Stochastik für Studierende der Informatik

Dr. Ernst August v. Hammerstein

Albert-Ludwigs-Universität Freiburg
Abteilung für Mathematische Stochastik

Sommersemester 2019

Organisatorisches

Übungsgruppen

Gruppe 1: Mi 08–10 Uhr, R 03 026 (Geb. 51)	(Saskia Glaffig)
Gruppe 2: Do 08–10 Uhr, R 00-031 (Geb. 51)	(Sebastian Stroppel)
Gruppe 3: Fr 08–10 Uhr, SR 01 009/13 (Geb. 101)	(Michaela Freitag)
Gruppe 4: Fr 14–16 Uhr, SR 01 016 (Geb. 101)	(Michaela Freitag)
Gruppe 5: Fr 14–16 Uhr, SR 01 018 (Geb. 101)	(Jasper Hoffmann)

Anmeldung: über **HISinOne**, möglichst bis Ende der ersten Vorlesung

In der zweiten Semesterwoche finden bereits Übungen statt, in denen Anwesenheitsaufgaben bearbeitet und besprochen werden.
Ersatztermin Gruppe 1: Di, 30.04., 08-10 Uhr, R 00-031 (Geb. 51)

Für die an Christi Himmelfahrt und Fronleichnam ausfallenden Übungen wird es Ersatztermine geben.

Übungsaufgaben, Studien- und Prüfungsleistung

Übungsaufgaben und ggf. weitere Vorlesungsmaterialien werden über ILIAS zur Verfügung gestellt.
Zugangspasswort für ILIAS: #St19IvH2t!

Neue Übungsblätter werden montags auf ILIAS hochgeladen,
Lösungen sind jeweils bis spätestens am Montag der Folgewoche **vor der Vorlesung** in die Briefkästen im EG Geb. 51 einzuwerfen.
Sie dürfen maximal zu zweit abgeben, dabei sollten die Abgabepartner sich in derselben Übungsgruppe befinden.

Voraussetzungen für die Zuerkennung der Studienleistung:

- aktive Teilnahme an den Übungsgruppen
- Vorrechnen mindestens einer Übungsaufgabe an der Tafel
- Mindestens **50%** der erreichbaren Punkte der Übungsaufgaben

Die **Prüfungsleistung** besteht in der erfolgreichen Teilnahme an der Abschlussklausur. Der genaue Termin hierfür steht noch nicht fest.

Literatur (Auswahl)

- **Dümbgen, L.** (2003), *Stochastik für Informatiker*, Springer
- **Henze, N.** (2017), *Stochastik für Einsteiger*, 11. Aufl., Springer
- **Kersting, G., Wakolbinger, A.** (2010), *Elementare Stochastik*, 2. Aufl., Birkhäuser

Kontakt

Fragen, Anregungen, Kritik (positive oder negative) können Sie neben den Übungsgruppen auch richten an

ernst.august.hammerstein@stochastik.uni-freiburg.de
pascal.beckedorf@stochastik.uni-freiburg.de

oder persönlich in den Sprechstunden:

E.A. v. Hammerstein: Mi, 10–11 Uhr, Raum 248, Ernst-Zermelo-Str. 1
P. Beckedorf: Di, 14–15 Uhr, Raum 227, Ernst-Zermelo-Str. 1

Zur Einführung

Was ist Stochastik?

Stochastik ist der Oberbegriff für Wahrscheinlichkeitsrechnung bzw. -Theorie und mathematische Statistik.

In der Stochastik werden mathematische Modelle von Zufallserscheinungen konstruiert, deren Gesetzmäßigkeiten studiert und ihre Anwendbarkeit auf reale Daten untersucht. Die Modelle basieren auf Zufallsbegriffen, wie z.B. dem der „Wahrscheinlichkeit".

Diese werden durch mathematische Axiome beschrieben (Kolmogorov 1933). Die Axiome erklären jedoch nicht das Wesen des Zufalls.

$$\begin{tikzcd} \boxed{\text{Modell}} \arrow[r, shift left, "\text{Wahrscheinlichkeitstheorie}"] & \boxed{\text{Daten}} \arrow[l, shift left, "\text{mathematische Statistik}"] \end{tikzcd}$$

Stochastik im Alltag

Entscheiden Auswahl von Kapitalanlagemöglichkeiten
Spiele, z.B. Schere-Stein-Papier
($\rightarrow$ Spieltheorie in den Wirtschaftswissenschaften)

Schätzen jährliches Steueraufkommen, Inflationsraten
Krankheitsaufkommen in der Bevölkerung (Inzidenzrate)

Vergleichen/Testen Ist eine Münze oder ein Würfel fair?
Ist ein Medikament besser/wirksamer als ein anderes?
Sind zwei Merkmale unabhängig oder korreliert?

Vorhersagen Wetter
Tippen: Toto, Lotto,...
Zukünftiger Kurs eines Wertpapiers

Messen physikalischer Größen (Messfehler $\rightarrow$ Fehlerausgleichsrechnung, Fehlergrenzen)
Quantenmechanik, Heisenbergsche Unschärferelation

Stochastik im Alltag (Forts.)

Mustererkennung, Fehlerkorrektur Stochastische Algorithmen zur Signalentstörung (Funk, Radar, DVD-Player)
Bildverschärfung
Gesichtserkennung in Fotoprogrammen

Verschlüsselungsverfahren stochastische Primzahltests

Analyse von Netzwerken und Algorithmen Graphentheorie, Netzwerkmodelle, Suchbäume, Quicksort

Versicherungsmathematik Prämienkalkulation in z.B. Haftpflicht-, Kranken- und Lebensversicherung (Aktuare)

Finanzmathematik Berechnung von Derivatpreisen (Optionen, Zertifikate, Swaps,. . .)
Optimale Handelsstrategien und Portfolios
Quantifizierung von Risiken (Markt-, Kredit-, Liquiditätsrisiken sowie operational risk)
Risikokapitalberechnung (Basel II, Basel III)

Genereller Ansatz und Verfahren

- Präzisiere, welche Ereignisse man betrachten will $\longrightarrow$ Modellbildung
- Ordne jedem Ereignis A eine Wahrscheinlichkeit $\mathbb{P}(A) \in [0, 1]$ zu. Mögliche Prinzipien zur Festsetzung von $\mathbb{P}(A)$:
  1. **subjektiv:** Maß des persönlichen Glaubens, dass A eintritt
  2. **frequentistisch:** Relative Häufigkeit bzw. deren Grenzwert bei beliebig vielen unabhängigen Wiederholungen
  3. **Gleichverteilung:** Quotient aus Anzahl der günstigen durch Anzahl der möglichen Fälle

Problem: Alle o.g. Festsetzungsmöglichkeiten für $\mathbb{P}(A)$ führen zu Schwierigkeiten

Ausweg: Axiomatischer Ansatz (Kolmogorov 1933)

Keine Hinterfragung der genauen Bedeutung von $\mathbb{P}(A)$ bzw. dessen Erhalt durch ein konkretes Zufallsexperiment.
Fordere lediglich gewisse (konsistente) Regeln, die für Wahrscheinlichkeiten $\mathbb{P}(A)$ gelten sollen.
Leite daraus Wahrscheinlichkeiten komplexerer Ereignisse ab.

3 Diskrete Wahrscheinlichkeitsräume

Definition 3.1 (Diskreter Wahrscheinlichkeitsraum)

Ein **diskreter Wahrscheinlichkeitsraum** *ist ein Tripel* $(\Omega, \mathcal{A}, \mathbb{P})$, *bestehend aus einer nicht-leeren, höchstens abzählbaren Menge* Ω *(Grundraum), der Potenzmenge* $\mathcal{A} = \mathcal{P}(\Omega)$ *und einer Abbildung* $\mathbb{P} : \mathcal{A} \to [0, 1]$ *(Wahrscheinlichkeitsmaß oder -Verteilung), die die folgenden Eigenschaften erfüllen muss:*

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

b) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ für jede Folge $(A_i)_{i \geq 1}$ paarweise disjunkter Mengen $A_i \in \mathcal{A}$ (d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$) $\quad$ (σ-Additivität).

Sprechweisen:

- Teilmengen $A \in \mathcal{P}(\Omega)$ von Ω heißen *Ereignisse*,
- $A = \Omega$ heißt *sicheres Ereignis*,
- $A = \emptyset$ heißt *unmögliches Ereignis*,
- $A = \{\omega\}$ mit $\omega \in \Omega$ heißt *Elementarereignis*,

- $A \cup B$ bedeutet, dass A oder B (oder beide) eintreten,
- $A \cap B$ bedeutet, dass sowohl A als auch B eintreten.

Einfache Folgerungen:

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

b) $\mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ für paarweise disjunkte Mengen $A_1, \ldots, A_n \in \mathcal{A}$ (endliche Additivität)
$\implies \mathbb{P}(A^C) = 1 - \mathbb{P}(A)$ und $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(B \cap A)$,

c) $A, B \in \mathcal{A}$, $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$ (Monotonie),

d) Seien $A_1, \ldots, A_n \in \mathcal{A}$, so gilt die sog. *Sieb- oder Einschluss-Ausschluss-Formel*

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \cdots \pm \mathbb{P}(A_1 \cap \cdots \cap A_n)$$

Speziell: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$,

e) $A_1, A_2, \cdots \in \mathcal{A} \Rightarrow \mathbb{P}(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mathbb{P}(A_i)$ (Sub-σ-Additivität).

Satz 3.2 (Eindeutige Festlegung von $\mathbb{P}$)

Sei Ω eine nicht-leere, höchstens abzählbare Menge und $(p_n)_{n\geq 1}$ eine Folge nicht-negativer Zahlen (d.h. $p_n \geq 0$ für alle n), für die gilt $\sum_{n=1}^{\infty} p_n = 1$.
Dann gibt es auf dem Grundraum Ω genau ein Wahrscheinlichkeitsmaß $\mathbb{P} : \mathcal{A} \to [0,1]$ mit $\mathbb{P}(\{\omega_n\}) = p_n$, d.h. das Wahrscheinlichkeitsmaß $\mathbb{P}$ ist durch Angabe der **Elementarwahrscheinlichkeiten** *$\mathbb{P}(\{\omega_n\})$, $n \geq 1$, bereits eindeutig festgelegt.*

Beweis: Anwesenheitsaufgabe

Einfache Beispiele: Würfel und Münzwurf

Einmaliger Würfelwurf: Grundraum der möglichen Ergebnisse $\Omega = \{1,2,3,4,5,6\}$

Bei einem fairen Würfel sollten alle möglichen Ergebnisse gleich wahrscheinlich sein

Laplace-Ansatz: *Bei endlichem Grundraum Ω sollen alle Elementarereignisse die gleiche Wahrscheinlichkeit haben.*

Im Fall des Würfels bedeutet das

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{i=1}^{6}\{i\}\right) = \sum_{i=1}^{6}\mathbb{P}(\{i\}) \overset{\text{Laplace}}{=} \sum_{i=1}^{6} p = 6p, \text{ d.h. } p = \mathbb{P}(\{i\}) = \frac{1}{6}.$$

Sei A_g das Ereignis, eine gerade Zahl zu würfeln, dann ist $A_g = \{2,4,6\}$ und

$$\mathbb{P}(A_g) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

Allgemein gilt unter dem Laplace Ansatz: Ist $\Omega = \{\omega_1, \ldots, \omega_n\}$ und somit $|\Omega| = \#\Omega = n$, dann ist $\mathbb{P}(\{\omega_k\}) = \frac{1}{|\Omega|} = \frac{1}{n}$ für $1 \le k \le n$, und für $A \subseteq \Omega$ gilt

$$\mathbb{P}(A) = \sum_{\omega_k \in A}\mathbb{P}(\{\omega_k\}) = \sum_{\omega_k \in A}\frac{1}{|\Omega|} = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl günstiger Fälle}}{\text{Anzahl möglicher Fälle}}.$$

Anderes Modell: *Manipulierter Würfel*

Bei diesem fällt die 6 mit Wahrscheinlichkeit $\mathbb{P}(\{6\}) = p$ für ein $0 < p < 1$ und die anderen Zahlen mit $\mathbb{P}(\{i\}) = \frac{1-p}{5}$, $1 \le i \le 5$.

Zweimaliger Würfelwurf:
$\Omega = \{(1,1),(1,2),\ldots,(6,6)\} = \{(\omega_1,\omega_2) \mid \omega_1,\omega_2 \in \{1,2,3,4,5,6\}\}$
Hier ist offensichtlich $|\Omega| = 6 \cdot 6 = 36$. Damit ergibt sich mit dem Laplace-Ansatz z.B. für die Wahrscheinlichkeit eines Paschs

$$\mathbb{P}(\{(1,1),(2,2),\ldots,(6,6)\}) = \sum_{i=1}^{6} \mathbb{P}(\{(i,i)\}) = \sum_{i=1}^{6} \frac{1}{36} = \frac{1}{6}.$$

n-maliger Münzwurf:
Mögliche Ergebnisse bei einmaligem Wurf: Kopf $= 1$, Zahl $= 0$
Grundraum: $\Omega = \{(\omega_1,\ldots,\omega_n) \mid \omega_i \in \{0,1\},\ 1 \leq i \leq n\} \implies |\Omega| = 2^n$
Laplace: $\mathbb{P}(\{(\omega_1,\ldots,\omega_n)\}) = \frac{1}{2^n}$

Allgemeiner: p-Münze, d.h. $\mathbb{P}(\{1\}) = p$, $0 < p < 1$, $\mathbb{P}(\{0\}) = 1 - p$.

Für das n-malige Würfeln mit einem manipulierten Würfel bzw. das n-malige Werfen einer p-Münze sind die Grundräume Ω dieselben wie zuvor, jedoch ist hier die Laplace-Annahme zur Festsetzung der Wahrscheinlichkeiten offensichtlich nicht gerechtfertigt. Hierzu benötigt man andere Annahmen (z.B. Unabhängigkeit).

Ziehen aus einer Urne

Eine Urne enthalte m weiße und n schwarze Kugeln. Wie groß ist die Wahrscheinlichkeit, eine weiße Kugel zu ziehen?

Grundraum: $\Omega = \{1, \ldots, n+m\}$, wobei $W = \{1, \ldots, m\}$ die weißen und $S = \{m+1, \ldots, n+m\}$ die schwarzen Kugeln seien.
Mit Laplace-Annahme gilt $\mathbb{P}(W) = \frac{|W|}{|\Omega|} = \frac{m}{m+n}$ und analog $\mathbb{P}(S) = \frac{n}{n+m}$.

Geburtstagsproblem

In einem Raum befinden sich N Personen. Mit wecher Wahrscheinlichkeit haben mindestens zwei der Anwesenden am gleichen Tag Geburtstag?

Grundraum: $\Omega = \{(\omega_1, \ldots, \omega_N) \mid \omega_i \in \{1, 2, \ldots, 365\}\} \implies |\Omega| = 365^N$

$$\begin{aligned} A &= \{\text{Mind. 2 Personen haben am gleichen Tag Geburstag}\} \\ &= \{\omega \in \Omega \mid \exists\, i \neq j : \omega_i = \omega_j\} \end{aligned}$$

Unter der Laplace-Annahme gilt $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$.
Problem: Bestimmung von $|A|$ kompliziert!

Ausweg: Nutze $\mathbb{P}(A) = 1 - \mathbb{P}(A^C) = 1 - \frac{|A^C|}{|\Omega|}$.

$$\text{Es ist } \ |A^C| = \begin{cases} \prod_{i=1}^{N}(366 - i), & N \leq 365, \\ 0, & N > 365. \end{cases}$$

Damit erhält man

N	10	20	23	30	40	50	60	100
$\mathbb{P}(A)$	0.1169	0.4114	0.5073	0.7063	0.8912	0.9704	0.9941	0.9999997

(Un)Faire Wetten

Man wettet mit einem Einsatz von 1€ auf ein Ereignis E, das mit Wahrscheinlichkeit p $(0 < p < 1)$ eintritt.
Prinzip des fairen Wettens: Der (Netto-)Gewinn G muss gerade so groß sein, dass man im Mittel nichts gewinnt, d.h. der mittlere Gewinn MG is gleich Null:

$$MG = G \cdot \mathbb{P}(E) - 1 \cdot \mathbb{P}(E^C) = G \cdot p - 1 \cdot (1-p) \stackrel{!}{=} 0 \implies G = \frac{1}{p} - 1$$

Auflösen nach p ergibt $p = \frac{1}{G+1}$, d.h. bei einer fairen Wette ist die Erfolgswahrscheinlichkeit der Kehrwert der Wettquote $(G+1):1$.

Solche Wettquoten werden u.a. bei Sportwetten (z.B. b-win, tipico) angegeben, z.B. für das Spiel SC Freiburg gegen Fortuna Düsseldorf

	Heimsieg	Unentschieden	Auswärtssieg
Freiburg vs Düsseldorf	2.1	3.6	3.4

Aus den obigen Quoten erhält man

$$\mathbb{P}(\text{Sieg Freiburg}) = \frac{1}{2.1} = \frac{10}{21} \approx 0.4762,\ \mathbb{P}(\text{Sieg Ddorf}) = \frac{1}{3.4} = \frac{5}{17} \approx 0.2941,$$
$$\mathbb{P}(\text{Unentsch.}) = \frac{1}{3.6} = \frac{5}{18} \approx 0.2778,$$

$$\text{Wegen } \mathbb{P}(\text{Sieg F}) + \mathbb{P}(\text{U}) + \mathbb{P}(\text{Sieg D}) = \frac{10}{21} + \frac{5}{18} + \frac{5}{17} \approx 1.048086 > 1$$

ist das Spiel nicht ganz fair (da der Wettanbieter Gewinn machen will). Die tatsächlichen (vermuteten) Wahrscheinlichkeiten erhält man hier durch Renormierung, z.B. $\mathbb{P}(\text{Sieg Freiburg}) = \frac{0.4762}{1.048086} \approx 0.4544$.

Urnenmodelle

Zur Anwendung des Laplace-Ansatzes muss man insbesondere die Mächtigkeit des Grundraumes Ω bestimmen, wozu häufig kombinatorische Überlegungen notwendig sind.

Fall 1: Anordnungen von m aus N Elementen mit Wiederholung

$A = \{a_1, \ldots, a_N\}$ sei endliche Menge mit unterscheidbaren Objekten

$\Omega := \{\omega = (\omega_1, \ldots, \omega_m) \,|\, \omega_i \in A, 1 \leq i \leq m\} = A^m \quad \Longrightarrow |\Omega| = N^m$

Äquivalentes Urnenmodell: m-faches Ziehen mit Zurücklegen und Beachtung der Reihenfolge aus einer Urne mit N unterscheidbaren Kugeln

Für $A = \{1, 2, \ldots, 6\}$ entspricht dies dem m-maligen Würfeln.

Weitere Anwendung: Belegung von Zellen/Schachteln mit unterscheidbaren Objekten

Beispiel: Für m verschiedene Elementarteilchen stehen N verschiedene Energiezustände zur Auswahl. Belegung der Energiezustände wird beschrieben durch $(\omega_1, \ldots, \omega_m)$, wobei ω_i den Energiezustand von Teilchen i angibt (mehrere Teilchen können denselben Energiezustand haben, Modell ohne Pauli-Prinzip)

Verallgemeinerung: m verschiedene Mengen A_i (Ziehen aus verschieden gefüllten Urnen)

$$\Omega := \{\omega = (\omega_1, \ldots, \omega_m) \,|\, \omega_i \in A_i,\, 1 \leq i \leq m\}, \quad |\Omega| = \prod_{i=1}^{m} |A_i|$$

Fall 2: Anordnungen von m aus N Elementen ohne Wiederholung

$A = \{a_1, \ldots, a_N\}$ sei endliche Menge mit unterscheidbaren Objekten

$$\Omega := \{\omega = (\omega_1, \ldots, \omega_m) \,|\, \omega_i \in A \text{ und } \omega_i \neq \omega_j \text{ für } i \neq j\}$$

$$\Longrightarrow |\Omega| = \prod_{i=1}^{m} (N - i + 1) = \frac{N!}{(N-m)!}, \text{ wobei } n! := n \cdot (n-1) \cdot \ldots \cdot 1$$

Äquivalentes Urnenmodell: m-faches Ziehen ohne Zurücklegen, aber mit Beachtung der Reihenfolge aus einer Urne mit N unterscheidbaren Kugeln

Spezialfall Permutation ($m = N$): Als *Permutation* bezeichnen wir eine bijektive Abbildung einer Menge A auf sich selbst. Bei einer endlichen Menge $A = \{a_1, \ldots, a_N\}$ entspricht eine Permutation einer Umordnung der Elemente von A.

