-1}} - \mathbb{E}[e^{iuz_{n-1}}])]\right|}_{\text{wende obige Schritte an}} \cdot \underbrace{\left|\mathbb{E}[e^{iuY_n}]\right|}_{\leq 1} \\
&\leq \sum_{k=1}^{n}\mathbb{E}[\left|\mathbb{E}[e^{iu\Delta_k} \mid \mathscr{F}_{k-1}] - \mathbb{E}[e^{uiY_k}]\right|]
\end{aligned}$$

Für $u = \frac{v}{s_n}$ folgt mit (*)

$$\begin{aligned}
\left|\varphi_{x_n/s_n}(v) - \varphi_{z_n/s_n}(v)\right| &= |\varphi_{x_n}(u) - \varphi_{z_n}(u)| \\
&\leq \sum_{k=1}^{n}\mathbb{E}[\left|\mathbb{E}[e^{iu\Delta_k} \mid \mathscr{F}_{k-1}] - \mathbb{E}[e^{iuY_k}]\right|] \\
&\leq \sum_{k=1}^{n}\mathbb{E}[u^2\mathbb{E}[\Delta_k^2\mathbb{1}_{|\Delta_k|>\varepsilon s_n} \mid \mathscr{F}_{k-1}] + u^2\mathbb{E}[Y_k^2\mathbb{1}_{|Y_k|>\varepsilon s_n}] + \frac{\varepsilon}{3}|u|^3 s_n\sigma_k^2] \\
&= \underbrace{v^2\mathbb{E}[\frac{1}{s_n^2}\sum_{k=1}^{n}\mathbb{E}[\Delta_k^2\mathbb{1}_{|\Delta_k|>\varepsilon s_n} \mid \mathscr{F}_{k-1}]]}_{\to 0 \text{ Lindeberg mit dom. Konvergenz}^{1}} \\
&+ \underbrace{\frac{v^2}{s_n^2}\sum_{k=1}^{n}\mathbb{E}[Y_k^2\mathbb{1}_{|Y_k|>\varepsilon s_n}]}_{\to 0} + \underbrace{\sum_{k=1}^{n}\frac{\varepsilon}{3}|v|^3\frac{\sigma_k^2}{s_n^2}}_{=\frac{\varepsilon}{3}|v|^3} \qquad (**) \\
&\to \frac{\varepsilon}{3}|v|^3 \\
&\to 0
\end{aligned}$$

Die Behauptung folgt mit dem Stetigkeitssatz , wenn (**) gilt.

Dazu

$$\begin{aligned}
\mathbb{E}[Y_k \mathbb{1}_{|Y_k|>\varepsilon s_n}] &= \frac{1}{\sqrt{2\pi\sigma_k^2}} \int_{|x|>\varepsilon s_n} x^2 \exp\left(\frac{-x^2}{2\sigma_k^2}\right) \,\mathrm{d}x \\
&\overset{y=\frac{x}{\sigma_k}}{=} \frac{\sigma_k^2}{\sqrt{2\pi}} \int_{|y|>\varepsilon s_n} y^2 \exp\left(\frac{-y^2}{2}\right) \,\mathrm{d}y \\
&\le \frac{\sigma_k^2}{\sqrt{2\pi}} \int_{|y|>\varepsilon \min_{k\le n}\{s_n/\sigma_k\}} y^2 \exp\left(\frac{-y^2}{2}\right) \,\mathrm{d}y \\
\Rightarrow \frac{1}{s_k^2} \sum_{k=1}^{n} \mathbb{E}[Y_k^2 \mathbb{1}_{|Y_k|>\varepsilon s_n}] &\le \frac{1}{\sqrt{2\pi}} \int_{|y|>\varepsilon \min_{k\le n}\{s_n/\sigma_k\}} y^2 \exp\left(\frac{-y^2}{2}\right) \,\mathrm{d}y \qquad (***)
\end{aligned}$$

Da

$$\begin{aligned}
\sigma_k^2 = \mathbb{E}[\Delta_k^2 \mid \mathscr{F}_{k-1}] &= \mathbb{E}[\Delta_k^2 \mathbb{1}_{|\Delta_k|\le\varepsilon s_n} \mid \mathscr{F}_{k-1}] + \mathbb{E}[\Delta_k^2 \mathbb{1}_{|\Delta_k|>\varepsilon s_n} \mid \mathscr{F}_{k-1}] \\
&\le \varepsilon^2 s_n^2 + \mathbb{E}[\Delta_k^2 \mathbb{1}_{|\Delta_k|>\varepsilon s_n} \mid \mathscr{F}_{k-1}]
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \max_{k\le n}\{\frac{\sigma_k^2}{s_n^2}\} \le \varepsilon^2 + \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}[\Delta_k^2 \mathbb{1}_{|\Delta_k|>\varepsilon s_n} \mid \mathscr{F}_{k-1}] &\xrightarrow{\text{Lindeberg}} \varepsilon^2 \to 0 \\
\Rightarrow \min_{k\le n}\{\frac{s_n}{\sigma_k}\} &\to \infty \\
\Rightarrow \left\{ |y| > \varepsilon \min_{k\le n}\{\frac{s_n}{\sigma_k}\} \right\} &\to \varnothing \\
\Rightarrow \text{mit dominierter Konvergenz folgt } (**) &
\end{aligned}$$

□

Anhang

Literaturverzeichnis

[1] Bauer, H. Wahrscheinlichkeitstheorie, 5 ed. De Gruyter, 2002.

[2] Dehling, H., and Haupt, B. Einführung in die Wahrscheinlichkeitstheorie und Statistik. Springer, 2003.

[3] Georgii, H.-O. Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik, 5 ed. De Gruyter, 2015.

[4] Krengel, U. Einführung in die Wahrscheinlichkeitstheorie und Statistik:. Vieweg, 2005.

[5] Schilling, R. L. Wahrscheinlichkeit: eine Einführung für Bachelor-Studenten, 1 ed. De Gruyter, 2017.

Index