Stochastik SS 2019

Dozent: Prof. Dr. Anita Behme

17. Mai 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. $A_1, A_2, \dots \in \mathscr{F} \Rightarrow \bigcup_{i \geq 1} \in \mathscr{F}$

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

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

σ-Additivität für paarweise disjunkte Ereignisse$A_1, A_2, \dots \in \mathscr{F} \Rightarrow \mathbb{P}\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} \mathbb{P}(A_i)$ (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:
$$\mathbb{P}(A_n) \longrightarrow \mathbb{P}(A), n \to \infty$$

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: $A \cap B = \varnothing \Longrightarrow \mathbb{P}(A \uplus B) = \mathbb{P}(A) + \mathbb{P}(B)$.
Es kann σ-Additivität verwendet werden, indem "fehlende" Mengen durch $\varnothing$ ergänzt werden:

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

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 $m \to \infty$
$$A_m = \biguplus_{i=1}^{m} F_i \Rightarrow A = \biguplus_{i=1}^{\infty} F_i = \biguplus_{i=1}^{\infty} A_i$$
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
$$\delta_\xi(A) := \begin{cases} 1 & \xi \in A \\ 0 & \text{sonst} \end{cases}$$
eine (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 $\mathscr{B}(\mathbb{R}_+) = \mathscr{F}$. Eine 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 ↗ 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

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

  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. $\{x \in \Omega \colon f(x) \leq c\} \in \mathscr{B}(\Omega)$ 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 bezeichnet wir als Dichte, Dichtefunktion/Wahrscheinlichkeitsdichte von $\mathbb{P}$ und nennen ein solches $\mathbb{P}$ (absolut)stetig (bzgl. denn Lebesgue-Maß).

Beweis.
- Der diskrete Fall ist klar.
- Im stetigen Fall folgt die Bahuptung aus den bekannten Eigenschaften des Lebesgue-Integrals (↗ 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 $\Omega \subset \mathbb{R}^n$ definiert 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.

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

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 $\mathscr{F} = \omega(\mathscr{G})$ für ein $\cap$-stabiles Erzeugendensystem $\mathscr{G} \subset \mathscr{P}(\Omega)$. 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 Wahrscheinlichkeitsver-

teilung 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

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

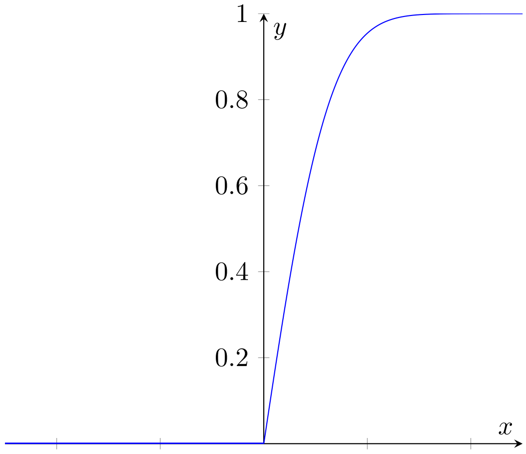
die (kumulative) Verteilungsfunktion von X.

■ **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

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



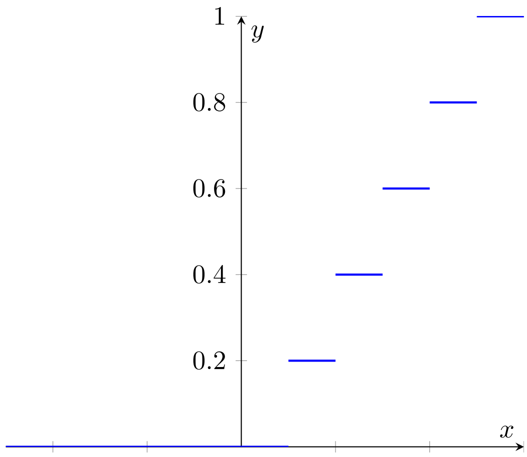
■ **Beispiel I.18**

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

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

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. $\lim\limits_{x\to-\infty} F(x) = 0,\ \lim\limits_{x\to\infty} F(x) = 1$

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

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

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 Definition I.16 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 $\Omega \rightarrow U(\Omega)$

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.

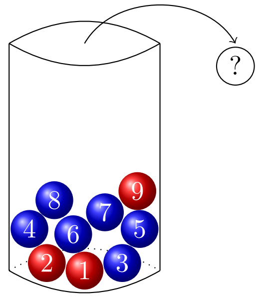


Abbildung II.1: Urnenmodell

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 <u>Produktdichten</u> 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

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

und

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

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

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

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

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

Multinomialverteilung mit Parametern n und p. Wir schreiben auch 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

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

Binomialverteilung auf $\{0, \ldots, n\}$ mit Parameter p (auch Erfolgswahrscheinlichkeit). Wir schreiben auch $\mathrm{Bin}(n,p)$. 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 $\mathrm{Bernoulli}(p)$.