Die Menge aller möglichen Permutationen von A ist damit

$$\Omega := \{\omega = (\omega_1, \ldots, \omega_N) \,|\, \omega_i \in A, \, \omega_i \neq \omega_j\} \quad \Longrightarrow |\Omega| = N!$$

Nach Obigem muss ferner $|\Omega| = N! = \frac{N!}{0!}$ sein, d.h. man setzt $0! := 1$.

Fall 3: Kombinationen von m aus N Elementen ohne Wiederholung

Äquivalentes Urnenmodell: m-faches Ziehen ohne Zurücklegen und ohne Beachtung der Reihenfolge aus einer Urne mit N unterscheidbaren Kugeln

Beispiel: Zahlenlotto „6 aus 49“. Hier kommt es nur darauf an, welche Zahlen gezogen werden, aber nicht auf die Ziehungsreihenfolge (Zahlen werden später aufsteigend sortiert).

Satz 4.1

Sei $A = \{a_1, \ldots, a_N\}$ eine endliche Menge unterscheidbarer Objekte. Dann gibt es

$$\frac{N!}{m!(N-m)!} =: \binom{N}{m}$$

verschiedene m-elementige Teilmengen von A ($m \leq N$). Man bezeichnet die (ungeordneten) Teilmengen als **Kombinationen (ohne Wiederholungen) der Größe m aus N Elementen**.

Beweis: Wählt man nacheinander m Elemente aus A unter Berücksichtigung der Reihenfolge aus, erhält man als mögliche Ergebnisse die Menge

$$\Omega := \{\omega = (\omega_1, \ldots, \omega_m) \in A^m \,|\, \omega_i \neq \omega_j \text{ für } i \neq j\}$$

mit $|\Omega| = \frac{N!}{(N-m)!}$ (vgl. Fall 2). Da es auf die Reihenfolge/Ordnung *nicht* ankommen soll, sind sämtliche möglichen Anordnungen/Permutationen derselben m gezogenen Objekte als äquivalent anzusehen.

Für ein m-Tupel $(\omega_1, \ldots, \omega_m)$ gibt es $m!$ verschiedene Anordnungen. Zusammenfassung aller Tupel, die die gleichen Objekte in unterschiedlicher Anordnung enthalten („Äquivalenzklassen"), ergibt $\frac{|\Omega|}{m!} = \frac{N!}{(N-m)!m!}$.

Für jede Äquivalenzklasse kann man einen *Repräsentanten*, d.h. ein Tupel mit spezieller Anordnung, auswählen, z.B. mit aufsteigender Reihenfolge (sofern auf A eine Ordnungsrelation existiert). Damit lässt sich die Menge Ω_m der m-elementigen Teilmengen von A darstellen als

$$\Omega_m = \{\omega \in A^m \,|\, \omega_1 < \omega_2 < \cdots < \omega_m\}, \; |\Omega_m| = \frac{N!}{m!(N-m)!} = \binom{N}{m}. \; \square$$

Bemerkung 4.2

Die Zahlen $\binom{n}{k} := \frac{n!}{k!(n-k)!}$, $0 \leq k \leq n$, heißen **Binomialkoeffizienten**.
Man setzt $\binom{n}{k} = 0$ für $k > n$.

Anwendungen:

a) Zahlenlotto „6 aus 49“. Es gibt $\binom{49}{6} = 13983816$ verschiedene mögliche Ziehungsergebnisse (ohne Superzahl).

b) Belegung vom N verschiedenen Energiezuständen durch m Elementarteilchen, wobei jeder Energiezustand nur von höchstens einem Teilchen angenommen werden darf (Pauli-Prinzip, gilt z.B. für Elektronen, Protonen und Neutronen).

Bemerkung 4.3

Äquivalent zu „unterscheidbare Objekte ohne Anordnung“ (d.h. eine Auswahl verschiedener Objekte, bei der es nicht auf die Reihenfolge ankommt) ist, *ununterscheidbare (gleiche) Objekte* zu betrachten, die man mangels charakteristischer Eigenheiten nicht auseinander halten und somit auch nicht anordnen kann.
(Dies ist z.B. bei Anwendung b) oben der Fall.)

Fall 4: Kombinationen von m aus N Elementen mit Wiederholung

Äquivalentes Urnenmodell: m-faches Ziehen mit Zurücklegen und ohne Beachtung der Reihenfolge aus einer Urne mit N unterscheidbaren Kugeln

$A = \{a_1, \dots, a_N\}$ sei endliche Menge mit unterscheidbaren Objekten (o.B.d.A. mit Ordnungsrelation)

Analog zum Beweis von Satz 4.1 lässt sich die Menge $\bar{\Omega}_m$ aller m-elementigen Auswahlen (Ziehungen) aus A mit Wiederholung, aber ohne Berücksichtigung der Reihenfolge, darstellen als

$$\bar{\Omega}_m = \{\omega \in A^m \,|\, \omega_1 \leq \omega_2 \leq \cdots \leq \omega_m\}$$

($\leq$, da nun Wiederholungen möglich sind). Es gilt $|\bar{\Omega}_m| = \binom{N+m-1}{m}$.

Beweis: O.B.d.A. sei $A = \{1, \dots, N\}$. Definiere ferner $\bar{A} := \{1, \dots, N+m-1\}$ sowie $\mathcal{P}_m(\bar{A}) = \{\bar{B} \subset \bar{A} \,|\, |\bar{B}| = m\}$ (m-elementige Teilmengen von $\bar{A}$) und

$f : \bar{\Omega}_m \to \mathcal{P}_m(\bar{A})$ durch $(\omega_1, \dots, \omega_m) \mapsto (\omega_1, \omega_2 + 1, \dots, \omega_m + m - 1)$.

Beachte, dass f bijektiv ist (die Injektivität ist trivial)!

f ist surjektiv, denn zu jeder m-elementigen Teilmenge $\bar{B}$ von $\bar{A}$ erhält man ein Urbild aus $\bar{\Omega}_m$, indem man die Elemente von $\bar{B}$ zunächst nach aufsteigender Größe ordnet und dann vom i-ten Glied der geordneten Folge $i-1$ subtrahiert.

Da f bijektiv ist, muss gelten $|\bar{\Omega}_m| = |\mathcal{P}_m(\bar{A})|$.

Da $\bar{A} = \{1, \ldots, N+m-1\}$ $N+m-1$ verschiedene Elemente hat, ist die Anzahl der m-elementigen Teilmengen von $\bar{A}$ nach Satz 4.1 gerade $\binom{N+m-1}{m}$, d.h. $|\bar{\Omega}_m| = |\mathcal{P}_m(\bar{A})| = \binom{N+m-1}{m}$. $\square$

Bemerkung: Die Funktion f oben stellt eine 1:1-Beziehung zwischen Ziehen mit und ohne Zurücklegen (jeweils ohne Reihenfolge) her:
Ziehen von m aus N Elementen mit Zurücklegen ohne Reihenfolge (Elemente von $\bar{\Omega}_m$) ist (vermöge der Abbildungsvorschrift f) äquivalent zu Ziehen von m aus $N+m-1$ Elementen ohne Zurücklegen und ohne Reihenfolge, d.h. Bildung einer m-elementigen Teilmenge aus $\bar{A}$ (Elemente von $\mathcal{P}_m(\bar{A})$)

Korollar 4.4 (Binomischer Lehrsatz)

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}, \qquad x, y \in \mathbb{R},\ n \geq 1.$$

Beweis:
$$(x+y)^n = (x+y) \cdot \ldots \cdot (x+y) = \sum_{A \subset \{1,\ldots,n\}} x^{|A|} y^{|A^c|}$$
$$= \sum_{k=0}^{n} \sum_{\substack{A \subset \{1,\ldots,n\} \\ |A|=k}} x^k y^{n-k} \overset{\text{Satz 4.1}}{=} \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \qquad \square$$

Anwendung: Ist $|\Omega| = n$, so gilt $|\mathcal{P}(\Omega)| = 2^n$.

Beweis: Sei $\mathcal{P}_k(\Omega) := \{A \subset \Omega \,|\, |A| = k\}$ die Menge der k-elementigen Teilmengen von Ω für $k = 0, 1, \ldots, n$. Dann gilt

$$|\mathcal{P}(\Omega)| = \sum_{k=0}^{n} |\mathcal{P}_k(\Omega)| \overset{\text{Satz 4.1}}{=} \sum_{k=0}^{n} \binom{n}{k} \overset{\text{Kor. 4.4}}{=} (1+1)^n = 2^n.$$

Zusammenfassung

Anordnungen/Kombinationen von m aus N Elementen

Statistische Physik	**unterscheidbare Objekte**	**ununterscheidbare (gleiche) Objekte**	
ohne Pauliprinzip	N^m „Maxwell-Boltzmann“	$\binom{N+m-1}{m}$ „Bose-Einstein“	mit Zurücklegen
mit Pauliprinzip	$\frac{N!}{(N-m)!}$	$\binom{N}{m}$ „Fermi-Dirac“	ohne Zurücklegen
	geordnete Stichproben	**ungeordnete Stichproben**	**Ziehen aus einer Urne**

Beispiel 4.5 (Fixpunkte in Permutationen)

Sei $\mathcal{S}_n$ die Menge aller Permutationen der Zahlen $\{1, \dots, n\}$. Eine Permutation $\sigma \in \mathcal{S}_n$ hat einen *Fixpunkt*, falls es ein $i \in \{1, \dots, n\}$ gibt mit $\sigma(i) = i$. Wie wahrscheinlich ist es, dass eine (zufällig ausgewählte) Permutation $\tau \in \mathcal{S}_n$ keinen Fixpunkt besitzt?

Sei $A_i = \{\sigma \in \mathcal{S}_n \mid \sigma(i) = i\}$ die Menge der Permutationen, die (mindestens) die Zahl i als Fixpunkt haben.
Dann gilt (unter der Laplace-Annahme)

$$\mathbb{P}(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

denn wenn i ein Fixpunkt ist, können effektiv nur noch $n-1$ Zahlen permutiert werden, so dass $|A_i| = (n-1)!$.
Analog ergibt sich für die Wahrscheinlichkeit, dass eine Permutation k Fixpunkte $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ hat

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \frac{(n-k)!}{n!} = \frac{1}{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}.$$

Die Wahrscheinlichkeit, dass die Permutation τ mindestens einen Fixpunkt hat, ist daher $\mathbb{P}(A_1 \cup \cdots \cup A_n)$.

Nach Siebformel und Satz 4.1 erhält man diese Wahrscheinlichkeit durch (beachte: $\mathbb{P}(A_i) = \mathbb{P}(A_1)\ \forall\, i$, ebenso $\mathbb{P}(A_{i_1} \cap A_{i_2}) = \mathbb{P}(A_1 \cap A_2)$ usw.)

$$\begin{aligned}\mathbb{P}(A_1 \cup \cdots \cup A_n) &= n\mathbb{P}(A_1) - \binom{n}{2}\mathbb{P}(A_1 \cap A_2) + \binom{n}{3}\mathbb{P}(A_1 \cap A_2 \cap A_3) \\ &\quad - \cdots \pm \binom{n}{n}\mathbb{P}(A_1 \cap \cdots \cap A_n) \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!}\end{aligned}$$

Damit ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Permutation τ keinen Fixpunkt besitzt, gegeben durch

$$\mathbb{P}\big((A_1 \cup \cdots \cup A_n)^C\big) = 1 - \mathbb{P}(A_1 \cup \cdots \cup A_n) = \frac{1}{2!} - \frac{1}{3!} + \cdots \mp \frac{1}{n!}$$

Für $n \to \infty$ erhält man

$$\mathbb{P}\big((A_1 \cup \cdots \cup A_n)^C\big) = \sum_{k=2}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \xrightarrow[n\to\infty]{} e^{-1}.$$

Hypergeometrische und Binomialverteilung

Betrachte Urne mit s schwarzen und w weißen Kugeln, $s + w = n$. Ziehe m Kugeln mit einem Griff heraus. Wie groß ist die Wahrscheinlichkeit, genau $k \leq \min(m, s)$ schwarze Kugeln zu ziehen?

Seien $A = \{1, \dots, n\}$ die Menge aller Kugeln in der Urne, $A_0 = \{1, \dots, s\}$ die schwarzen und $A_0^C = \{s + 1, \dots, n\}$ die weißen Kugeln. Da es nur auf die Anzahlen der gezogenen schwarzen bzw. weißen Kugeln ankommt, aber nicht auf deren Reihenfolge innerhalb der Ziehung, liegt eine Ziehung ohne Anordnung und ohne Wiederholung vor bzw. eine Kombination von m aus n Elementen ohne Wiederholung.

$$\Omega = \{(\omega_1, \dots, \omega_m) \,|\, \omega_i \in A,\, \omega_1 < \omega_2 < \dots < \omega_m\}, \quad |\Omega| = \binom{n}{m}.$$

$$\begin{aligned} E &:= \text{„genau } k \text{ schwarze unter den } m \text{ gezogenen Kugeln"} \\ &= \{\omega \in \Omega \,|\, \omega_i \in A_0,\, 1 \leq i \leq k,\, \omega_i \in A_0^C,\, i > k\} \end{aligned}$$

Betrachte

$$\Omega' := \{\omega' = (\omega'_1, \dots, \omega'_k) \,|\, \omega'_i \in A_0,\, \omega'_1 < \cdots < \omega'_k\}, \quad |\Omega'| = \binom{s}{k},$$

$$\Omega'' := \{\omega'' = (\omega''_1, \dots, \omega''_{m-k}) \,|\, \omega''_i \in A_0^C,\, \omega''_1 < \cdots < \omega''_{m-k}\}, \quad |\Omega'| = \binom{w}{m-k}.$$

Definiere

$$\varphi : E \to \Omega' \times \Omega'',\ \varphi((\omega_1, \dots, \omega_m)) = \big((\omega_1, \dots, \omega_k), (\omega_{k+1}, \dots, \omega_m)\big)$$

Offensichtlich ist φ bijektiv, daher gilt $|E| = |\Omega'| \cdot |\Omega''| = \binom{s}{k}\binom{w}{m-k}$.

Unter der Laplace-Annahme gilt

$$\mathbb{P}(E) = \frac{|E|}{|\Omega|} = \frac{\binom{s}{k}\binom{w}{m-k}}{\binom{n}{m}}.$$

Definition 5.1 (Hypergeometrische Verteilung)

Die Wahrscheinlichkeitsverteilung $\mathbb{P}$ *auf* $\{\max(0, m-w), \dots, \min(m, s)\}$, *gegeben durch*

$$\mathbb{P}(\{k\}) = \frac{\binom{s}{k}\binom{w}{m-k}}{\binom{n}{m}} = \frac{\binom{s}{k}\binom{n-s}{m-k}}{\binom{n}{m}},$$

heißt **hypergeometrische Verteilung** *zu den Parametern n, s und m.*

Bemerkung: Die Verteilung ist auch auf der u.U. größeren Menge $\{0, \dots, m\}$ definiert, da für $k < m - w$ und $k > s$ jeweils ein Binomialkoeffizient im Zähler 0 wird.

Anwendungen:

a) Lotto „6 aus 49“: $n = 49$ Kugeln, $s = 6$ schwarze (Richtige, d.h. zuvor getippte Zahlen), $m = 6$ Kugeln werden gezogen, $k = 0, 1, \dots, 6$. p_k ist die Wahrscheinlichkeit, dass genau k der getippten Zahlen gezogen werden: $p_k = \frac{\binom{6}{k}\binom{43}{6-k}}{\binom{49}{6}}$

k	0	1	2	3	4	5	6
p_k	0.4359	0.4130	0.1324	0.01765	$0.9686 \cdot 10^{-3}$	$0.1845 \cdot 10^{-4}$	$0.715 \cdot 10^{-7}$

b) Qualitätskontrolle: n Werkstücke, s defekt, $w = n - s$ ok. Für Stichprobe der Größe m kann mit hypergeom. Vert. die Wahrscheinlichkeit berechnet werden, dass Stichprobe genau k defekte Stücke enthält.

Was passiert, wenn der Gesamtumfang n der Urne immer größer wird ($n \to \infty$), dabei aber der relative Anteil der schwarzen Kugeln $\frac{s_n}{n}$ nahezu konstant bleibt bzw. gegen ein festes Verhältnis strebt ($\frac{s_n}{n} \to p$)?

Satz 5.2

Sei $m \in \mathbb{N}$ beliebig, aber fest gewählt. Gilt $\frac{s_n}{n} \to p$ für $n \to \infty$ und $0 < p < 1$, so folgt für $0 \leq k \leq m$, $k \in \mathbb{N}$,

$$\frac{\binom{s_n}{k}\binom{n-s_n}{m-k}}{\binom{n}{m}} \xrightarrow[n\to\infty]{} \binom{m}{k} p^k (1-p)^{m-k}.$$

Interpretation: Ist n (und damit auch s_n) groß gegenüber m, besteht nahezu kein Unterschied zwischen Ziehen mit und ohne Zurücklegen. $p \approx \frac{s_n}{n}$ ist dann (nach Laplace-Annahme) die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen. Die rechte Seite entspricht somit der Wahrscheinlichkeit, bei m Ziehungen von einer Kugel aus der Urne mit jeweils anschließendem Zurücklegen genau k schwarze Kugeln zu erhalten.

Definition 5.3 (Binomialverteilung)

Sei $n \geq 1$ und $0 \leq p \leq 1$. Die auf $\Omega = \{0, 1, \ldots, n\}$ durch

$$p_k = b_{n,p}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}$$

definierte Wahrscheinlichkeitsverteilung heißt **Binomialverteilung zu den Parametern** n **und** p*. Sie wird oft auch mit $B(n, p)$ bezeichnet.*

Bemerkung: Dass die $p_k = b_{n,p}(\{k\})$ eine Wahrscheinlichkeitsverteilung auf $\Omega = \{0, 1, \ldots, n\}$ definieren, folgt aus Satz 3.2 und Korollar 4.4:

$$\sum_{k=0}^{n} b_{n,p}(\{k\}) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \overset{\text{Kor. 4.4}}{=} (p + (1-p))^n = 1.$$

Bemerkung 5.4

Bei der Definition von hypergeometrischen Verteilung kam es nur auf die Gesamtzahlen der gezogenen schwarzen und weißen Kugeln an, aber nicht auf die genaue Reihenfolge. Zieht man z.B. 3 Kugeln ohne Zurücklegen, aber *mit* Beachtung der Reihenfolge, so erhält man nach der Verallgemeinerung von Fall 1 und Fall 2 aus dem vorigen Abschnitt

$$\mathbb{P}(\{(S, W, S)\}) = \frac{s \cdot w \cdot (s-1)}{n \cdot (n-1) \cdot (n-2)}$$

Allgemeiner gilt, falls man m Kugeln zieht und sich darunter k schwarze (in einer bestimmten Reihenfolge) befinden sollen,

$$\begin{aligned}&\mathbb{P}(\{(S, W, W, \dots, S)\}) = \\ &= \frac{s \cdot (s-1) \cdot \ldots \cdot (s-k+1) \cdot w \cdot (w-1) \cdot \ldots \cdot (w-(m-k)+1)}{n \cdot (n-1) \cdot \ldots \cdot (n-m+1)}\end{aligned}$$

Wird wie in Satz 5.2 der Umfang der Urne immer größer ($n \to \infty$), wobei der relative Anteil der schwarzen Kugeln konstant bleibt ($\frac{s_n}{n} \to p$), gilt für die obige Wahrscheinlichkeit

$$\lim_{n\to\infty} \mathbb{P}(\{(S, W, W, \dots, S)\}) = p^k (1-p)^{m-k}.$$

Im Vergleich zur Binomialverteilung fällt der Binomialkoeffizient $\binom{m}{k}$ weg, da hier die Ziehungsreihenfolge berücksichtigt wird.

Anwendungsbeispiele

Beispiel 5.5

a) In einer Keksdose befinden sich 20 Kekse, davon 6 mit und 14 ohne Schokolade. Wie wahrscheinlich ist es, bei Entnahme von 5 Keksen (ohne hinzusehen!) genau einen Schokokeks zu erwischen?

Ziehen ohne Zurücklegen, Gesamgröße der Urne/Keksdose ist endlich
$\Longrightarrow$ Hypergeometrische Verteilung

$$\mathbb{P}(\{1 \text{ Schokokeks bei 5 Ziehungen}\}) = \frac{\binom{6}{1}\binom{14}{4}}{\binom{20}{5}} \approx 0.387$$

b) Erfahrungsgemäß sind bei der Produktion elektronischer Bauteile 3% der Teile fehlerhaft. Wenn man aus der Gesamtproduktion 10 Teile herausgreift, mit welcher Wahrscheinlichkeit ist dann höchstens eines davon fehlerhaft?

Urne (Gesamtproduktion) sehr groß, daher Ziehen ohne Zurücklegen
$\approx$ Ziehen mit Zurücklegen $\Longrightarrow$ Binomialverteilung

„Höchstens eins“ bedeutet entweder kein oder genau ein fehlerhaftes Teil, also ist

$$\begin{aligned}&\mathbb{P}(\{\text{höchstens eins von 10 Teilen fehlerhaft}\})\\&\qquad=\mathbb{P}(\{\text{0 Teile fehlerhaft}\})+\mathbb{P}(\{\text{1 Teil fehlerhaft}\})\\&\qquad=\binom{10}{0}\cdot 0.03^0\cdot 0.97^{10}+\binom{10}{1}\cdot 0.03^1\cdot 0.97^9\approx 0.965\end{aligned}$$

c) Von k Personen werden in einer anonymen Befragung die Geburtsmonate festgestellt. Wie viele verschiedene Ergebnisse sind bei einer solchen Befragung möglich?
Durch die Anonymisierung hat man k ununterscheidbare Objekte (Personen), die auf 12 verschiedene Schachteln (Geburtsmonate) zu verteilen sind, wobei Mehrfachbelegungen (gleiche Geburtsmonate verschiedener Personen) möglich sind. Das entspricht einer Ziehung von k Elementen aus einer Menge von $N = 12$ Elementen mit Wiederholung und ohne Reihenfolge (für jedes Objekt wird eine Schachtelnummer bzw. Geburtsmonat gezogen).
Dafür gibt es nach Fall 4 $\binom{12+k-1}{k} = \binom{k+11}{k}$ Möglichkeiten.

Satz 5.6

Zu einer endlichen Menge $A = \{a_1, \dots, a_n\}$ und ganzen Zahlen $n_1, \dots, n_r \geq 0$ mit $\sum_{i=1}^r n_i = n$ gibt es genau $\frac{n!}{n_1!n_2!\dots n_r!}$ Zerlegungen von A in Teilmengen $A_1, \dots, A_r$ derart, dass A_i genau n_i Elemente enthält. Die Zahlen $\frac{n!}{n_1!n_2!\dots n_r!} =: \binom{n}{n_1,\dots,n_r}$ heißen ***Multinomialkoeffizienten****.*

Beweis: Man erhält eine Partition von A mit den gewünschten Eigenschaften durch Auswahl der n_1 Elemente für A_1 ($\binom{n}{n_1}$ Möglichkeiten nach Satz 4.1), dann der nächsten n_2 Elemente von A_2 ($\binom{n-n_1}{n_2}$ Möglichkeiten nach Satz 4.1) usw. Die Gesamtzahl der möglichen Partitionen von A in Teilmengen der gewünschten Größe ist dann
$\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \ldots \cdot \binom{n_r}{n_r} = \frac{n!}{n_1!(n-n_1)!} \cdot \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \cdot \ldots \cdot \frac{n_r!}{n_r!0!} = \frac{n!}{n_1!\dots n_r!}$ □

Korollar 5.7 (Multinomialsatz)

$$(x_1 + \dots + x_r)^n = \sum_{\substack{n_1,\dots,n_r \geq 0 \\ \sum_{i=1}^r n_i = n}} \frac{n!}{n_1! \dots n_r!} \cdot x_1^{n_1} \cdot \ldots \cdot x_r^{n_r}, \quad x_1, \dots, x_r \in \mathbb{R},\ n, r \in \mathbb{N}.$$

Beweis:

$$(x_1 + \cdots + x_r)^n = \sum_{\substack{(A_1,\ldots,A_r)\\ \text{Zerlegung}\\ \text{von}\{1,\ldots,n\}}} \prod_{i=1}^{r} x_i^{|A_i|} = \sum_{\substack{n_1,\ldots,n_r \geq 0\\ \sum n_i = n}} \sum_{\substack{(A_1,\ldots,A_r)\\ \text{Zerlegung}\\ \text{mit}|A_i|=n_i}} \prod_{i=1}^{r} x_i^{n_i}$$

$$\overset{\text{Satz 5.6}}{=} \sum_{\substack{n_1,\ldots,n_r \geq 0\\ \sum n_i = n}} \frac{n!}{n_1!\ldots n_r!} \prod_{i=1}^{r} x_i^{n_i} \qquad \square$$

Folgerung: Für Parameter $p_1, \ldots, p_r \geq 0$ mit $\sum_{i=1}^{r} p_i = 1$ und $n, r \in \mathbb{N}$ ist eine Wahrscheinlichkeitsverteilung $\mathbb{P} = M(n, r, p_1, \ldots, p_r)$ auf dem Raum $\Omega = \{(n_1, \ldots, n_r) \mid n_i \geq 0, \sum_{i=1}^{r} n_i = n\}$ gegeben durch

$$\mathbb{P}\big(\{(n_1, \ldots, n_r)\}\big) = \frac{n!}{n_1!\ldots n_r!} \cdot p_1^{n_1} \cdot \ldots \cdot p_r^{n_r}.$$

Diese Verteilung heißt **Multinomialverteilung**.

Beweis: $\sum_{\substack{n_1,\ldots,n_r \geq 0\\ \sum n_i = n}} \frac{n!}{n_1!\ldots n_r!} \cdot p_1^{n_1} \cdot \ldots \cdot p_r^{n_r} \overset{\text{Kor. 5.7}}{=} (p_1 + \cdots + p_r)^n = 1^n = 1.$

Beispiel: Wie groß ist die Wahrscheinlichkeit, bei n Würfen mit einem fairen Würfel n_1-mal die 1, n_2-mal die $2, \dots, n_6$-mal die 6 zu erhalten, wobei $n_i \geq 0$ und $\sum_{i=1}^6 n_i = n$?

Setze $\Omega := \big\{(\omega_1, \dots, \omega_n) \,|\, \omega_i \in \{1, \dots, 6\}\big\}$
und $A := \big\{\omega \in \Omega \,\big|\, |\{i \,|\, \omega_i = j\}| = n_j,\, 1 \leq j \leq 6\big\}$.
Jedes $\omega \in A$ definiert eine geordnete Zerlegung von $\{1, \dots, n\}$ in 6 Teilmengen mit $|A_1| = n_1, \dots |A_6| = n_6$: A_1 enthält die Indizes aller ω_i aus ω mit $\omega_i = 1$, A_2 die Indizes aller ω_i aus ω mit $\omega_i = 2$ usw.
Nach Satz 5.6 ist $|A| = \frac{n!}{n_1! \dots n_6!}$ und nach dem Laplace-Ansatz somit
$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{n!}{n_1! \dots n_6! \cdot 6^n}$

Die entsprechende Verteilung auf $\{(n_1, \dots, n_6) \,|\, n_i \geq 0,\, \sum_{i=1}^6 n_i = n\}$ ist also eine Multinomialverteilung mit den Parametern
$n, 6, p_1 = \dots = p_6 = \frac{1}{6}$.

Bedingte Wahrscheinlichkeit

Einführendes Beispiel: Dreimaliges Werfen einer fairen Münze

„Kopf"$= 1$, „Zahl"$= 0$. Dann ist $\Omega = \{0, 1\}^3$, und mit der Laplace-Annahme gilt $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{8}$ für jedes $\omega \in \Omega$.

Betrachte Ereignis

$A = \{$„mindestens zweimal Kopf"$\} = \{(1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}$,

dann ist $\mathbb{P}(A) = \frac{4}{8} = \frac{1}{2}$.

Angenommen, wir wissen bereits, dass das Ergebnis des ersten Wurfs „Kopf" ist. Wie ändert sich dann unsere Einschätzung für die Wahrscheinlichkeit des Eintretens von A?

Wir wissen also, dass eines der Ereignisse von
$B = \{(1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$ eintritt (und kein anderes!).
Daher können wir B als neuen Grundraum $\tilde{\Omega}$ ansehen mit
$\tilde{\mathbb{P}}(\{\tilde{\omega}\}) = \frac{1}{|\tilde{\Omega}|} = \frac{1}{4}$ für alle $\tilde{\omega} \in \tilde{\Omega} = B$ (Laplace-Annahme).
Auf dem neuen Grundraum ist
$\{$„mindestens zweimal Kopf"$\} = \{(1, 1, 0), (1, 0, 1), (1, 1, 1)\} = A \cap B$.

Folglich ist

$$\tilde{\mathbb{P}}(\{\text{„mindestens zweimal Kopf"}\}) = \frac{3}{4} = \frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|\Omega|}}{\frac{|B|}{|\Omega|}} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

$\tilde{\mathbb{P}}$ kann als *bedingte Wahrscheinlichkeit* aufgefasst werden, die die Wahrscheinlichkeiten von Ereignissen unter der Voraussetzung/Bedingung angibt, dass das Ereignis B in jedem Fall eintrit.

Definition 6.1 (Bedingte Wahrscheinlichkeit)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum,* $B \in \mathcal{A}$ *mit* $\mathbb{P}(B) > 0$*, dann heißt*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (A \in \mathcal{A})$$

die **bedingte Wahrscheinlichkeit von** A **gegeben** B*.*

Weiteres Beispiel: Zweimaliger Würfelwurf

$\Omega = \{(i,j) \mid 1 \leq i,j \leq 6\}$, $\mathbb{P}(\{\omega\}) = \frac{1}{36}\ \forall\, \omega \in \Omega$ (Laplace-Annahme).
$A = \{6 \text{ im ersten Wurf}\} = \{(6,j) \mid 1 \leq j \leq 6\} \quad \Longrightarrow \quad \mathbb{P}(A) = \frac{1}{6}$.
$B = \{\text{Augensumme ist } 11\} = \{(6,5),(5,6)\} \quad \Longrightarrow \quad \mathbb{P}(B) = \frac{1}{18}$.
$\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\{(6,5)\})}{\mathbb{P}(B)} = \frac{\frac{1}{36}}{\frac{1}{18}} = \frac{1}{2}$.

Satz 6.2

$\mathbb{P}(\cdot|B) : \mathcal{A} \to [0,1]$ *ist eine auf B konzentrierte Wahrscheinlichkeitsverteilung.*

Spezialfälle:

$$A \supset B \quad \Longrightarrow \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1,$$

$$A \subset B^C \quad \Longrightarrow \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(B)} = 0,$$

$$B = \Omega \quad \Longrightarrow \quad \mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(\Omega)} = \mathbb{P}(A).$$

Korollar 6.3 (Multiplikationsformel)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum und* $A_1, \ldots, A_n \in \mathcal{A}$ *Ereignisse mit* $\mathbb{P}\big(\bigcap_{i=1}^{n-1} A_i\big) > 0$*, dann gilt*

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

Satz 6.4 (Satz von der totalen Wahrscheinlichkeit)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum und* $B_1, B_2, \dots \in \mathcal{A}$ *eine endliche oder abzählbare Zerlegung von* Ω *(d.h.* $B_i \cap B_j = \emptyset$ *für* $i \neq j$ *und* $\bigcup_{i \geq 1} B_i = \Omega$*) mit* $\mathbb{P}(B_i) > 0$ *für alle* $i \geq 1$*. Dann gilt für jedes* $A \in \mathcal{A}$

$$\mathbb{P}(A) = \sum_{i \geq 1} \mathbb{P}(A | B_i) \cdot \mathbb{P}(B_i).$$

Satz 6.5 (Bayes'sche Regel)

Unter den Voraussetzungen von Satz 6.4 gilt für jedes $A \in \mathcal{A}$ *mit* $\mathbb{P}(A) > 0$*, dass*

$$\mathbb{P}(B_i | A) = \frac{\mathbb{P}(A | B_i) \cdot \mathbb{P}(B_i)}{\sum_{j \geq 1} \mathbb{P}(A | B_j) \cdot \mathbb{P}(B_j)}.$$

Anwendungsbeispiele

Beispiel 6.6

a) Ein elektronisches Gerät enthält zwei Schaltkreise I und II. Schaltkreis I fällt mit Wahrscheinlichkeit 0.1 aus. Fällt Schaltkreis I aus, so fällt Schaltkreis II mit Wahrscheinlichkeit 0.2 ebenfalls aus. Bleibt Schalkreis I intakt, fällt Schaltkreis II mit Wahrscheinlichkeit 0.05 aus. Mit welcher Wahrscheinlichkeit fallen beide Schaltkreise aus? Mit welcher Wahrscheinlichkeit fällt auch Schaltkreis I aus, wenn Schaltkreis II ausfällt?

Seien $A = \{\text{Schaltkreis I fällt aus}\}$ und $B = \{\text{Schaltkreis II fällt aus}\}$, dann folgt aus obigen Angaben

$$\mathbb{P}(A) = 0.1, \quad \mathbb{P}(B|A) = 0.2, \quad \mathbb{P}(B|A^C) = 0.05.$$

Damit ist die Wahrscheinlichkeit für den Ausfall beider Schaltkreise

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A) \cdot \mathbb{P}(A) = 0.2 \cdot 0.1 = 0.02.$$

Die in der zweiten Frage gesuchte Wahrscheinlichkeit ist $\mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$. Hierfür benötigen wir $\mathbb{P}(B)$. Nach Satz 6.4 ist

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(B|A)\cdot\mathbb{P}(A) + \mathbb{P}(B|A^C)\cdot\mathbb{P}(A^C) \\ &= \mathbb{P}(B|A)\cdot\mathbb{P}(A) + \mathbb{P}(B|A^C)\cdot(1-\mathbb{P}(A)) \\ &= 0.2\cdot 0.1 + 0.05\cdot 0.9 = 0.065. \end{aligned}$$

$$\Longrightarrow \quad \mathbb{P}(A|B) = \frac{0.02}{0.065} \approx 0.3078$$

b) Bei der Massenproduktion eines Computerchips ist erfahrungsgemäß 1% der Produktion fehlerhaft. Daher wird jeder einzelne Chip vor der Auslieferung überprüft; Chips, bei denen die Prüfung einen Fehler anzeigt, werden aussortiert, die anderen ausgeliefert. Da auch das Prüfverfahren nicht perfekt ist, zeigt es mit Wahrscheinlichkeit 0.1 bei an sich einwandfreien Chips fälschlich einen Defekt an. Umgekehrt wird bei tatsächlich fehlerhaften Chips mit Wahrscheinlichkeit 0.05 irrtümlich kein Defekt erkannt. Mit welcher Wahrscheinlichkeit ist ein ausgelieferter Chip tatsächlich fehlerfrei? Mit welcher Wahrscheinlichkeit ist ein aussortierter Chip wirklich defekt?

Sei $A = \{\text{Chip ist fehlerfrei}\}$ und $B = \{\text{Prüfung zeigt Fehler an}\}$, dann gilt n.V.

$$\mathbb{P}(A) = 0.99, \quad \mathbb{P}(B|A) = 0.1, \quad \mathbb{P}(B^C|A^C) = 0.05.$$

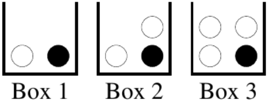
Dann ist die Wahrscheinlichkeit, dass ein ausgelieferter Chip tatsächlich fehlerfrei ist, nach Satz 6.5

$$\begin{aligned}\mathbb{P}(A|B^C) &= \frac{\mathbb{P}(B^C|A)\cdot\mathbb{P}(A)}{\mathbb{P}(B^C|A)\cdot\mathbb{P}(A)+\mathbb{P}(B^C|A^C)\cdot\mathbb{P}(A^C)}\\ &= \frac{(1-\mathbb{P}(B|A))\cdot\mathbb{P}(A)}{(1-\mathbb{P}(B|A))\cdot\mathbb{P}(A)+\mathbb{P}(B^C|A^C)\cdot\mathbb{P}(A^C)}\\ &= \frac{0.9\cdot 0.99}{0.9\cdot 0.99+0.05\cdot 0.01}\approx 0.999\end{aligned}$$

Die Wahrscheinlichkeit, dass ein aussortierter Chip tatsächlich defekt ist, beträgt

$$\begin{aligned}\mathbb{P}(A^C|B) &= \frac{\mathbb{P}(B|A^C)\cdot\mathbb{P}(A^C)}{\mathbb{P}(B|A^C)\cdot\mathbb{P}(A^C)+\mathbb{P}(B|A)\cdot\mathbb{P}(A)}\\ &= \frac{0.95\cdot 0.01}{0.95\cdot 0.01+0.1\cdot 0.99}\approx 0.0876.\end{aligned}$$

c) Von den unten abgebildeten Urnen wird eine zufällig ausgewählt und dann eine Kugel daraus gezogen.



Die gezogene Kugel wird anschließend gezeigt, aber nicht die Urne, aus der sie gezogen wurde. Angenommen, die gezogene Kugel sei weiß, mit welcher Wahrscheinlichkeit stammt sie aus Urne 2, d.h. was ist $\mathbb{P}(\text{Urne } 2|\text{weiß})$?

Aus der obigen Skizze folgt

$$\mathbb{P}(\text{Urne 2 und weiß}) = \mathbb{P}(\text{weiß}|\text{Urne } 2) \cdot \mathbb{P}(\text{Urne } 2) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}.$$

Analog folgt $\mathbb{P}(\text{Urne 1 und weiß}) = \frac{1}{6}$, $\mathbb{P}(\text{Urne 3 und weiß}) = \frac{1}{4}$. Damit ist

$$\mathbb{P}(\text{weiß}) = \frac{2}{9} + \frac{1}{6} + \frac{1}{4} = \frac{23}{36} \quad \Longrightarrow \quad \mathbb{P}(\text{Urne } 2|\text{weiß}) = \frac{\frac{2}{9}}{\frac{23}{36}} = \frac{8}{23}.$$

Unabhängigkeit

Intuitiv ist naheliegend, zwei Ereignisse A und B als unabhängig anzusehen, wenn das Vorliegen des Ereignisses A die Wahrscheinlichkeit für das Auftreten von B nicht beeinträchtigt und umgekehrt, d.h. wenn gilt

$$\mathbb{P}(B|A) = \mathbb{P}(B) \quad \text{und} \quad \mathbb{P}(A|B) = \mathbb{P}(A).$$

Daraus folgt aber wegen $\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$

$$\mathbb{P}(B \cap A) = \mathbb{P}(B) \cdot \mathbb{P}(A) \quad \text{und} \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Definition 6.7 (Unabhängigkeit)

a) *Zwei Ereignisse $A, B \in \mathcal{A}$ eines Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, \mathbb{P})$ heißen* **unabhängig**, *falls $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.*

b) *Eine (endliche oder abzählbare) Folge von Ereignissen $(A_n)_{n \geq 1} \subset \mathcal{A}$ heißt* **paarweise unabhängig**, *falls $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$ für alle $i, j \geq 1$, $i \neq j$.*

Definition 6.7 (Forts.)

c) *Eine (endliche oder abzählbare) Folge von Ereignissen* $(A_n)_{n\geq 1} \subset \mathcal{A}$ *heißt* **unabhängig**, *falls für jede endliche Teilmenge* $T \subset \mathbb{N}$ *gilt, dass* $\mathbb{P}\left(\bigcap_{j\in T} A_j\right) = \prod_{j\in T} \mathbb{P}(A_j)$.

Bemerkung 6.8

a) Die Ereignisse A und Ω bzw. A und $\emptyset$ sind stets unabhängig, denn
$\mathbb{P}(A \cap \Omega) = \mathbb{P}(A) = \mathbb{P}(A) \cdot \mathbb{P}(\Omega)$ und
$\mathbb{P}(A \cap \emptyset) = \mathbb{P}(\emptyset) = 0 = \mathbb{P}(\emptyset) \cdot \mathbb{P}(A)$.

b) Unabhängigkeit einer Folge von Ereignissen $(A_n)_{n\geq 1}$ impliziert stets paarweise Unabhängigkeit (wähle speziell $T = \{i, j\}$), aber die Umkehrung gilt i.A. nicht!

Gegenbeispiel: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $\mathbb{P}(\{\omega_i\}) = \frac{1}{4}$, $1 \leq i \leq 4$.
$A = \{\omega_1, \omega_3\}$, $B = \{\omega_2, \omega_3\}$, $C = \{\omega_3, \omega_4\}$. Dann ist
$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ und
$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \mathbb{P}(\{\omega_3\}) = \frac{1}{4}$, also sind A, B, C paarweise unabhängig.