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 $\{0, \ldots, n\}$ mit Zähldichte

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

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

3. Poisson-Approximation und Poisson-Verteilung

$\mathrm{Bin}(n,p)$ 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

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

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 $a(l) \overset{n\to\infty}{\sim} b(l) \Leftrightarrow \frac{a(l)}{b(l)} \xrightarrow{n\to\infty} 1$. 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 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: $\mathbb{P}_B = 1$ (R)

Proportionalität: Für alle$A \subseteq \mathscr{F}$ 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 $\mathbb{P}_B$ (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 <u>bedingte Wahrscheinlichkeit von A gegeben B</u>. 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.

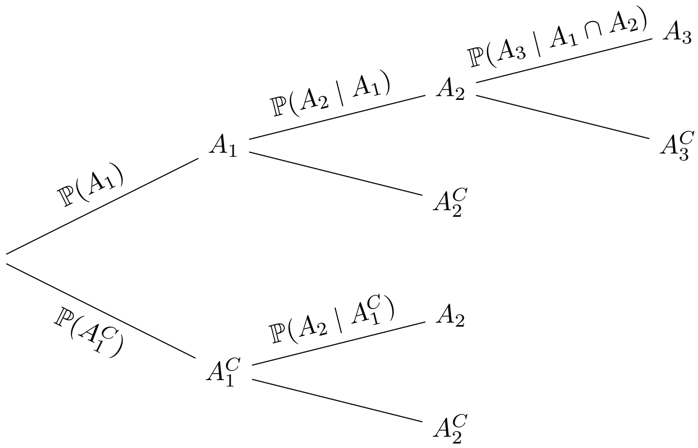


Abbildung III.1: Lemma III.5

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

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

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

$$W = \{\text{richtige Antwort gewusst}\}$$
$$R = \{\text{Richtige Antwort gewählt}\}$$

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:

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

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)

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

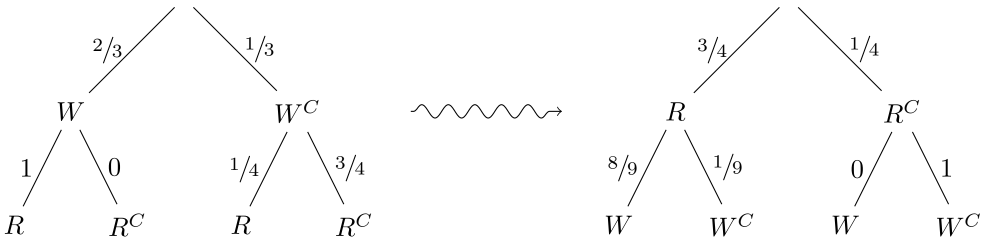


Abbildung III.2: Satz III.3

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 <u>nicht</u> 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 <u>unabhängig</u>, 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 <u>unabhängig</u>, 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 $I = \{1, \ldots, n\}$ 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

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

  □

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 $I = \{1, \dots, n\}$ 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 $F_i \in \mathscr{F}_i, i = 2, \dots, n$ und defniere für $F \in \sigma(\mathscr{F}_i)$ die endlichen Maße

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

  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

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

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

Also sind

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

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

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

unabhängig.

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

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

$\cap$-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 $X_1, \dots, X_n : (\Omega, \mathscr{F}) \to (E, \mathbb{E})$ Zufallsvariablen. Dann sind die folgenden Aussagen äquivalent:

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

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

Beweis. Per Ringschluss:

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

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

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 Satz I.I.9.

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

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

Dann

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

□

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

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

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

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.

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 $\mathbb{P}_1 * \mathbb{P}_2 = (\mathbb{P}_1 \otimes \mathbb{P}_2) \circ A^{-1}$ 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 $(\mathbb{Z}, \mathscr{B}(\mathbb{Z}))$ 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

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

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

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

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 $\rho(x) = {}^1\!/_{(b-a)}, a \le x \le b$ und Verteilungsfunktion

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

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 $\operatorname{negBin}(r,p)$. 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 $\operatorname{Geom}(p)$.

2.1. Exponential- und Gammaverteilung

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

   Die zweite Annahme macht Sinn, denn
   - Poissonverteilung ist Modell für Anzahl seltener Ereignisse
   - Nach Beispiel 3.26:

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

     Die Linearität des Parameters entspricht also einer Stationaritätsvorrausetzung:

Modelliert

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

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

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

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

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

$$f(x) = \frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x},$$

wobei

$$\Gamma(r) = \int_0^\infty y^{r-1}e^{-y}\,\mathrm{d}y, r > 0$$

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

Insbesondere ist $\text{Gamma}(\lambda, r)$ 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

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

□

Exponentialverteilungen sind zudem gedächtnislos:

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

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

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 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 $X \in \mathcal{L}^1(\mathbb{P})$.
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: $X \in \mathcal{L}^1(\mathbb{P})$ 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

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

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 ($\nearrow$ Schilling MINT 18.1). $\square$

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 $(\mathbb{Z}, \mathbb{P}(\mathbb{Z}))$ 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. $\square$

■ **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 $\mathcal{L}^1(\mathbb{P})$ und $a, b \in \mathbb{R}$ konstant.