Bemerkung 6.8 (Forts.)

Wegen $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{\omega_3\}) = \frac{1}{4} \neq \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{1}{8}$ sind die drei Ereignisse jedoch nicht unabhängig.

Beispiel: Zweimaliger Würfelwurf

$\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, \ldots, 6\}\}$, $\mathbb{P}(\{(\omega_1, \omega_2)\}) = \frac{1}{36}$ (Laplace-Ann.)
$A_i^1 := \{(i, \omega_2) \mid \omega_2 \in \{1, \ldots, 6\}\}$ (Ergebnis i im ersten Wurf),
$A_j^2 := \{(\omega_1, j) \mid \omega_1 \in \{1, \ldots, 6\}\}$ (Ergebnis j im zweiten Wurf),
dann ist $\mathbb{P}(A_i^1) = \mathbb{P}(A_j^2) = \frac{1}{6}$ und
$\mathbb{P}(A_i^1 \cap A_j^2) = \mathbb{P}(\{(i, j)\}) = \frac{1}{36} = \mathbb{P}(A_i^1) \cdot \mathbb{P}(A_j^2)$, also sind die Ergebnisse des ersten und des zweiten Würfelwurfs unter der Laplace-Annahme voneinander unabhängig.

Das lässt sich vollkommen analog auch auf den n-maligen Würfelwurf und den n-maligen Münzwurf übertragen: Unter der Laplace-Annahme sind die Ausgänge verschiedener Würfel- oder Münzwürfe stets unabhängig.

Korollar 6.9

Sind die Ereignisse $A_1, \dots, A_n$ unabhängig, dann gilt

$$\mathbb{P}\left(\bigcap_{i=1}^{n} B_i\right) = \prod_{i=1}^{n} \mathbb{P}(B_i),$$

wobei für jedes $i \in \{1, \dots, n\}$ $B_i = A_i$ oder $B_i = A_i^C$ sei.
Insbesondere sind auch die Komplemente $A_1^C, \dots, A_n^C$ unabhängig.

Betrachte n Experimente mit zufälligem Ausgang, wobei jedes Einzelexperiment durch einen diskreten Wahrscheinlichkeitsraum $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$, $1 \leq i \leq n$, beschrieben werde. Bezeichne $\Omega_i = \{\omega_{ij}, j \geq 1\}$ und $\mathbb{P}_i(\{\omega_{ij}\}) = p_{ij}$, so dass $p_{ij} \geq 0$ und $\sum_{j \geq 1} p_{ij} = 1$ für alle $1 \leq i \leq n$.

Das zusammengesetzte Experiment ist dann $\Omega = \prod_{i=1}^{n} \Omega_i$ und $\mathcal{A} := \mathcal{P}(\Omega)$.

$A_{ij} := \Omega_1 \times \cdots \times \Omega_{i-1} \times \{\omega_{ij}\} \times \Omega_{i+1} \times \cdots \times \Omega_n$ ist dann das Ereignis, dass der Ausgang des i-ten Experiments ω_{ij} ist.

Satz 6.10 (Produktraum und Produktmaß)

Es existiert genau ein Wahrscheinlichkeitsmaß $\mathbb{P}$ *auf* $(\Omega, \mathcal{A}) = (\prod_{i=1}^n \Omega_i, \mathcal{P}(\prod_{i=1}^n \Omega_i))$, *so dass jede Familie* $\{A_{1j_1}, \ldots, A_{nj_n}\}$ *unabhängig ist.* $\mathbb{P}$ *ist gegeben durch*

$$\mathbb{P}(\{(\omega_{1j_1}, \ldots, \omega_{nj_n})\}) = \prod_{i=1}^n p_{ij_i}.$$

Der so definierte diskrete Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ *wird das* **Produkt der** $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ *genannt und das obige* $\mathbb{P}$ **Produktmaß**.

Spezialfall: $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1) = \ldots = (\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$, d.h. n-maliges Wiederholen desselben Experiments.

Beispiel 6.11

a) **n-maliger Münzwurf:** Kopf $= 1$, Zahl $= 0$

$\Omega_1 = \ldots = \Omega_n = \{0, 1\}$, $\mathbb{P}_1 = \ldots = \mathbb{P}_n$,

$$\mathbb{P}_i(\{\omega_i\}) = \begin{cases} p, & \omega_i = 1, \\ 1 - p, & \omega_i = 0. \end{cases}$$

Beispiel 6.11 (Forts.)

$$\begin{aligned}\mathbb{P}(\{(\omega_1,\dots,\omega_n)\}) &= p^{\text{Anzahl Einsen}}(1-p)^{\text{Anzahl Nullen}}\\ &= p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}\end{aligned}$$

b) Seien $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ diskrete Wahrscheinlichkeitsräume mit Laplace-Verteilungen $\mathbb{P}_i$, dann ist auch die Produktverteilung wieder eine Laplace-Verteilung, denn:
$\Omega = \Omega_1 \times \cdots \times \Omega_n$, $\omega = (\omega_1, \dots, \omega_n)$, dann ist

$$\mathbb{P}(\{\omega\}) \underset{\text{Def. Prod.vert.}}{=} \prod_{i=1}^n \mathbb{P}_i(\{\omega_i\}) = \prod_{i=1}^n \frac{1}{|\Omega_i|} = \frac{1}{\prod_{i=1}^n |\Omega_i|} = \frac{1}{|\Omega|}.$$

Umgekehrt folgt aus der obigen Gleichung: Sind die Verteilungen $\mathbb{P}_i$ auf den einzelnen Komponenten $(\Omega_i, \mathcal{A}_i)$ Laplace-Verteilungen, und definiert man auf dem Produktraum $(\Omega, \mathcal{P}(\Omega))$ das Wahrscheinlichkeitsmaß $\mathbb{P}$ mittels Laplace-Annahme (sofern Ω endlich ist), sind die einzelnen Komponenten unabhängig voneinander.

Zufallsvariablen

Definition 7.1 (Zufallsvariable)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum. Eine (reellwertige, diskrete) Zufallsvariable ist eine Funktion $X : \Omega \to \mathbb{R}$.
Mit $X(\Omega) = \{x_1, x_2, \ldots\}$ wird der Wertebereich von X bezeichnet.

Satz 7.2 (Verteilung einer Zufallsvariablen)

Durch die Festsetzung

$$\mathbb{P}_X(A) := \mathbb{P}\big(X^{-1}(A)\big) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\})$$

wird eine Wahrscheinlichkeitsverteilung $\mathbb{P}_X$ auf dem Wertebereich $(X(\Omega), \mathcal{P}(X(\Omega)))$ von X definiert. $\mathbb{P}_X$ wird die **Verteilung von** *X oder auch* **das Bildmaß von $\mathbb{P}$ unter der Abbildung** *X genannt.*

Beweis: Für $A \in \mathcal{P}(X(\Omega))$ ist $\mathbb{P}_X(A) = \mathbb{P}(\underbrace{X^{-1}(A)}_{\in \mathcal{P}(\Omega)}) \geq 0$, ferner ist

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

Seien $A_1, A_2, \ldots \in \mathcal{P}(X(\Omega))$ paarweise disjunkt, so gilt weiterhin

$$\begin{aligned}\mathbb{P}_X\left(\bigcup_{i\geq 1} A_i\right) &= \mathbb{P}\left(X^{-1}\left(\bigcup_{i\geq 1} A_i\right)\right) = \mathbb{P}\left(\bigcup_{i\geq 1} \underbrace{X^{-1}(A_i)}_{\text{disjunkt!}}\right)\\ &= \sum_{i\geq 1} \mathbb{P}\big(X^{-1}(A_i)\big) = \sum_{i\geq 1} \mathbb{P}_X(A_i)\end{aligned}$$

$\square$

Notation: Für die Urbildmenge von $A \subset \mathbb{R}$ unter X schreibt man oft auch

$$\{X \in A\} := X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}.$$

Falls $A = (-\infty, x]$, ist $\{X \leq x\} := \{X \in A\}$ (analog sind $\{X < x\}$, $\{X \geq x\}$ und $\{X > x\}$ zu verstehen).

Nach Definition von $\mathbb{P}_X$ ist dann $\mathbb{P}_X(A) = \mathbb{P}(X \in A)$.
Für $\Omega = \{\omega_1, \omega_2, \ldots\}$ und $X^{-1}(x_i) = \omega_i$ ist
$\mathbb{P}_X(\{x_i\}) = \mathbb{P}(X = x_i) = \mathbb{P}(\{\omega_i\})$.

Beispiel 7.3

a) Für das zweimalige Werfen eines fairen Würfels ist $\Omega = \{(\omega_1, \omega_2) \mid \omega_1, \omega_2 \in \{1, \dots, 6\}\}$, $\mathbb{P}(\{\omega\}) = \frac{1}{36}$ (Laplace-Ann.).

Definiere $X : \Omega \to \mathbb{R}$, $X(\omega) = \omega_1 + \omega_2$ (Augensumme beider Würfe), dann ist offensichtlich $X(\Omega) = \{2, 3, \dots, 12\}$ und

$$\mathbb{P}_X(\{2\}) = \mathbb{P}(X^{-1}(2)) = \mathbb{P}(\{\omega \in \Omega \mid \omega_1 + \omega_2 = 2\}) = \mathbb{P}(\{(1,1)\}) = \frac{1}{36},$$

$$\mathbb{P}_X(\{3\}) = \mathbb{P}(X^{-1}(3)) = \mathbb{P}(\{(1,2),(2,1)\}) = \frac{2}{36}, \dots$$

Analog ergibt sich damit die Verteilung $\mathbb{P}_X$ zu

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}_X(\{k\})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Das Bildmaß $\mathbb{P}_X$ ist also im Gegensatz zu $\mathbb{P}$ keine Laplace-Verteilung mehr!

b) n-facher Münzwurf (vgl. Beispiel 6.11 a)), Kopf $= 1$, Zahl $= 0$, $\Omega = \{0,1\}^n$, $\mathbb{P}(\{(\omega_1,\dots,\omega_n)\}) = p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}$.

Definiere $X : \Omega \to \mathbb{R}$, $X(\omega) = \sum_{i=1}^n \omega_i$ (Anzahl Auftreten „Kopf“ in n Würfen), dann ist $X(\Omega) = \{0,1,\dots,n\}$ und $X^{-1}(k) = \{\omega \in \Omega \mid \sum_{i=1}^n \omega_i = k\},\ 0 \le k \le n$.

Da für alle $\omega \in \Omega$ mit $\sum_{i=1}^n \omega_i = k$ gilt $\mathbb{P}(\{\omega\}) = p^k(1-p)^{n-k}$ (s.o.), folgt

$$\begin{aligned}\mathbb{P}_X(\{k\}) &= \mathbb{P}(X^{-1}(k)) = \mathbb{P}(\{\omega \in \Omega \mid \textstyle\sum_{i=1}^n \omega_i = k\}) \\ &= \#\{\omega \in \Omega \mid \textstyle\sum_{i=1}^n \omega_i = k\} \cdot p^k(1-p)^{n-k} \\ &= \binom{n}{k} p^k(1-p)^{n-k} = b_{n,p}(\{k\}),\end{aligned}$$

denn es gibt $\binom{n}{k}$ verschiedene Möglichkeiten, k Einsen auf n Plätze zu verteilen (vgl. Satz 4.1 und Bemerkung 4.3).

Die Anzahl des Auftretens von „Kopf“ (und analog von „Zahl“) in n Münzwürfen ist also binomialverteilt.

Frage: Wie verändert sich bzw. wohin strebt die Verteilung von X aus Beispiel 7.3 b), wenn die Anzahl der Würfe beliebig groß wird ($n \to \infty$), aber gleichzeitig die Wahrscheinlichkeit p für das Auftreten von „Kopf" immer kleiner wird ($p = p_n \to 0$ für $n \to \infty$)?
Allgemeiner formuliert: Wenn die Erfolgswahrscheinlichkeit in einem Versuch immer kleiner wird, man den Versuch aber beliebig oft wiederholen kann, wie sieht dann die Verteilung der Anzahl der Erfolge X aus?

Satz 7.4 (Poissons Gesetz der kleinen Zahlen)

Falls $0 < p_n < 1$ *und* $\lim_{n\to\infty} np_n = \lambda > 0$ (*was* $p_n \to 0$ *impliziert!*), *dann gilt für jedes* $k = 0, 1, 2, \dots$

$$\lim_{n\to\infty} b_{n,p_n}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} =: P_\lambda(\{k\}).$$

Die durch die $P_\lambda(\{k\})$ *auf* $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ *definierte Wahrscheinlichkeitsverteilung heißt* **Poisson-Verteilung**.

Ist bei einer Binomialvereilung der Parameter n sehr groß und p ziemlich klein, kann man somit die (dann schwierig zu berechnenden) Binomialwahrscheinlichkeiten $b_{n,p_n}(\{k\})$ durch die Poisson-Wahrscheinlichkeiten $P_\lambda(\{k\})$ (mit $\lambda = np$) approximieren.

Definition 7.5 (Unabhängigkeit von Zufallsvariablen)

Die auf einem (diskreten) Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ *definierten Zufallsvariablen* $X_1, \ldots, X_n$ *heißen* **unabhängig**, *falls für jede Auswahl von Ereignissen* $A_1 \in \mathcal{P}(X_1(\Omega)), A_2 \in \mathcal{P}(X_2(\Omega)), \ldots, A_n \in \mathcal{P}(X_n(\Omega))$ *gilt*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i),$$

d.h. die zugehörigen Urbilder $\{X_1 \in A_1\}, \ldots, \{X_n \in A_n\}$ *müssen stets unabhängig sein.*

Satz 7.6

Seien $X_1, \ldots, X_n$ *Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, \mathbb{P})$ *mit Wertebereichen* $X_i(\Omega) = \{x_{i1}, x_{i2}, \ldots\}$, $1 \leq i \leq n$. *Dann sind* $X_1, \ldots, X_n$ *unabhängig genau dann, wenn*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i = x_{ij_i}\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_{ij_i}) \quad \text{für alle } j_1, \ldots, j_n.$$

Anwendung: Betrachte den n-fachen Münzwurf aus Beispiel 6.11 a) und 7.3 b) mit $\Omega = \{0,1\}^n$, $\mathbb{P}(\{(\omega_1,\dots,\omega_n)\}) = p^{\sum_{i=1}^n \omega_i}(1-p)^{n-\sum_{i=1}^n \omega_i}$.

Definiere $X_1,\dots,X_n$ durch $X_i(\omega) = \omega_i$, $1 \leq i \leq n$, (Ergebnis des i-ten Wurfs/i-te Projektion)

Dann sind die X_i unabhängig nach Satz 7.6, denn für ein beliebiges $\omega' = (\omega_1',\dots,\omega_n') \in \Omega$ ist

$$\mathbb{P}\left(\bigcap_{i=1}^n \{X_i = \omega_i'\}\right) = \mathbb{P}(\{(\omega_1',\dots,\omega_n')\}) = p^{\sum_{i=1}^n \omega_i'}(1-p)^{n-\sum_{i=1}^n \omega_i'}$$
$$= \prod_{i=1}^n \mathbb{P}(X_i = \omega_i').$$

Satz 7.6 besagt also insbesondere, dass die Würfe bzw. Projektionen genau dann unabhängig sind, wenn das Wahrscheinlichkeitsmaß auf $(\Omega,\mathcal{A})$ das Produktmaß ist.

Wartezeitverteilungen

Wir betrachten noch einmal das Münzwurfmodell mit
$\mathbb{P}(\text{Kopf im } i\text{-ten Wurf}) = \mathbb{P}(X_i = 1) = p > 0$ und $\mathbb{P}(X_i = 0) = 1 - p > 0$.

Nun sei aber die Anzahl der Würfe nicht festgelegt, sondern die Münze wird unabhängig voneinander so oft geworfen, bis das erste Mal „Kopf" erscheint. Wie sehen der Grundraum und die zugehörige Wahrscheinlichkeitsverteilung $\mathbb{P}_X$ der „Wartezeit" X auf das erste Mal „Kopf" aus?

$\Omega = \{(\omega_1, \dots, \omega_n) \mid \omega_1 = \omega_2 = \dots = \omega_{n-1} = 0,\, \omega_n = 1,\, n \geq 1\}$,
die Wartezeit ist $X : \Omega \to \mathbb{N}$, $X((\omega_1, \dots, \omega_n)) = n$.

Wegen der Unabhängigkeit der Würfe ist

$$\mathbb{P}_X(\{n\}) = \mathbb{P}(X = n) = \mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = (1-p)^{n-1} p.$$

Dadurch wird nach Satz 3.2 eine Wahrscheinlichkeitsverteilung auf $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ definiert, denn

$$\sum_{n \geq 1} \mathbb{P}_X(\{n\}) = \sum_{n \geq 1} (1-p)^{n-1} p = p \sum_{n \geq 0} (1-p)^n \underset{\text{geom. Reihe}}{=} p \cdot \frac{1}{1 - (1-p)} = 1.$$

Diese Verteilung heißt **geometrische Verteilung** $Geo(p)$.

Verallgemeinerung: Man wirft die Münze so lange unabhängig voneinander, bis das r-te Mal „Kopf“ erscheint ($r \geq 1$). Dann ist $\Omega = \{(\omega_1, \ldots, \omega_n) \mid \omega_i \in \{0,1\}, \, \omega_n = 1, \, \sum_{i=1}^n \omega_i = r, \, n \geq r\}$, und die Wartezeit ist $X : \Omega \to \{r, r+1, \ldots\}$, $X((\omega_1, \ldots, \omega_n)) = n$.

Sei $A_n = \{(\omega_1, \ldots, \omega_n) \mid \omega_i \in \{0,1\}, \, \omega_n = 1, \, \sum_{i=1}^n \omega_i = r\}$ für ein festes $n \geq r$, dann gilt wegen der Unabhängigkeit der Würfe für alle $\omega \in A_n$ $\mathbb{P}(\{\omega\}) = p^r(1-p)^{n-r}$ (r Erfolge und $n-r$ Misserfolge in n Würfen).

$$\Longrightarrow \; \mathbb{P}_X(\{n\}) = \mathbb{P}(A_n) = |A| \cdot p^r(1-p)^{n-r} = \binom{n-1}{r-1} p^r (1-p)^{n-r},$$

denn der r-te Erfolg muss im letzten (n-ten) Wurf auftreten, die übrigen $r-1$ Erfolge können beliebig auf die ersten $n-1$ Würfe verteilt werden, wofür es nach Satz 4.1 und Bemerkung 4.3 $\binom{n-1}{r-1}$ Möglichkeiten gibt.

Diese auf $(\{r, r+1, \ldots\}, \mathcal{P}(\{r, r+1, \ldots\}))$ definierte Wahrscheinlichkeitsverteilung wird **Pascalsche Verteilung** genannt.

In der Literatur findet man auch oft die auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ durch $\mathbb{P}(\{k\}) = \binom{n+k-1}{k} p^n (1-p)^k$ definierte **negative Binomialverteilung** $BN(n,p)$, die die Wahrscheinlichkeit von n Erfolgen in $n+k$ Würfen ($k \geq 0$) angibt.

Erwartungswert und Varianz

Im Folgenden sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit Wertebereich $X(\Omega) = \{x_1, x_2, \ldots\}$.

Definition 8.1 (Erwartungswert und Varianz)

a) *Falls* $\sum_{i \geq 1} |x_i| \cdot \mathbb{P}(X = x_i) < \infty$, *heißt*

$$\mathbb{E}[X] := \sum_{i \geq 1} x_i \cdot \mathbb{P}_X(\{x_i\}) = \sum_{i \geq 1} x_i \cdot \mathbb{P}(X = x_i) \overset{(*)}{=} \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\{\omega\})$$

der **Erwartungswert von** X.

b) *Für eine Funktion* $f : \mathbb{R} \to \mathbb{R}$ *definiert man analog*

$$\mathbb{E}[f(X)] := \sum_{i \geq 1} f(x_i) \mathbb{P}_X(\{x_i\}) = \sum_{i \geq 1} f(x_i) \mathbb{P}(X = x_i) = \sum_{\omega \in \Omega} f(X(\omega)) \mathbb{P}(\{\omega\}),$$

sofern $\sum_{i \geq 1} |f(x_i)| \mathbb{P}(X = x_i) < \infty$.

Speziell für $f(x) = x^r$ *bzw.* $f(x) = |x|^r$ *heißt* $\mathbb{E}[X^r]$ $(\mathbb{E}[|X|^r])$ r-**tes (absolutes) Moment von** X.

Definition 8.1 (Forts.)

c) *Gilt* $\mathbb{E}[X^2] < \infty$, *so heißt die mittlere quadratische Abweichung vom Erwartungswert*

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

die **Varianz von** X (*hier ist* $f(x) = (x - \mathbb{E}[X])^2$).

Die Wurzel $\sigma_X := \sqrt{\mathrm{Var}[X]}$ *aus der Varianz wird* **Standardabweichung** *genannt.*

Bemerkung 8.2

a) Die mit $(*)$ bezeichnete Gleichung in Teil a) von Definition 8.1 wird auch als *Transformationssatz* bezeichnet. Sie folgt ganz einfach aus der Tatsache, dass $X(\omega) \equiv x_i$ für $\omega \in \{X = x_i\}$ und $\mathbb{P}(X = x_i) = \sum_{\omega \in \{X = x_i\}} \mathbb{P}(\{\omega\})$.

b) Ist $X(\Omega) = \{x_1, \dots, x_n\}$ und $\mathbb{P}_X(\{x_i\}) = \frac{1}{n}$ für alle $1 \leq i \leq n$ (Laplace-Verteilung), so ist der Erwartungswert $\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n x_i$ das arithmetische Mittel der Werte von X. Allgemeiner ist der Erwartungswert ein gewichtetes Mittel der Werte von X, wobei die Gewichte die Elementarwahrscheinlichkeiten von $\mathbb{P}_X$ sind.

c) Nach Definition 8.1 hängen $\mathbb{E}[X]$, $\mathrm{Var}[X]$ und r-te (absolute) Momente nur von der Verteilung $\mathbb{P}_X$ der Zufallsvariablen ab. Sie sind Kenngrößen der Verteilung $\mathbb{P}_X$.

Satz 8.3 (Eigenschaften des Erwartungswertes)

Seien X, Y Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit existierendem Erwartungswert, dann gilt:

a) *Seien $a, b \in \mathbb{R}$, so ist $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ (Linearität).*

b) *Gilt $X \geq Y$ (punktweise, d.h. $X(\omega) \geq Y(\omega)$ für alle $\omega \in \Omega$), so ist $\mathbb{E}[X] \geq \mathbb{E}[Y]$. Insbesondere gilt $\mathbb{E}[X] \geq 0$, falls $X \geq 0$.*

c) *Für alle $A \in \mathcal{A}$ ist $\mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A)$.*

Satz 8.4 (Jensensche Ungleichung)

Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $f : \mathbb{R} \to \mathbb{R}$ eine konvexe Funktion, so dass $\mathbb{E}[X]$ und $\mathbb{E}[f(X)]$ existieren, dann gilt

$$\mathbb{E}[f(X)] \geq f\big(\mathbb{E}[X]\big).$$

Ist die Funktion f konkav, gilt $\mathbb{E}[f(X)] \leq f\big(\mathbb{E}[X]\big)$.

Satz 8.5 (Eigenschaften der Varianz)

Ist X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit $\mathbb{E}[X^2] < \infty$, dann gilt:

a) $0 \leq \mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 < \infty$.

b) $\mathrm{Var}[X] = \mathrm{Var}[X + c]$ *für alle* $c \in \mathbb{R}$ (*Translationsinvarianz*).
Insbesondere ist $\mathrm{Var}[X] = \mathrm{Var}[X - \mathbb{E}[X]]$.

c) $\mathrm{Var}[cX] = c^2 \cdot \mathrm{Var}[X]$ *für alle* $c \in \mathbb{R}$.

Beispiel 8.6

a) **Laplace-Verteilung:** Sei $X(\Omega) = \{k, \ldots, \ell\}$ für $k, \ell \in \mathbb{N}$, $k < \ell$ und $\mathbb{P}_X(\{i\}) = \frac{1}{\ell - k + 1}$ für $i \in \{k, \ldots, \ell\}$. Dann ist

$$\mathbb{E}[X] \underset{\text{Linearität}}{=} \mathbb{E}[X - k] + k = \frac{1}{\ell - k + 1} \cdot \sum_{i=0}^{\ell - k} i \; + k$$
$$= \frac{(\ell - k)(\ell - k + 1)}{2(\ell - k + 1)} + k = \frac{\ell - k}{2} + k = \frac{\ell + k}{2},$$

wobei wir die Summenformel $\sum_{i=0}^{n} i = \frac{n(n+1)}{2}$ benutzt haben.

Für die Varianz erhalten wir aus Satz 8.5 b) und a), dass $\mathrm{Var}(X) = \mathrm{Var}(X-k) = \mathbb{E}[(X-k)^2] - (\mathbb{E}[X-k])^2$. Nun ist

$$\begin{aligned}\mathbb{E}[(X-k)^2] &= \frac{1}{\ell-k+1} \cdot \sum_{i=0}^{\ell-k} i^2 = \frac{(\ell-k)(\ell-k+1)(2\ell-2k+1)}{6(\ell-k+1)} \\ &= \frac{(\ell-k)(2\ell-2k+1)}{6} = \frac{(\ell-k)^2}{3} + \frac{\ell-k}{6},\end{aligned}$$

denn $\sum_{i=0}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$. Ferner ist nach der vorherigen Rechnung $\mathbb{E}[X-k] = \frac{\ell-k}{2}$ und damit

$$\begin{aligned}\mathrm{Var}[X] &= \frac{(\ell-k)^2}{3} + \frac{\ell-k}{6} - \frac{(\ell-k)^2}{4} = \frac{(\ell-k)^2}{12} + \frac{\ell-k}{6} \\ &= \frac{(\ell-k)^2 + (2\ell-2k)}{12} = \frac{(\ell-k)(\ell-k+2)}{12}.\end{aligned}$$

Für einen fairen Würfel ist $k=1$, $\ell=6$, und somit $\mathbb{E}[X] = \frac{1+6}{2} = 3.5$ sowie $\mathrm{Var}[X] = \frac{(6-1)(6-1+2)}{12} = \frac{35}{12}$.

b) **Binomialverteilung:** Nach Beispiel 7.3 b) lässt sich eine binomialverteile Zufallsvariable $X \sim B(n, p)$ darstellen als Summe $\sum_{i=1}^n X_i$, mit identisch nach $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$ verteilten X_i. Daher ist $\mathbb{E}[X_i] = 1 \cdot p + 0 \cdot (1-p) = p$ und wegen der Linearität des Erwartungswertes (Satz 8.3 a))

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p = np.$$

Das kann man (etwas mühsamer) auch direkt nachrechnen:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^n k \cdot \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \underset{\text{Kor. 4.4}}{=} np \cdot (p + (1-p))^{n-1} = np
\end{aligned}$$

Zur Berechnung der Varianz nutzen wir Satz 8.5 a) und berechnen zunächst (mit $q := (1 - p)$)

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} = np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} q^{n-k} \\
&= np \sum_{k=0}^{n-1} (k+1) \binom{n-1}{k} p^k q^{n-k-1} \\
&= np \Bigg[\underbrace{\sum_{k=0}^{n-1} k \binom{n-1}{k} p^k q^{n-k-1}}_{\mathbb{E}[B(n-1,p)]} + \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k q^{n-k-1}}_{=1} \Bigg] \\
&= np \cdot [(n-1)p + 1]
\end{aligned}$$

$$\begin{aligned}
\Longrightarrow \ \mathrm{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = np \cdot [(n-1)p + 1] - (np)^2 \\
&= np - np^2 = np(1-p)
\end{aligned}$$

c) **Poisson-Verteilung:** Wir erinnern an Satz 7.4, nach dem die Poisson-Verteilung sich als Grenzfall der Binomialverteilung für $p \to 0$ und $np \to \lambda > 0$ ergibt. Da Erwartungswert und Varianz Verteilungskenngrößen sind, sollte für gelten, dass

$$\mathbb{E}[B(n,p)] = np \underset{n\to\infty}{\longrightarrow} \lambda = \mathbb{E}[\mathrm{Pois}(\lambda)],$$
$$\mathrm{Var}[B(n,p)] = \underbrace{np}_{\to\lambda}\underbrace{(1-p)}_{\to 1} \underset{n\to\infty}{\longrightarrow} \lambda = \mathrm{Var}[\mathrm{Pois}(\lambda)].$$

Das ist auch für $X \sim \mathrm{Pois}(\lambda)$ der Fall, denn:

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda$$

$$\mathbb{E}[X^2] = \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \cdot \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}$$
$$= \lambda \Bigg[\underbrace{\sum_{k=1}^{\infty} (k-1) \cdot \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}}_{=\mathbb{E}[X]=\lambda} + \underbrace{\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}}_{=1} \Bigg] = \lambda^2 + \lambda$$

Erneut nach Satz 8.5 a) ist somit

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

d) **Geometrische Verteilung:** Für $X \sim \mathit{Geo}(p)$ erhält man mit $q := (1 - p)$

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^{\infty} kpq^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1} = p \cdot \frac{d}{dq} \left(\sum_{k=0}^{\infty} q^k \right) \\ &= p \cdot \frac{d}{dq} \left(\frac{1}{1-q} \right) = p \cdot \frac{1}{(1-q)^2} = p \cdot \frac{1}{p^2} = \frac{1}{p}.\end{aligned}$$

Der Erwartungswert ist hier also gerade der Kehrwert der Erfolgswahrscheinlichkeit. Für einen fairen Würfel bedeutet dies, dass man im Mittel sechs mal würfeln muss, um die erste 6 zu erhalten. Eine faire Münze muss man durchschnittlich zweimal werfen, um das erste Mal „Kopf“ zu erhalten.

Für die Varianz erhält man mit ähnlichen Argumenten, aber länglichen Rechnungen $\mathrm{Var}[X] = \frac{1}{p^2} - \frac{1}{p}$.

Exkurs: Die Laufzeit von QuickSort

Gegeben: Liste von reellen Zahlen $z = (z_1, z_2, \ldots, z_n)$, die aufsteigend geordnet werden sollen.

Naiver Algorithmus (BubbleSort): Bestimme durch $n-1$ Paarvergleiche das Minimum $z_{(1)}$ aller z_i, dann durch $n-2$ Paarvergleiche das Minimum $z_{(2)}$ der übrigen $n-1$ Zahlen usw. Damit benötigt man insgesamt $(n-1)+(n-2)+\ldots+2+1=\frac{(n-1)n}{2}$ Paarvergleiche.

Divide-and-Conquer-Algorithmus (QuickSort): Wähle zufällig einen Laplace-verteilten Index $J \in \{1,\ldots,n\}$ ($\mathbb{P}(J=j)=\frac{1}{n}$ für $1 \leq j \leq n$) und teile z anhand des „Pivot-Elements“ z_J auf in $(z^{(L)}, z_J, z^{(R)})$, wobei

$z^{(L)}$: Liste mit Einträgen z_i von z, so dass $z_i < z_J$,
$z^{(R)}$: Liste mit Einträgen z_i von z, so dass $z_i \geq z_J$ und $i \neq J$.

Damit steht nach einem Durchlauf ($n-1$ Vergleiche aller übrigen z_i mit z_J) das Element z_J an der richtigen Stelle. Falls $z^{(L)}$ und $z^{(R)}$ mehr als ein Element besitzen, verfahre in weiteren Durchläufen mit beiden Teillisten analog.

Als Maß für die Laufzeit von QuickSort betrachten wir die Anzahl $V(z) = V(z, \omega)$ aller Paarvergleiche, die QuickSort zum Sortieren von z benötigt. Diese Größe ist zufällig, da sie von der zu sortierende Folge z und der zufälligen Auswahl des Pivotelements abhängt!

Nach Konstruktion des Algorithmus werden zwei Elemente z_i und z_j $(i \neq j)$ höchstens einmal miteinander verglichen, daher ist $V(z) \leq \frac{n(n-1)}{2}$.

Worst-Case-Laufzeit: $V(z) = \frac{n(n-1)}{2}$. Diese stellt sich ein, wenn bei jeder Umordnung einer Teilliste ihr kleinstes oder größtes Element als Pivoelement gewählt wird.

Frage: *Wie sieht die Average-Case-Laufzeit aus unter der Annahme paarweise verschiedener Komponenten von z?*

Wir bezeichnen mit $z_{(1)} < z_{(2)} < \ldots < z_{(n)}$ die aufsteigend sortierten Elemente von z, d.h. $z_{(1)}$ ist das kleinste, $z_{(2)}$ das zweitkleinste usw.

Satz 8.7

Beim Sortieren von z mit QuickSort gilt für beliebige $1 \leq i < j \leq n$

$$\mathbb{P}(z_{(i)} \text{ und } z_{(j)} \text{ werden verglichen}) = \frac{2}{j - i + 1}.$$

Beweis: Bei der Umordnung $z \rightsquigarrow (z^{(L)}, z_J, z^{(R)})$ können drei Fälle auftreten:

a) $z_J \in \{z_{(i)}, z_{(j)}\}$: Bei der Umordnung werden $z_{(i)}$ und $z_{(j)}$ miteinander verglichen.

b) $z_{(i)} < z_J < z_{(j)}$: Bei Umordnung gilt $z_{(i)} \in z^{(L)}$ und $z_{(j)} \in z^{(R)}$, daher kein direkter Vergleich von $z_{(i)}$ und $z_{(j)}$.

c) $z_J \notin [z_{(i)}, z_{(j)}]$: Bei Umordnung gilt entweder $z_{(i)}, \ldots, z_{(j)} \in z^{(L)}$ oder $z_{(i)}, \ldots, z_{(j)} \in z^{(R)}$, Vergleich von $z_{(i)}$ und $z_{(j)}$ fand noch nicht statt.

Für den Speziallfall $i = 1$ und $j = n$ folgt

$$\mathbb{P}(\text{Vergleich von } z_{(i)} \text{ und } z_{(j)}) = \mathbb{P}(z_J \in \{z_{(1)}, z_{(n)}\}) = \frac{2}{n} = \frac{2}{j - i + 1}$$

Insbesondere ist damit die Behauptung des Satzes für $n = 2$ wahr.
Für $n > 2$ beweisen wir sie mit Induktion und nehmen an, sie gelte für Listen mit $m < n$ paarweise verschiedener Komponenten. Dann gilt

$$\mathbb{P}(\text{Vergleich von } z_{(i)} \text{ und } z_{(j)} \mid z_J = z_{(k)}) = \begin{cases} 1, & k \in \{i,j\}, \\ 0, & i < k < j, \\ \frac{2}{j-i+1}, & k \notin [i,j]. \end{cases}$$

Dann ist

$$\begin{aligned}
&\mathbb{P}(\text{Vergleich von } z_{(i)} \text{ und } z_{(j)}) \\
&= \sum_{k=1}^{n} \mathbb{P}(\text{Vergleich von } z_{(i)} \text{ und } z_{(j)} \mid z_J = z_{(k)}) \cdot \mathbb{P}(z_J = z_{(k)}) \\
&= \frac{1}{n} \sum_{k=1}^{n} \mathbb{P}(\text{Vergleich von } z_{(i)} \text{ und } z_{(j)} \mid z_J = z_{(k)}) \\
&= \frac{2}{n} + \frac{i-1}{n} \cdot \frac{2}{j-i+1} + \frac{n-j}{n} \cdot \frac{2}{j-i+1} \\
&= \frac{2}{j-i+1} \cdot \frac{j-i+1}{n} + \frac{n-(j-i+1)}{n} \cdot \frac{2}{j-i+1} = \frac{2}{j-i+1}.
\end{aligned}$$

□

Satz 8.8 (Average-Case-Laufzeit von QuickSort)

Sei $z = (z_1, \dots, z_n)$ mit paarweise verschiedenen Elementen, dann ist

$$\mathbb{E}[V(z)] = \sum_{d=1}^{n-1} \frac{2(n-d)}{d+1} \begin{cases} \leq 2n\log(n), \\ \geq 2n\log(n) - 4n. \end{cases}$$

Beweis: Sei $i < j$ und $Y_{ij} = \mathbb{1}_{\{z_{(i)} \text{ und } z_{(j)} \text{ werden beim Sortieren von } z \text{ verglichen}\}}$, dann ist $\mathbb{E}[Y_{ij}] = \mathbb{P}(Y_{ij} = 1) = \frac{2}{j-i+1}$ nach Satz 8.7 und damit

$$\mathbb{E}[V] = \mathbb{E}\left[\sum_{1\leq i<j\leq n} Y_{ij}\right] = \sum_{1\leq i<j\leq n} \mathbb{E}[Y_{ij}] = \sum_{1\leq i<j\leq n} \frac{2}{j-i+1}$$

$$= \sum_{d=1}^{n-1}\sum_{i=1}^{n-d} \frac{2}{d+1} = \sum_{d=1}^{n-1} \frac{2(n-d)}{d+1}.$$

Die Abschätzungen folgen nun mittels Integralvergleichskriterium:

Offensichtlich ist

$$\mathbb{E}[V] \leq 2n \sum_{d=1}^{n-1} \frac{1}{d+1} \leq 2n \sum_{d=1}^{n-1} \int_d^{d+1} \frac{1}{x}\, dx = 2n \int_1^n \frac{1}{x}\, dx = 2n \log(n).$$

Andererseits gilt

$$\begin{aligned} \mathbb{E}[V] &\geq 2n \sum_{d=1}^{n-1} \frac{1}{d+1} - 2n \geq 2n \sum_{d=1}^{n-1} \int_{d+1}^{d+2} \frac{1}{x}\, dx - 2n \\ &= 2n(\log(n+1) - \log(2)) - 2n \geq 2n \log(n) - 4n. \end{aligned}$$

□

Man kann ferner zeigen (was deutlich aufwändiger ist), dass gilt

Satz 8.9

Sei $z = (z_1, \ldots, z_n)$ mit paarweise verschiedenen Elementen, dann ist

$$\mathrm{Var}[V(z)] \leq 3n(n-1).$$

Standardisierung von Zufallsvariablen

Eine diskrete Zufallsvariable X heißt *standardisiert*, falls $\mathbb{E}[X] = 0$ und $\mathrm{Var}[X] = 1$. Jede nicht-standardisierte Zufallsvariable X mit $\mathrm{Var}[X] > 0$ kann standardisiert werden durch

$$X^* := \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]}} = \frac{X - \mathbb{E}[X]}{\sigma_X},$$

denn wegen der Linearität des Erwartungswertes (Satz 8.3 a)) ist

$$\mathbb{E}[X^*] = \frac{1}{\sigma_X}\mathbb{E}[X - \mathbb{E}[X]] = \frac{1}{\sigma_X}(\mathbb{E}[X] - \mathbb{E}[X]) = 0,$$

und aus Satz 8.5 b) und c) folgt

$$\mathrm{Var}[X^*] = \frac{1}{\sigma_X^2}\mathrm{Var}[X - \mathbb{E}[X]] = \frac{1}{\sigma_X^2}\mathrm{Var}[X] = \frac{\mathrm{Var}[X]}{\mathrm{Var}[X]} = 1.$$

Bemerkung: Ist $\mathrm{Var}[X] = 0$, so folgt aus Definition 8.1 c) und Satz 8.3 b), dass $\mathbb{P}(X = \mathbb{E}[X]) = 1$, d.h. in diesem Fall ist X konstant ($X \equiv a = \mathbb{E}[X]$) und nicht mehr zufällig, sondern deterministisch.

Seien X, Y zwei diskrete, auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definierte Zufallsvariablen mit Wertebereichen $X(\Omega) = \{x_1, x_2, \ldots\}$ und $Y(\Omega) = \{y_1, y_2, \ldots\}$, dann ist die **gemeinsame Verteilung** $\mathbb{P}_{(X,Y)}$ **von** X **und** Y gegeben durch die Elementarwahrscheinlichkeiten

$$p_{ij} := \mathbb{P}_{(X,Y)}(\{(x_i, y_j)\}) = \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(\{X = x_i\} \cap \{Y = y_j\}).$$

Die **Randverteilungen** $\mathbb{P}_X$ **und** $\mathbb{P}_Y$ erhält man aus der gemeinsamen Verteilung durch Aufsummieren:

$$p_i := \mathbb{P}_X(\{x_i\}) = \sum_{j \geq 1} p_{ij} = \sum_{j \geq 1} \mathbb{P}_{(X,Y)}(\{(x_i, y_j)\}),$$

analog für $\mathbb{P}_Y$.

Bemerkung: Falls $X : \Omega_1 \to \mathbb{R}$ und $Y : \Omega_2 \to \mathbb{R}$, kann man als gemeinsamen Grundraum $(\Omega, \mathcal{P}(\Omega)) = (\Omega_1 \times \Omega_2, \mathcal{P}(\Omega_1 \times \Omega_2))$ nehmen und $X(\omega) := X(\omega_1)$ bzw. $Y(\omega) := Y(\omega_2)$ setzen. Das Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{P}(\Omega))$ muss dabei nicht unbedingt das Produktmaß sein!