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

Beispiel V.7
Da für $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ unabhängig gilt, dass

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

Satz V.8 (Produktformel für Erwartungswerte)
Seien $(\Omega, \mathscr{F}, \mathbb{P})$ Wahrscheinlichkeitsraum un $X_1, \dots, X_n : \Omega \to \mathbb{R}^d$ 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 $f_i(X_i) \in \mathcal{L}^1(\mathbb{P}), i = 1, \dots, n$, 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 $h(X,Y) \in \mathcal{L}^1(\mathbb{P})$, 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

$$\begin{aligned}
\mathbb{E}[h(X,Y)] &= \int_{\Omega} h(X(\omega), Y(\omega)) \, \mathrm{d}\mathbb{P}(\omega) \\
&\overset{V.3}{=} \int_{\mathbb{R}^d} h(x,y) \mathbb{P}(X \in \mathrm{d}x, Y \in \mathrm{d}y) \\
&\overset{X \perp\!\!\!\perp Y, III.21}{=} \int_{\mathbb{R}^{2d}} h(x,y) \mathbb{P}_X \otimes \mathbb{P}_Y(\mathrm{d}x, \mathrm{d}y) \\
&\overset{\text{Fubini}}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x,y) \mathbb{P}_X(\mathrm{d}x) \mathbb{P}_Y(\mathrm{d}y) \\
&\overset{V.3}{=} \int_{\Omega} \int_{\mathbb{R}^d} h(x, Y(\omega)) \mathbb{P}_X(\mathrm{d}x) \mathbb{P}(\omega) \\
&= \mathbb{E}[\int_{\mathbb{R}^d} h(x,Y) \mathbb{P}_X(\mathrm{d}x)].
\end{aligned}$$ □

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

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

also $f(X)g(Y) \in \mathcal{L}^1(\mathbb{P})$. 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 k-te Moment von X (sofern definiert).

▶ **Bemerkung**

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

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

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

Von Interesse ist insbesondere das zweite Moment.

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

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}X^2 - \underbrace{(\mathbb{E}X)^2}_{=\sim p}$$

mit

$$\begin{aligned}\mathbb{E}X^2 &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}\\ &= np\sum_{i=1}^{n} k \frac{(n-1)!}{(n-1-(k-1))!(k-1)!} p^{k-1}(1-p)^{n-1-(k-1)}\\ &= np\sum_{l=0}^{n-1}(l+1)\binom{n-1}{l} p^l (1-p)^{n-1-l}\\ &= np(1+\sum_{l=0}^{n-1} l\binom{n-1}{l} p^l(1-p)^{n-1-l})\\ &= np(1+(n-1)p)\cdot 1)\\ &= np + n(n-1)p^2\\ &\implies \mathbb{V}\mathrm{ar}X = np + n^2p^2 - np^2 - n^2p^2\\ &= np(1-p).\end{aligned}$$

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

1. $\mathbb{V}\mathrm{ar}(aX + b) = a^2 \mathbb{V}\mathrm{ar}(X)$
2. Sei

$$(\mathbb{C}\mathrm{ov}(X, Y))^2 \leq \mathbb{V}\mathrm{ar}X \mathbb{V}\mathrm{ar}Y$$

   und insbesondere

$$|\mathbb{C}\mathrm{orr}(X, Y)| \leq 1$$

3. $\mathbb{V}\mathrm{ar}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i) + \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i, X_j)$ Sind die $X_1, \ldots, X_n$ 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 $\varepsilon > 0$

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) \leq \frac{\mathbb{V}\mathrm{ar}X}{\varepsilon^2}$$

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

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

   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

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

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 $X \sim \mathrm{Bin}(n,p)$, so folgt

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

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

$$\mathbb{E}[X^n] < \infty \quad nge1, \text{ genau dann, wenn } \psi_X^{(n)} = \lim_{s \uparrow 1} \psi_X^{(n)}(s) < \infty.$$

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

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

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 $X \sim \operatorname{Bin}(n,p)$, 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 $\mathbb{P}(* \mid B)$ 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 bedingten Erwartungswert 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 $(X \in \mathcal{L}^1, \mathbb{P}(B) > 0)$

Frage: (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

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

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

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

  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 absolutstetig bzgl. μ $(\nu \ll \mu)$ genau dann wenn ν eine messbare Dichte f bzüglich 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 $\mathbb{P}_Y$-Nullmengen. 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 Wahrscheinlihckeitsraum $([0,1], \mathscr{B}([0,1]), \mathrm{U}([0,1])$ 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 geeignten $\mathbb{P}_Y$-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 $\tilde{\mathbb{P}}(X \in \cdot \mid Y = y)$ nichtnegativ 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. □

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.