Beispiel 8.10 (Gemeinsame und Randverteilungen)

		Y			
		0	1	2	$\mathbb{P}_X(\{x_i\})$
	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
X	1	0	0	$\frac{1}{4}$	$\frac{1}{4}$
	3	0	$\frac{1}{4}$	0	$\frac{1}{4}$
$\mathbb{P}_Y(\{y_j\})$		$\frac{1}{6}$	$\frac{5}{12}$	$\frac{5}{12}$	

Aus Satz 7.6 ergibt sich dann leicht

Korollar 8.11

Die auf einem gemeinsamen Wahrscheinlichkeitsraum definierten Zufallsvariablen $X_1, \ldots, X_n$ *sind unabhängig genau dann, wenn ihre gemeinsame Verteilung* $\mathbb{P}_{(X_1,\ldots,X_n)}$ *gleich dem Produkt der Verteilungen* $\mathbb{P}_{X_i}$ *ist.*

Beweis: Die Unabhängigkeit impliziert, dass die gemeinsame Verteilung das Produktmaß ist, denn

$$\mathbb{P}_{(X_1,\dots,X_n)}(\{(x_{1j_1},\dots,x_{nj_n})\}) = \mathbb{P}\Big(\bigcap_{i=1}^n \{X_i = x_{ij_i}\}\Big) \overset{\text{Unabh.}}{=} \prod_{i=1}^n \mathbb{P}(\{X_i = x_{ij_i}\})$$

$$= \prod_{i=1}^n \mathbb{P}_{X_i}(\{x_{ij_i}\}).$$

Ist die gemeinsame Verteilung das Produktmaß, folgt

$$\prod_{i=1}^n \mathbb{P}_{X_i}(\{x_{ij_i}\}) \overset{\text{gem. Vert.}}{=} \mathbb{P}_{(X_1,\dots,X_n)}(\{(x_{1j_1},\dots,x_{nj_n})\} = \mathbb{P}\Big(\bigcap_{i=1}^n \{X_i = x_{ij_i}\}\Big),$$

also sind die $X_1,\dots,X_n$ unabhängig. □

Die Zufallsvariablen aus Beispiel 8.10 sind also nicht unabhängig, denn z.B. ist $\mathbb{P}_{(X,Y)}(\{(0,0)\}) = \frac{1}{6} \neq \mathbb{P}_X(\{0\}) \cdot \mathbb{P}_Y(\{0\}) = \frac{1}{2} \cdot \frac{1}{6}$.

Satz 8.12 (Produktformel)

Sind X, Y auf einem gemeinsamen Wahrscheinlichkeitsraum definierte, unabhängige Zufallsvariablen mit existierenden Erwartungswerten, dann existiert auch $\mathbb{E}[XY]$, und es gilt $\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

Beweis: Die Existenz von $\mathbb{E}[XY]$ folgt aus folgender Rechnung:

$$\mathbb{E}[|XY|] = \sum_{i,j\geq 1} |x_i y_j| \, \mathbb{P}_{(X,Y)}(\{(x_i, y_j)\}) \underset{\text{Unabh.}}{=} \sum_{i,j\geq 1} |x_i y_j| \, \mathbb{P}_X(\{x_i\})\mathbb{P}_Y(\{y_j\})$$

$$= \left(\sum_{i\geq 1} |x_i| \cdot \mathbb{P}_X(\{x_i\})\right) \cdot \left(\sum_{j\geq 1} |y_j| \cdot \mathbb{P}_Y(\{y_j\})\right) = \mathbb{E}[|X|] \cdot \mathbb{E}[|Y|]$$

Die gleiche Rechnung ohne Beträge liefert dann die Behauptung. $\square$

Definition 8.13 (Kovarianz)

Seien X, Y zwei auf einem gemeinsamen Wahrscheinlichkeitsraum definierte Zufallsvariablen mit $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$, dann heißt

$$\operatorname{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

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

X *und* Y *heißen* **unkorreliert**, *falls* $\operatorname{Cov}[X, Y] = 0$.

Aus Definition 8.13 und den Eigenschaften des Erwartungswertes sieht man leicht, dass

$$\mathrm{Cov}[X,Y] = \mathrm{Cov}[Y,X], \quad \mathrm{Cov}[X,X] = \mathrm{Var}[X],$$
$$\mathrm{Cov}[X+a, Y+b] = \mathrm{Cov}[X,Y] \quad \text{für alle } a,b \in \mathbb{R}.$$

Bemerkung 8.14

Aus Satz 8.12 folgt unmittelbar, dass unabhängige Zufallsvariablen auch unkorreliert sind. *Die Umkehrung gilt jedoch i.A. nicht!* Ein Gegenbeispiel liefern z.B. die Zufallsvariablen aus Beispiel 8.10. Für diese gilt

$$\begin{aligned}
\mathbb{E}[X] &= 0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} + 3\cdot\frac{1}{4} = 1,\\
\mathbb{E}[Y] &= 0\cdot\frac{1}{6} + 1\cdot\frac{5}{12} + 2\cdot\frac{5}{12} = \frac{15}{12} = \frac{5}{4},\\
\mathbb{E}[XY] &= 1\cdot 2\cdot\frac{1}{4} + 3\cdot 1\cdot\frac{1}{4} = \frac{5}{4},
\end{aligned}$$

also ist $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, d.h. X und Y sind unkorreliert. Wir haben aber bereits gesehen, dass X und Y nicht unabhängig sind.

Korollar 8.15

Seien $X_1, \ldots, X_n$ auf einem gemeinsamen Wahrscheinlichkeitsraum definierte Zufallsvariablen mit $\mathbb{E}[X_i^2] < \infty$, $1 \leq i \leq n$. Dann existiert auch die Varianz ihrer Summe $\mathrm{Var}[X_1 + \ldots + X_n]$, und es gilt

$$\mathrm{Var}[X_1 + \ldots + X_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + 2 \cdot \sum_{i,j=1, i<j}^{n} \mathrm{Cov}[X_i, X_j].$$

Sind die X_i zusätzlich paarweise unkorreliert, gilt

$$\mathrm{Var}[X_1 + \ldots + X_n] = \sum_{i=1}^{n} \mathrm{Var}[X_i]. \qquad \text{(Bienaymé-Gleichung)}$$

Beispiel 8.16

Sei X_1 eine Zufallsvariable mit $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = 0) = 1 - p$, dann ist $\mathbb{E}[X_1] = 1 \cdot p + 0 \cdot (1 - p) = p$ und $\mathbb{E}[X_1^2] = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$, also ist $\mathrm{Var}[X_1] = \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2 = p - p^2 = p(1 - p)$.

Beispiel 8.16 (Forts.)

Sei nun X definiert als die Anzahl der Erfolge („Kopf") beim n-fachen Münzwurf, wobei die Erfolgswahrscheinlichkeit bei einem einzelnen Wurf gleich p sei. Dann gilt nach den Beispielen 6.11 a) und 7.3 b) sowie Folie 61, dass X binomialverteilt zu den Parametern n und p ist und sich schreiben lässt als $X = \sum_{i=1}^n X_i$, wobei die X_i unabhängig und identisch verteilt sind wie X_1 oben. Nach Korollar 8.15 gilt daher

$$\mathrm{Var}[X] = \sum_{i=1}^n \mathrm{Var}[X_i] = \sum_{i=1}^n p(1-p) = np(1-p),$$

was mit unserer Rechnung aus Beispiel 8.6 b) übereinstimmt.

Definition 8.17 (Korrelationskoeffizient)

Seien X, Y zwei auf einem gemeinsamen Wahrscheinlichkeitsraum definierte Zufallsvariablen mit $0 < \mathrm{Var}[X], \mathrm{Var}[Y] < \infty$, dann heißt

$$\rho_{X,Y} := \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X \sigma_Y}$$

der **Korrelationskoeffizient von X und Y**.

Seien X, Y wie in Definition 8.17 und $X_* := X - \mathbb{E}[X]$, $Y_* := Y - \mathbb{E}[Y]$, dann gilt $\mathbb{E}[X_*] = \mathbb{E}[Y_*] = 0$ und $\mathrm{Var}[X_*] = \mathbb{E}[X_*^2]$, $\mathrm{Var}[Y_*] = \mathbb{E}[Y_*^2]$. Aus den Eigenschaften von Varianz und Kovarianz folgt dann

$$\begin{aligned}\rho_{X,Y} &= \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}} = \frac{\mathrm{Cov}[X_*, Y_*]}{\sqrt{\mathrm{Var}[X_*]}\sqrt{\mathrm{Var}[Y_*]}} \\ &= \frac{\mathbb{E}[X_* Y_*]}{\sqrt{\mathbb{E}[X_*^2]}\sqrt{\mathbb{E}[Y_*^2]}} = \rho_{X_*,Y_*}.\end{aligned}$$

Aus der Satz 8.18 unten folgt damit $-1 \leq \rho_{X,Y} \leq 1$.

Satz 8.18 (Cauchy-Schwarzsche Ungleichung)

Seien X, Y zwei auf einem gemeinsamen Wahrscheinlichkeitsraum definierte Zufallsvariablen mit $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$, dann gilt

$$(\mathbb{E}[XY])^2 \leq (\mathbb{E}[|XY|])^2 \leq \mathbb{E}[X^2] \cdot \mathbb{E}[Y^2].$$

Die Gleichheit gilt genau dann, wenn X und Y mit Wahrscheinlichkeit 1 linear abhängig sind, d.h. wenn $\mathbb{P}(aX + bY = 0) = 1$ für geeignete $a, b \in \mathbb{R}$ (und nicht $a = b = 0$ ist).

Stetige Wahrscheinlichkeitsverteilungen

Bislang: Betrachtung diskreter Wahrscheinlichkeitsräume $(\Omega, \mathcal{A}, \mathbb{P})$, wobei Ω eine höchstens abzählbare Menge ist, $\mathcal{A} = \mathcal{P}(\Omega)$, und $\mathbb{P}$ durch die Elementarwahrscheinlichkeiten $\mathbb{P}(\{\omega\})$, $\omega \in \Omega$, eindeutig festgelegt ist.

Für reelle Zufallsvariable $X : \Omega \to \mathbb{R}$ ist dann auch $(X(\Omega), \mathcal{P}(X(\Omega)), \mathbb{P}_X)$ ein diskreter Wahrscheinlichkeitsraum, insbesondere ist auch $X(\Omega)$ höchstens abzählbar.

Künftig wollen wir auch reelle Zufallsvariablen $X : \Omega \to \mathbb{R}$ betrachten, deren Wertebereich $X(\Omega)$ ganz $\mathbb{R}$ oder eine überabzählbare Teilmenge von $\mathbb{R}$ (z.B. $\mathbb{R}_+$) sein kann. (Hier muss dann auch der Grundraum Ω bereits überabzählbar sein, ansonsten gäbe es mehr Bilder als Urbilder, so dass X keine Funktion wäre.) In diesem Fall lässt sich das Wahrscheinlichkeitsmaß $\mathbb{P}_X$ auf $X(\Omega)$ nicht mehr über Elementarwahrscheinlichkeiten festlegen. Stattdessen kann man eine Dichte nehmen gemäß folgender

Definition 9.1 (Wahrscheinlichkeitsdichte)

Eine **Wahrscheinlichkeitsdichte** *ist eine nicht-negative Funktion* $f : \mathbb{R} \to [0, \infty)$, *für die gilt* $\int_{\mathbb{R}} f(y)\, dy = 1$.

Definition 9.2 (Stetiges Wahrscheinlichkeitsmaß)

Ist $\mathbb{P}_X$ ein Wahrscheinlichkeitsmaß auf $(X(\Omega), \mathcal{B})$, dann heißt f_X **die zu $\mathbb{P}_X$ gehörige Wahrscheinlichkeitsdichte**, *falls gilt*

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \int_B f_X(y)\,dy = \int_{\mathbb{R}} \mathbb{1}_B(y)\,f_X(y)\,dy \quad \text{für alle } B \in \mathcal{B}.$$

Umgekehrt lässt sich bei gegebener Dichte f_X durch obige Gleichung auch ein Wahrscheinlichkeitsmaß $\mathbb{P}_X$ auf $(X(\Omega), \mathcal{B})$ definieren, sofern $\int_{X(\Omega)} f_X(y)\,dy = 1$ (d.h. die Dichte ist auf $X(\Omega)$ konzentriert).
Ein durch eine Dichte definiertes Wahrscheinlichkeitsmaß nennen wir **stetiges Wahrscheinlichkeitsmaß oder auch stetige Wahrscheinlichkeitsverteilung**.

Bemerkung 9.3

a) Ein stetiges Wahrscheinlichkeitsmaß erfüllt ebenfalls die Bedingungen a) und b) aus Definition 3.1. Normierung und Nicht-Negativität des Wahrscheinlichkeitsmaßes folgen aus den definierenden Eigenschaften einer Wahrscheinlichkeitsdichte, die σ-Additivität folgt aus den entsprechenden Eigenschaften des Integrals.

b) Anstelle der Elementarwahrscheinlichkeiten ist ein stetiges Wahrscheinlichkeitsmaß $\mathbb{P}_X$ durch die Werte $\mathbb{P}_X([a,b]) = \int_a^b f_X(y)\,dy = \mathbb{P}(a \leq X \leq b)$ mit $a \leq b \in \mathbb{R}$ eindeutig festgelegt.

c) Für ein stetiges Wahrscheinlichkeitsmaß gilt stets $\mathbb{P}(X = a) = \mathbb{P}_X(\{a\}) = \mathbb{P}_X([a,a]) = \int_a^a f_X(y)\,dy = 0$ für alle $a \in \mathbb{R}$.
Aus der σ-Additivität folgt damit für $a < b$

$$\begin{aligned}\mathbb{P}_X((a,b)) &= \mathbb{P}_X\big([a,b] \setminus (\{a\} \cup \{b\})\big) = \mathbb{P}_X([a,b]\big) - \mathbb{P}_X(\{a\}) - \mathbb{P}_X(\{b\}) \\ &= \mathbb{P}_X([a,b]) = \int_a^b f_X(y)\,dy,\end{aligned}$$

d.h. $\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b)$.
Analog zeigt man $\mathbb{P}_X((a,b]) = \mathbb{P}_X([a,b)) = \mathbb{P}_X([a,b])$.

d) Das Mengensystem $\mathcal{B}$ von Ereignissen $B \subseteq X(\Omega)$, denen man durch $\mathbb{P}_X(B) = \int_B f_X(y)\,dy$ eine Wahrscheinlichkeit zuordnen kann, ist i.A. echt kleiner als die Potenzmenge $\mathcal{P}(X(\Omega))$, da sich das Integral $\int_B f_X(y)\,dy$ nicht für beliebige Teilmengen $B \subset \mathbb{R}$ definieren lässt (es gibt „nicht integrierbare" Teilmengen von $\mathbb{R}$).

Wir gehen hierauf nicht weiter ein, sondern nehmen im Folgenden stets an, dass $\mathcal{B}$ alle Teilmengen von $X(\Omega)$ enthält, für die man $\mathbb{P}_X(B) = \int_B f_X(y)\,dy$ sinnvoll definieren kann.

Insbesondere enthält $\mathcal{B}$ alle Intervalle der Form $[a, b]$, (a, b), $(a, b]$, $[a, b)$ mit $-\infty \leq a \leq b \leq \infty$ sowie deren Komplemente, abzählbare Vereinigungen und abzählbare Schnitte.

e) Stetige Wahrscheinlichkeitsverteilungen sind meistens auf dem Bildraum $(X(\Omega), \mathcal{B}, \mathbb{P}_X)$ definiert und geben die Verteilung einer Zufallsvariablen an. Der Urbild- oder Grundraum $(\Omega, \mathcal{A}, \mathbb{P})$, von dem aus die Zufallsvariable X nach $\mathbb{R}$ abbildet, ist in diesem Fall oft nicht so wichtig und tritt in den Hintergrund.

Bei Bedarf kann man sich aber leicht einen passenden Grundraum definieren, indem man $(\Omega, \mathcal{A}, \mathbb{P}) = (X(\Omega), \mathcal{B}, \mathbb{P}_X)$ wählt und als Zufallsvariable X die Identität, d.h. $X(\omega) = \omega$ für alle $\omega \in \Omega$.

Verteilungsfunktionen

Definition 9.4 (Verteilungsfunktion)

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $X : \Omega \to \mathbb{R}$ *eine reelle Zufallsvariable, dann heißt die durch*

$$F_X(z) := \mathbb{P}_X((-\infty, z]) = \mathbb{P}(X \leq z)$$

definierte Funktion $F_X : \mathbb{R} \to [0,1]$ **Verteilungsfunktion von X bzw. der Verteilung $\mathbb{P}_X$ von X**.

Bemerkung 9.5

a) Für stetige Verteilungen $\mathbb{P}_X$ mit Dichte f_X ist

$$F_X(z) = \int_{-\infty}^{z} f_X(y)\, dy.$$

Hier ist die Verteilungsfunktion offensichtlich stetig. Ist $C_{f_X} := \{z \in \mathbb{R} \mid f_X \text{ stetig in } z\}$ die Menge der Stetigkeitsstellen von f_X, so gilt ferner $F_X'(z) = f_X(z)$ für alle $z \in C_{f_X}$.

b) Ist $\mathbb{P}_X$ diskret und $X(\Omega) = \{x_1, x_2, \ldots\}$ der höchstens abzählbare, aufsteigend geordnete Wertebereich von X (d.h. $x_1 < x_2 < \ldots$), ist

$$F_X(z) := \sum_{x_k \leq z} \mathbb{P}_X(\{x_k\}) = \sum_{k=0}^{k^*} \mathbb{P}(X = x_k),$$

wobei $k^* = \max\{k \geq 0 \,|\, x_k \leq z\}$. In diesem Fall ist die Verteilungsfunktion stückweise konstant und springt nur jeweils an den Stellen x_k um den Wert $\mathbb{P}(X = x_k)$.

Satz 9.6 (Eigenschaften von Verteilungsfunktionen)

Jede Verteilungsfunktion F_X besitzt die folgenden Eigenschaften:

F_X ist monoton wachsend, rechtsseitig stetig, und es gilt
$\lim_{z\to-\infty} F_X(z) = 0, \quad \lim_{z\to\infty} F_X(z) = 1.$

Beweis: Ist $z_1 \leq z_2$, gilt $(-\infty, z_1] \subseteq (-\infty, z_2]$ und damit wegen der Monotonie von Wahrscheinlichkeitsmaßen

$$F_X(z_1) = \mathbb{P}_X((-\infty, z_1]) \leq \mathbb{P}_X((-\infty, z_2]) = F_X(z_2).$$

Die rechtsseitige Stetigkeit folgt aus der sog. „Stetigkeit von oben“, die man aus der σ-Additivität eines Wahrscheinlichkeitsmaßes ableiten kann: Ist $A_1 \supset A_2 \supset \dots$ eine absteigende Folge von Ereignissen aus $\mathcal{A}$, so gilt $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\bigcap_{n=1}^{\infty} A_n)$.
Ist nun $z_1 > z_2 > \dots$ eine absteigende Folge reeller Zahlen mit $z = \lim_{n\to\infty} z_n$, so ist $\bigcap_{n=1}^{\infty}(-\infty, z_n] = (-\infty, z]$, so dass man aus der Stetigkeit von oben erhält

$$\begin{aligned}\lim_{n\to\infty} F_X(z_n) = \lim_{n\to\infty} \mathbb{P}_X((-\infty, z_n]) &= \mathbb{P}_X\left(\bigcap_{n=1}^{\infty}(-\infty, z_n]\right)\\ &= \mathbb{P}_X((-\infty, z]) = F_X(z).\end{aligned}$$

Für stetige Wahrscheinlichkeitsverteilungen ist

$$\lim_{z\to-\infty} F_X(z) = \lim_{z\to-\infty} \int_{-\infty}^{z} f_X(y)\,dy = \int_{-\infty}^{-\infty} f_X(y)\,dy = 0$$

und

$$\lim_{z\to\infty} F_X(z) = \int_{-\infty}^{\infty} f_X(y)\,dy = 1$$

nach Definition der Wahrscheinlichkeitsdichte.

Für eine diskrete Zufallsvariable X mit Wertebereich $X(\Omega) = \{x_1, x_2, \dots\} \subset \mathbb{R}$ gilt für $z \uparrow \infty$, dass $\{x_k \in X(\Omega) \mid x_k \leq z\} \uparrow X(\Omega)$. Folglich ist

$$\lim_{z \to \infty} F_X(z) = \lim_{z \to \infty} \sum_{x_k \leq z} \mathbb{P}_X(\{x_k\}) = \sum_{k \geq 1} \mathbb{P}_X(\{x_k\}) = 1.$$

Die zweite Behauptung folgt analog aus $\{x_k \in X(\Omega) \mid x_k \leq z\} \downarrow \emptyset$ für $z \downarrow -\infty$. $\square$

Bemerkung 9.7

Die Verteilung $\mathbb{P}_X$ ist durch die Verteilungsfunktion F_X bereits eindeutig festgelegt: Für stetige Verteilungen bzw. Verteilungsfunktionen ist

$$\begin{aligned} F_X(b) - F_X(a) &= \mathbb{P}_X((-\infty, b]) - \mathbb{P}_X((-\infty, a]) = \mathbb{P}_X((-\infty, b] \setminus (-\infty, a]) \\ &= \mathbb{P}_X((a, b]) \underset{9.5\text{ c)}}{=} \mathbb{P}_X([a, b]), \end{aligned}$$

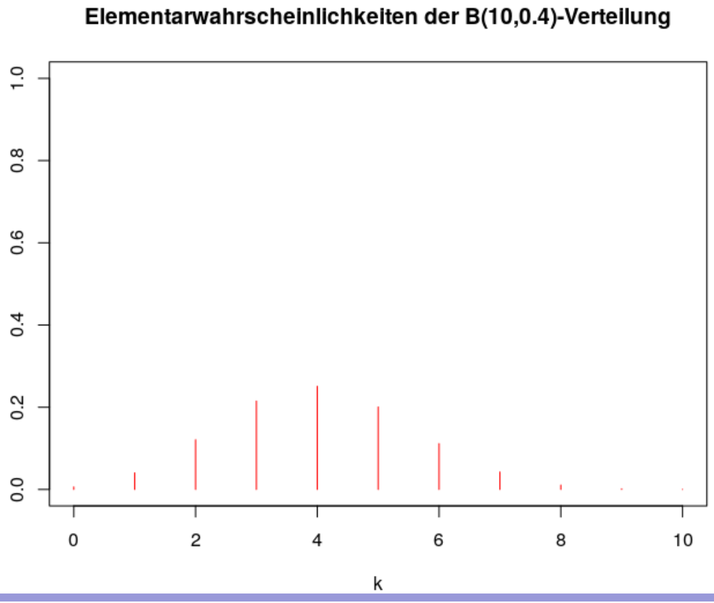
also sind alle Wahrscheinlichkeiten $\mathbb{P}_X([a, b])$ und damit nach Bemerkung 9.5 b) bereits $\mathbb{P}_X$ durch F_X eindeutig bestimmt.

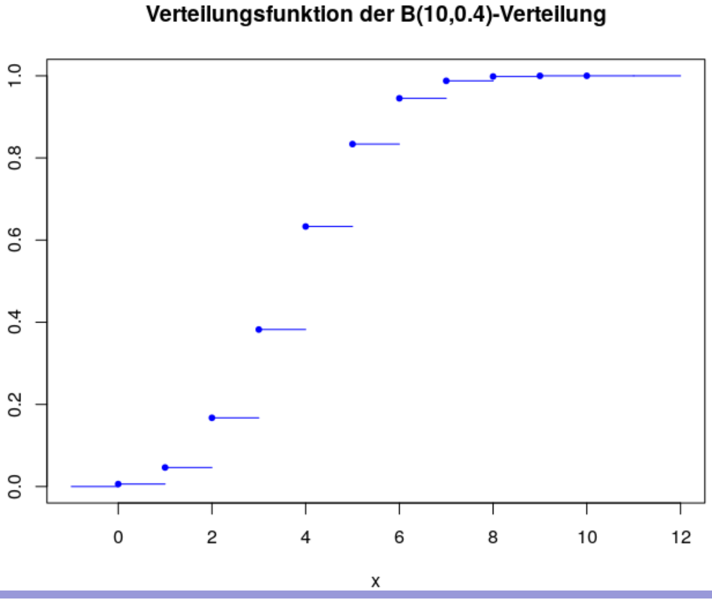
Bemerkung 9.7 (Forts.)

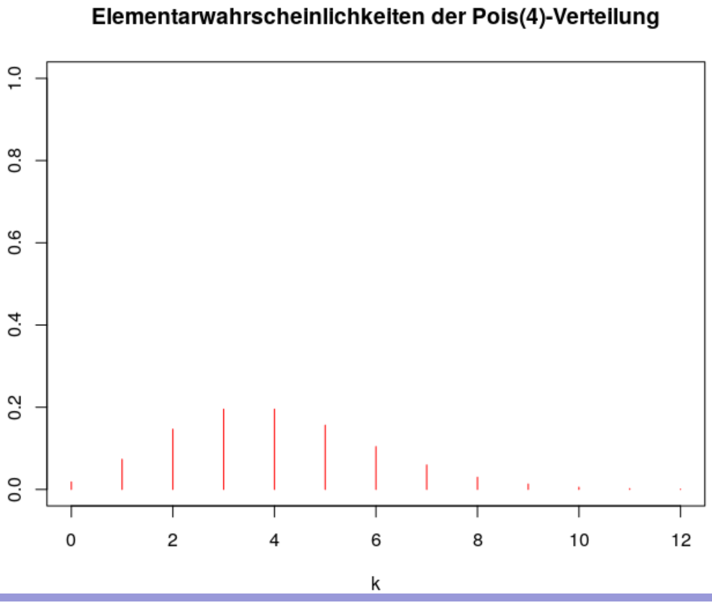
Für eine diskrete Zufallsvariable X mit $X(\Omega) = \{x_1, x_2, \ldots\} \subset \mathbb{R}$ wähle ein x_k sowie eine Folge $z_n \uparrow x_k$, $z_n < x_k$ für alle n, dann folgt mit Stetigkeit von oben

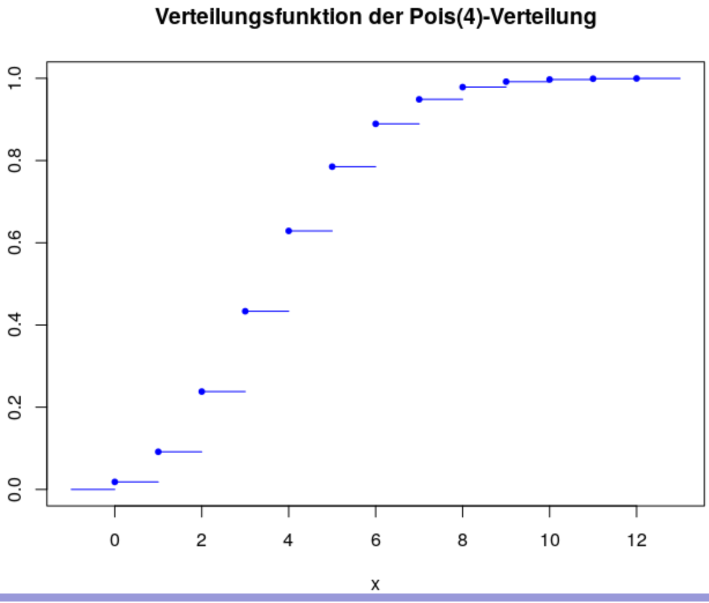
$$F_X(x_k) - \lim_{n\to\infty} F_X(z_n) = \lim_{n\to\infty} \left(F_X(x_k) - F_X(z_n)\right) = \lim_{n\to\infty} \mathbb{P}_X((z_n, x_k])$$
$$= \mathbb{P}_X\left(\bigcap_{n=1}^{\infty}(z_n, x_k]\right) = \mathbb{P}_X(\{x_k\}),$$

d.h. die Elementarwahrscheinlichkeiten $\mathbb{P}_X(\{x_k\}) = \mathbb{P}(X = x_k)$ erhält man aus den Sprunghöhen $F_X(x_k) - F_X(x_k-)$ an den Sprungstellen x_k von F_X (dabei bezeichnet $F_X(z-) := \lim_{z_n\uparrow z} F_X(z_n)$ den linksseitigen Grenzwert von F_X an der Stelle z).
Das ist konsistent mit Bemerkung 9.5 b).


Elementarwahrscheinlichkeiten der B(10,0.4)-Verteilung
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
k


Verteilungsfunktion der B(10,0.4)-Verteilung
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
x


Elementarwahrscheinlichkeiten der Pois(4)-Verteilung
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
k


Verteilungsfunktion der Pois(4)-Verteilung
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
x

Beispiel 9.8 (Stetige Verteilungen)

a) **Gleichverteilung auf einem Intervall** $[a, b]$ Die Gleich- oder auch Uniformverteilung $U([a, b])$ (mit $a < b$) ist das stetige Analogon zur diskreten Laplace-Verteilung. Sie hat die Dichte

$$f_{U([a,b])}(y) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(y)$$

und die Verteilungsfunktion

$$F_{U([a,b])}(z) = \begin{cases} 0, & z < a, \\ \frac{z-a}{b-a}, & a \leq z \leq b, \\ 1, & z > b. \end{cases}$$

b) **Exponentialverteilung zum Parameter** $\lambda > 0$: Die Exponentialverteilung $\mathrm{Exp}(\lambda)$ ist auf $\mathbb{R}_+$ definiert. Die zugehörige Dichte ist

$$f_{\mathrm{Exp}(\lambda)}(y) = \lambda e^{-\lambda y} \mathbb{1}_{[0,\infty)}(y),$$

und für die Verteilungsfunktion erhält man (für $z \geq 0$)

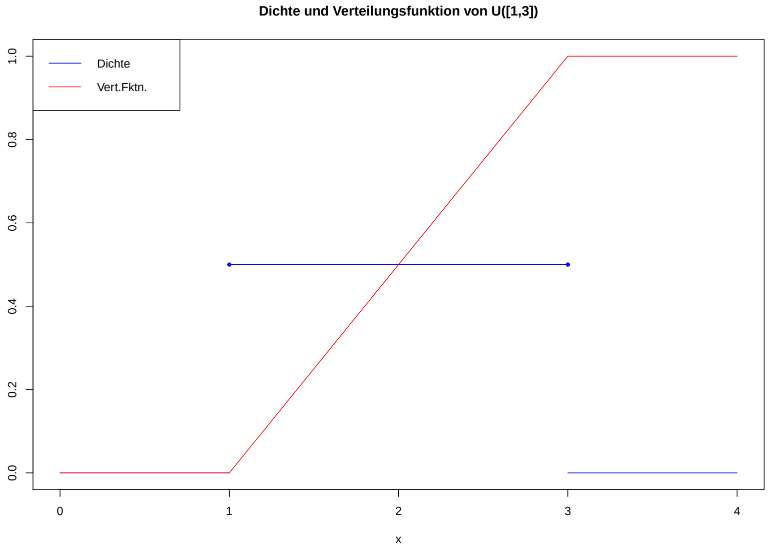
$$F_{\mathrm{Exp}(\lambda)}(z) = \int_{-\infty}^{z} f_{\mathrm{Exp}(\lambda)}(y)\,dy = \int_{0}^{z} \lambda e^{-\lambda y}\,dy = -e^{-\lambda y}\Big|_{0}^{z} = 1 - e^{-\lambda z}$$

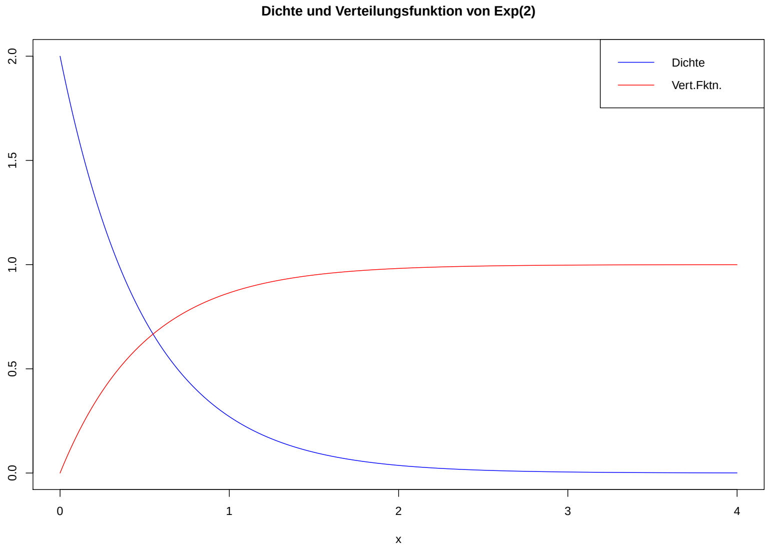
(für $z < 0$ ist $F_{\mathrm{Exp}(\lambda)}(z) \equiv 0$).

Anwendung: Modell für Warte- oder Lebenszeiten

- Bei gleichartigen, unabhängig voneinander auftretenden Ereignissen: Wartezeit bis zum nächsten Ereignis (Anruf in einer Telefonzentrale, Zerfall eines Atoms in einer radioaktiven Materialprobe, Autounfall an verkehrsreicher Kreuzung, ...)
- Lebensdauer von Geräten, Glühbirnen, elektronischen Bauteilen (=Wartezeit bis zum (ersten) Ausfall)

In diesem Fall kann der Parameter λ als Ereignisrate (Anzahl Ereignisse pro Zeiteinheit) aufgefasst werden.


Dichte und Verteilungsfunktion von U([1,3])
Dichte
Vert.Fktn.
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
x



Dichte und Verteilungsfunktion von Exp(2)
Dichte
Vert.Fktn.
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
x

Die Exponentialverteilung kann auch als „stetige Version der geometrischen Verteilung“ aufgefasst werden:

Erinnerung: Die geometrische Verteilung beschreibt die Wartezeit bis zum ersten Erfolg bei unabhängigen Wiederholungen eines Experiments mit Erfolgswahrscheinlichkeit p:
$\mathbb{P}(\{n\}) = p \cdot (1-p)^{n-1}$.

Annahme: Das Experiment wird in kurzen Zeitintervallen Δt wiederholt mit Erfolgswahrscheinlichkeit $\lambda \Delta t$.

Sei T die Wartezeit bis zum ersten Erfolg. Wenn dieser zur Zeit t auftritt, hat man ungefähr $n \approx \frac{t}{\Delta t}$ Versuche benötigt, d.h.

$$\mathbb{P}(T = t) \approx \lambda \Delta t (1 - \lambda \Delta t)^{\frac{t}{\Delta t} - 1}$$

$$\text{bzw.} \quad \mathbb{P}(t < T \leq t + \Delta t) \approx \lambda \Delta t (1 - \lambda \Delta t)^{\frac{t}{\Delta t}}$$

Durch Grenzübergang $\Delta t \to dt$ (d.h. das Zeitintervall wird infinitesimal klein) erhält man (formal)

$$\mathbb{P}(T \in (t, t+dt]) = \lambda e^{-\lambda t} dt \quad \text{und damit} \quad \mathbb{P}(s \leq T \leq t) = \int_s^t \lambda e^{-\lambda y}\, dy.$$

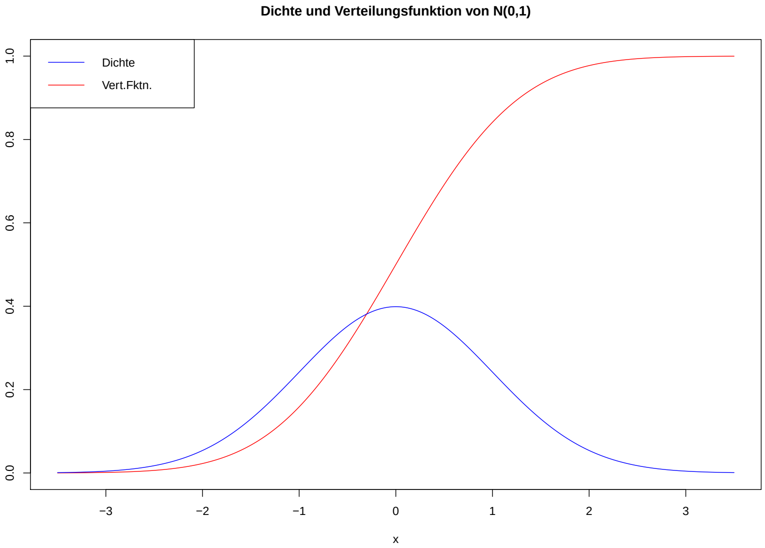
Bemerkung 9.8 (Forts.)

c) **Normalverteilung** $N(\mu, \sigma^2)$: Die Normalverteilung ist eine der wichtigsten Grenzwertverteilungen und tritt im Zentralen Grenzwertsatz auf, den wir später noch kennenlernen werden. Sie ist auf ganz $\mathbb{R}$ definiert durch die Dichte

$$f_{N(\mu,\sigma^2)}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}} \qquad \mu \in \mathbb{R}, \sigma > 0.$$

Mit $\mu = 0$, $\sigma = 1$ erhält man die **Standard-Normalverteilung** $\mathbf{N(0,1)}$ mit Dichte $\varphi(y) := f_{N(0,1)}(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$ und Verteilungsfunktion $\Phi(z) := \int_{-\infty}^{z} \varphi(y)\, dy$. Diese liegt tabelliert vor. Ist $X \sim N(\mu, \sigma^2)$, dann ist nach dem Trafo-Satz für Integrale

$$\begin{aligned} \mathbb{P}(a \le X \le b) \quad &= \quad \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy \\ &\underset{x=\frac{y-\mu}{\sigma}}{=} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$


Dichte und Verteilungsfunktion von N(0,1)
Dichte
Vert.Fktn.
0.0
0.2
0.4
0.6
0.8
1.0
−3
−2
−1
0
1
2
3
x

Beweis, dass $f_{N(\mu,\sigma^2)}(y)$ für alle $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ eine Wahrscheinlichkeitsdichte ist:

Es genügt der Nachweis für $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$, denn sei $y = \sigma x + \mu$, so ist $x = \frac{y-\mu}{\sigma}$ und $dx = \sigma^{-1} dy$ und nach der Trafo-Regel für die Integration

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

Offensichtlich ist $\varphi(x)$ stetig, beschränkt auf $[-1, 1]$ und $\varphi(x) \leq \frac{1}{\sqrt{2\pi}}\, e^{-\frac{|x|}{2}}$ für $|x| \geq 1$, daher ist $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx < \infty$.

Ferner $\int_{-\infty}^{\infty} \varphi(x)\, dx = 1 \quad \Longleftrightarrow \quad \left(\int_{-\infty}^{\infty} \varphi(x)\, dx\right) \cdot \left(\int_{-\infty}^{\infty} \varphi(x)\, dx\right) = 1.$

$$\left(\int_{-\infty}^{\infty} \varphi(x)\, dx\right) \cdot \left(\int_{-\infty}^{\infty} \varphi(x)\, dx\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x)\, \varphi(y)\, dx\, dy$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}}\, dx\, dy.$$

Quantile

Übergang zu Polarkoordinaten: $x = r \cdot \cos(\phi)$, $y = r \cdot \sin(\phi)$ mit $0 \leq r < \infty$, $0 \leq \phi < 2\pi$. Es gilt (Trafo-Satz) $dx\, dy = r\, dr\, d\phi$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx\, dy = \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-\frac{r^2}{2}} r\, dr\, d\phi$$
$$= \int_{0}^{\infty} e^{-\frac{r^2}{2}} r\, dr = -e^{-\frac{r^2}{2}} \Big|_0^{\infty} = 1.$$

Definition 9.9 (Quantilfunktion und Quantile)

Sei F_X Verteilungsfunktion einer (stetigen oder diskreten) Verteilung, dann ist die zugehörige **Quantilfunktion** $F_X^{-1}(p) : (0,1] \to \mathbb{R}$ *definiert durch die verallgemeinerte Inverse* $F_X^{-1}(p) = \inf\{z \in \mathbb{R} \mid F_X(z) \geq p\}$.
$F_X^{-1}(p)$ *heißt* p-**Quantil der Verteilung** F_X.

Bemerkung: In anderen Worten besagt obige Definition, dass das p-Quantil $F_X^{-1}(p)$ die kleinste Zahl $z_0 \in \mathbb{R}$ ist, an der die Verteilungsfunktion $F_X(z)$ den Wert p erstmalig erreicht oder überschreitet.

Wegen der rechtsseitigen Stetigkeit von Verteilungsfunktionen ist das p-Quantil für alle $p \in (0, 1]$ eindeutig bestimmt. (Es gibt auch andere Quantildefinitionen, bei denen p-Quantile nicht immer eindeutig sind.)

Falls die Verteilungsfunktion F_X stetig und streng monoton wachsend ist, ist die Quantilfunktion F_X^{-1} die aus der Analysis bekannte gewöhnliche Inverse oder Umkehrfunktion. In diesem Fall gilt $F_X\big(F_X^{-1}(p)\big) = p$.

Beispiel 9.10

a) **Quantilfunktion der Gleichverteilung** $U([a, b])$

Die Verteilungsfunktion $F_{U([a,b])}$ ist auf $[a, b]$ stetig und streng monoton wachsend, also ist $F_{U([a,b])}^{-1}$ auf $[a, b]$ die gewöhnliche Umkehrfunktion. Auflösen von

$$p = F_{U([a,b])}(z) = \frac{z-a}{b-a} \quad \text{ergibt} \quad z = p(b-a) + a,$$

$F_{U([a,b])}^{-1}(p) = p(b-a) + a.$

Aus Konsistenzgründen definiert man $F_{U([a,b])}^{-1}(0) =: a$.

b) Die p-Quantile $F^{-1}_{N(0,1)}(p)$ der Standard-Normalverteilung kann man den Tabellen mit den Werten der Verteilungsfunktion entnehmen.

Aus $\Phi\big(F^{-1}_{N(0,1)}(p)\big) = p = F_{N(\mu,\sigma^2)}\big(F^{-1}_{N(\mu,\sigma^2)}(p)\big) = \Phi\left(\frac{F^{-1}_{N(\mu,\sigma^2)}(p)-\mu}{\sigma}\right)$

folgt $F^{-1}_{N(\mu,\sigma^2)}(p) = \sigma F^{-1}_{N(0,1)}(p) + \mu$,

d.h. die Quantile der $N(\mu,\sigma^2)$-Verteilung erhält man durch lineare Transformation der Quantile von $N(0,1)$.

Zur Erzeugung von Zufallszahlen beliebiger Verteilungen muss man im Prinzip nur unabhängige, nach $U([0,1])$ verteilte Zufallszahlen generieren, denn es gilt

Satz 9.11 (Simulationslemma)

Sei X eine reellwertige Zufallsvariable mit (stetiger oder diskreter) Verteilung $\mathbb{P}_X$ und zugehöriger Verteilungsfunktion F_X sowie Quantilfunktion F_X^{-1}.
Ist $U \sim U([0,1])$ eine auf $[0,1]$ gleichverteilte Zufallsvariable, dann ist $Y = F_X^{-1}(U)$ genauso verteilt wie X, d.h. es gilt $\mathbb{P}_Y = \mathbb{P}_X$ bzw. $F_Y = F_X$.

Beweis: Nach Definition 9.9 gilt $F_X\big(F_X^{-1}(p)\big) \geq p$. Ist also $F_X^{-1}(U) \leq x$, so gilt wegen der Monotonie von F_X, dass $F_X\big(F_X^{-1}(U)\big) \leq F_X(x)$, also auch $U \leq F_X\big(F_X^{-1}(U)\big) \leq F_X(x)$. Ist umgekehrt $U \leq F_X(x)$, so muss gelten $x \geq \inf\{z \in \mathbb{R} \mid F_X(z) \geq U\} = F_X^{-1}(U)$.
Zusammen folgt $F_X^{-1}(U) \leq x \iff U \leq F_X(x)$ und damit $\{Y \leq x\} = \{F_X^{-1}(U) \leq x\} = \{U \leq F_X(x)\}$, so dass

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}\big(U \leq F_X(x)\big) = F_{U([0,1])}\big(F_X(x)\big) = F_X(x),$$

also haben Y und X die gleiche Verteilungsfunktion, so dass nach Bemerkung 9.7 gilt $\mathbb{P}_Y = \mathbb{P}_X$. □

Bemerkung: Sind $U_1, \ldots, U_n$ unabhängige, identisch nach $U([0,1])$ verteilte Zufallsvariablen, so sind nach Satz 9.11 und Übungsblatt 4, Aufgabe 3 a), $F_X^{-1}(U_1), \ldots, F_X^{-1}(U_n)$ unabhängig und identisch nach $\mathbb{P}_X$ verteilt. Ist $u_1, \ldots, u_n$ eine (simulierte) Realisierung von $U_1, \ldots, U_n$, so sind $F_X{}^-1(u_1), \ldots, F_X^{-1}(u_n)$ n unabhängige Realisierungen (simulierte Werte) von X. Das erklärt den Namen „Simulationslemma“.

Beispiel 9.12

a) **Simulation binomialverteilter Zufallsgrößen:** Sei X eine Bernoulli-Variable mit $\mathbb{P}(X = 1) = p_0$, $\mathbb{P}(X = 0) = 1 - p_0$, dann hat X die Verteilungsfunktion

$$F_X(z) = \begin{cases} 0, & z < 0, \\ 1 - p_0, & 0 \leq z < 1, \\ 1, & z \geq 1. \end{cases}$$

Nach Definition 9.9 ist daher die 0 das p-Quantil für alle $0 \leq p \leq 1 - p_0$ und die 1 das p-Quantil für alle $1 - p_0 < p \leq 1$, d.h. die Quantilfunktion lässt sich schreiben als $F_X^{-1}(p) = \mathbb{1}_{(1-p_0,1]}(p)$.

Sind also $U_1, \ldots, U_n$ unabhängige, identisch nach $U([0,1])$ verteilte Zufallsvariablen, so sind $\mathbb{1}_{(1-p_0,1]}(U_1), \ldots, \mathbb{1}_{(1-p_0,1]}(U_n)$ unabhängige, identisch verteilte Bernoulli-Variablen und damit (vgl. Beispiel 8.16) $X = \sum_{i=1}^n \mathbb{1}_{(1-p_0,1]}(U_i)$ eine binomialverteilte Zufallsvariable zu den Parametern n und p_0 ($X \sim B(n, p_0)$).

Beispiel 9.12 (Forts.)

b) **Fairer Würfelwurf:** Diesen kann man nach dem gleichen Prinzip wie bei Bernoulli-Variablen simulieren. Hier ist die 1 das p-Quantil für alle $0 \le p \le \frac{1}{6}$, die 2 das p-Quantil für alle $\frac{1}{6} < p \le \frac{2}{6}$ usw.

Da $\frac{j-1}{6} < U \le \frac{j}{6}$ für $1 \le j \le 6 \iff j - 1 < 6U \le j$, ist $X = \lceil 6U \rceil$ mit $\lceil x \rceil := \min\{n \in \mathbb{Z} \mid n \ge x\}$ in Verteilung ein fairer Würfel.

Oft gibt es einfachere Alternativen zur Simulation von Zufallszahlen mit vorgegebener Verteilung als Satz 9.11, wie schon Beispiel 9.12 oben zeigt. Für die Normalverteilung, deren Verteilungs- und damit auch Quantilfunktion nicht in geschlossener Form angebbar ist, kann man den folgenden Algorithmus verwenden, den wir ohne Beweis angeben:

Satz 9.13 (Box-Muller-Methode)

Seien U_1, U_2 unabhängige, identisch $U([0,1])$-verteilte Zufallsgrößen und X_1, X_2 definiert durch

$$X_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2), \qquad X_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2).$$

Dann sind X_1, X_2 unabhängig und identisch $N(0,1)$-verteilt.

Kenngrößen stetiger Verteilungen

Definition 10.1

Sei X eine reelle Zufallsvariable mit stetiger Verteilung $\mathbb{P}_X$ und zugehöriger Dichte f_X, dann ist der Erwartungswert von X definiert durch

$$\mathbb{E}[X] := \int_{\mathbb{R}} y\, f_X(y)\, dy, \quad \textit{sofern} \quad \int_{\mathbb{R}} |y|\, f_X(y)\, dy < \infty,$$

und allgemeiner das r-te (bzw. das r-te absolute) Moment von X durch

$$\mathbb{E}[X^r] := \int_{\mathbb{R}} y^r\, f_X(y)\, dy \quad \textit{bzw.} \quad \mathbb{E}[|X|^r] := \int_{\mathbb{R}} |y|^r\, f_X(y)\, dy.$$

Die Varianz von X wird wie im diskreten Fall definiert durch

$$\mathrm{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{\mathbb{R}} (y - \mathbb{E}[X])^2\, f_X(y)\, dy.$$

Bemerkung 10.2

Die Sätze 8.3, 8.4 und 8.5 gelten genauso auch für Zufallsvariablen mit stetigen Verteilungen:

Zu Satz 8.3: $\mathbb{E}[aX] = a\mathbb{E}[X]$ folgt direkt aus der Lineraiät des Integrals, die Gleichung $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ zeigen wir später.

Ist $X \geq 0$, gilt für *alle* integrierbaren Mengen $A \subseteq \mathbb{R}_-$, dass $0 = \mathbb{P}(X \in A) = \int_A f_X(y)\,dy$, also muss $f_X(y) \equiv 0$ für $y \in \mathbb{R}_-$ gelten, folglich ist $\mathbb{E}[X] = \int_{\mathbb{R}} y\, f_X(y)\,dy = \int_{\mathbb{R}_+} y\, f_X(y)\,dy \geq 0$. Daraus folgt wie im Beweis von Satz 8.3 $\mathbb{E}[X] \geq \mathbb{E}[Y]$, falls $X \geq Y$.

Ferner ist $\mathbb{E}[\mathbb{1}_A] = \int_{\mathbb{R}} \mathbb{1}_A(y)\, f_X(y)\,dy = \int_A f_X(y)\,dy = \mathbb{P}_X(A)$.

Zu Satz 8.4: Der Beweis benutzte nur die Linearität des Erwartungswertes und gilt daher unverändert auch im stetigen Fall.

Zu Satz 8.5: Die Eigenschaften ergeben sich vollkommen analog zum diskreten Fall aus der Linearität des Integrals und des Erwartungswertes.

Beispiel 10.3

a) **Gleichverteilung** $U([a,b])$**:** Für den Erwartungswert einer gleichverteilten Zufallsvariable $X \sim U([a,b])$ erhält man

$$\begin{aligned}\mathbb{E}[X] &= \int_{\mathbb{R}} y\, f_{U([a,b])}(y)\, dy = \int_{\mathbb{R}} y\, \frac{1}{b-a} \mathbb{1}_{[a,b]}(y)\, dy = \frac{1}{b-a} \int_a^b y\, dy \\ &= \frac{1}{b-a}\, \frac{y^2}{2}\bigg|_a^b = \frac{1}{2}\, \frac{b^2-a^2}{b-a} = \frac{b+a}{2},\end{aligned}$$

also ein zum diskreten Fall (Laplace-Verteilung, vgl. Beispiel 8.6 a)) analoges Ergebnis. Für die Varianz ergibt sich

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}[(X-\mathbb{E}(X))^2] = \int_{\mathbb{R}} \left(y - \frac{a+b}{2}\right)^2 f_{U([a,b])}(y)\, dy \\ &= \frac{1}{b-a} \int_a^b \left(y - \frac{a+b}{2}\right)^2 dy = \frac{1}{3}\, \frac{1}{b-a} \left(y - \frac{a+b}{2}\right)^3 \bigg|_a^b\end{aligned}$$

$$= \frac{1}{3}\frac{1}{b-a}\left[\left(\frac{b}{2}-\frac{a}{2}\right)^3-\left(\frac{a}{2}-\frac{b}{2}\right)^3\right] = \frac{2}{3}\frac{1}{b-a}\left(\frac{b}{2}-\frac{a}{2}\right)^3$$

$$= \frac{1}{12}(b-a)^2.$$

Im Vergleich zur diskreten Laplace-Verteilung fällt hier der Summand $\frac{a-b}{6}$ weg.

b) **Exponentialverteilung** $\mathrm{Exp}(\lambda)$**:** Für $X \sim \mathrm{Exp}(\lambda)$ ist

$$\mathbb{E}[X] = \int_{\mathbb{R}} y\, f_{\mathrm{Exp}(\lambda)}(y)\, dy = \int_0^\infty y \cdot \lambda\, e^{-\lambda y}\, dy$$

$$= -y \cdot e^{-\lambda y}\Big|_0^{+\infty} + \int_0^\infty e^{-\lambda y}\, dy = -\frac{1}{\lambda} e^{-\lambda y}\Big|_0^{+\infty} = \frac{1}{\lambda}.$$

Das ist auch plausibel, wenn man sich daran erinnert, dass die Exponentialverteilung eine stetige Wartezeitverteilung ist und λ als Ereignisrate aufgefasst werden kann: Je größer λ, desto mehr Ereignisse passieren pro Zeiteinheit, also sollte die (durchschnittliche) Wartezeit bis zum nächsten Ereignis kleiner werden.

Für das zweite Moment ergibt sich

$$\begin{aligned}\mathbb{E}[X^2] &= \int_0^\infty y^2 \cdot \lambda\, e^{-\lambda y}\, dy = -y^2 \cdot e^{-\lambda y}\Big|_0^{+\infty} + \int_0^\infty 2y \cdot e^{-\lambda y}\, dy \\ &= -\frac{2y}{\lambda} \cdot e^{-\lambda y}\Big|_0^{+\infty} + \frac{2}{\lambda}\int_0^\infty e^{-\lambda y}\, dy = \frac{2}{\lambda} \cdot \frac{1}{\lambda},\end{aligned}$$

also ist $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$.

c) **Normalverteilung** $N(\mu, \sigma^2)$**:** Für $X \sim N(\mu, \sigma^2)$ ist

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{+\infty} y \cdot f_{N(\mu,\sigma^2)}(y)\, dy = \int_{-\infty}^{+\infty} y \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy \\ &\underset{z=y-\mu}{=} \int_{-\infty}^{+\infty} (z+\mu) \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz \\ &= \underbrace{\int_{-\infty}^{+\infty} z \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz}_{=0 \text{ (Integrand ungerade: } g(-z) = -g(z))} + \mu \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz}_{=1 \text{ (Dichteeigenschaft!)}} = \mu\end{aligned}$$

Für das zweite Moment erhält man

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} y^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(y-\mu)^2}{2\sigma^2}}\, dy \\
&\underset{z=y-\mu}{=} \int_{-\infty}^{+\infty} (z+\mu)^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz \\
&= \int_{-\infty}^{+\infty} (z^2 + 2\mu z + \mu^2) \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz \\
&= \int_{-\infty}^{+\infty} z^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz + \mu^2
\end{aligned}
$$

Partielle Integration liefert wegen $\left(e^{-\frac{z^2}{2\sigma^2}}\right)' = -\frac{z}{\sigma^2}\, e^{-\frac{z^2}{2\sigma^2}}$

$$
\begin{aligned}
\int_{-\infty}^{+\infty} z^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz &= \int_{-\infty}^{+\infty} z \cdot \frac{z}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz \\
&= -z \cdot \frac{\sigma^2}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\Big|_{-\infty}^{+\infty} + \sigma^2 \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{z^2}{2\sigma^2}}\, dz = 0 + \sigma^2
\end{aligned}
$$

d.h. $\mathbb{E}[X^2] = \sigma^2 + \mu^2$ und somit
$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$.

Die Parameter der Normalverteilung $N(\mu, \sigma^2)$ geben also direkt Erwartungswert und Varianz an. Für $X \sim N(\mu, \sigma^2)$ ist somit die standardisierte Zufallsvariable $X^* = \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]}} = \frac{X-\mu}{\sigma}$ standard-normalverteilt, d.h. $X^* \sim N(0,1)$.

Damit ergibt sich eine alternative Herleitung der Gleichung auf Folie 105: Ist $X \sim N(\mu, \sigma^2)$, dann ist

$$\begin{aligned}
\mathbb{P}(a \leq X \leq b) &= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= \mathbb{P}\left(\frac{a-\mu}{\sigma} \leq X^* \leq \frac{b-\mu}{\sigma}\right) \\
&\underset{X^* \sim N(0,1)}{=} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, dy = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)
\end{aligned}$$

Definition 10.4 (Stetige gemeinsame Verteilung)

Die gemeinsame Verteilung $\mathbb{P}_{(X_1,\ldots,X_n)}$ *der reellen Zufallsvariablen* $X_1, \ldots, X_n$ *heißt* **stetige gemeinsame Verteilung**, *falls eine nicht-negative Funktion* $f_{(X_1,\ldots,X_n)} : \mathbb{R}^n \to [0, \infty)$ *existiert, so dass für jede beliebige Auswahl von Intervallen* $[a_1, b_1], \ldots, [a_n, b_n]$ *gilt*

$$\begin{aligned}\mathbb{P}(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) &= \mathbb{P}_{(X_1,\ldots,X_n)}([a_1, b_1] \times \cdots \times [a_n, b_n]) \\ &= \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_{(X_1,\ldots,X_n)}(y_1, \ldots, y_n)\, dy_1 \ldots dy_n.\end{aligned}$$

$f_{(X_1,\ldots,X_n)}$ *heißt dann Dichte der gemeinsamen Verteilung* $\mathbb{P}_{(X_1,\ldots,X_n)}$.

Bemerkung 10.5

a) Wegen

$$\begin{aligned}1 &= \mathbb{P}(X_1 \in \mathbb{R}, \ldots, X_n \in \mathbb{R}) = \mathbb{P}_{(X_1,\ldots,X_n)}(\mathbb{R} \times \cdots \times \mathbb{R}) \\ &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_{(X_1,\ldots,X_n)}(y_1, \ldots, y_n)\, dy_1 \ldots dy_n\end{aligned}$$

impliziert die obige Definition auch die Normiertheit der Dichte.

b) Sind $X_1, \dots, X_n$ reelle Zufallsvariablen und $g : \mathbb{R}^n \to \mathbb{R}$ eine (stetige) Funktion, so ist auch $g(X_1, \dots, X_n)$ eine reelle Zufallsvariable. Ist die gemeinsame Verteilung $\mathbb{P}_{(X_1,\dots,X_n)}$ stetig und hat die Dichte $f_{(X_1,\dots,X_n)}$, erhält man den Erwartungswert von $g(X_1, \dots, X_n)$ durch

$$\mathbb{E}[g(X_1, \dots, X_n)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y_1, \dots, y_n)\, f_{(X_1,\dots,X_n)}(y_1, \dots, y_n)\, dy_1 \dots dy_n,$$

falls $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |g(y_1, \dots, y_n)|\, f_{(X_1,\dots,X_n)}(y_1, \dots, y_n)\, dy_1 \dots dy_n < \infty$.

Wir betrachten den Fall $n = 2$ genauer: Für beliebige $a_1, a_2 \in \mathbb{R}$ gilt nach Definition 10.4

$$\begin{aligned}
\mathbb{P}(X_1 \le a_1, X_2 \le a_2) &= \mathbb{P}_{(X_1,X_2)}((-\infty, a_1] \times (-\infty, a_2]) \\
&= \int_{-\infty}^{a_2} \int_{-\infty}^{a_1} f_{(X_1,X_2)}(y_1, y_2)\, dy_1 dy_2 \\
&\underset{\text{Fubini}}{=} \int_{-\infty}^{a_1} \left(\int_{-\infty}^{a_2} f_{(X_1,X_2)}(y_1, y_2)\, dy_2 \right) dy_1
\end{aligned}$$

(der Satz von Fubini ist anwendbar, da die Dichte $f_{(X_1,X_2)}$ nicht-negativ ist).

Damit folgt insbesondere

$$\begin{aligned} F_{X_1}(a_1) &= \mathbb{P}(X_1 \leq a_1) = \mathbb{P}(X_1 \leq a_1, X_2 \in \mathbb{R}) = \mathbb{P}_{(X_1,X_2)}((-\infty, a_1] \times (-\infty, \infty)) \\ &= \int_{-\infty}^{a_1} \left(\int_{-\infty}^{\infty} f_{(X_1,X_2)}(y_1, y_2)\, dy_2 \right) dy_1 = \int_{-\infty}^{a_1} f_{X_1}(y_1)\, dy_1, \end{aligned}$$

d.h. die Verteilung $\mathbb{P}_{X_1}$ ist ebenfalls stetig und hat die sog. *Randdichte* $f_{X_1}(y_1) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(y_1, y_2)\, dy_2$.

Analog folgt, dass auch $\mathbb{P}_{X_2}$ stetig ist und die Dichte $f_{X_2}(y_2) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(y_1, y_2)\, dy_1$ hat.

Man erinnere sich, dass wir bei diskreten gemeinsamen Verteilungen die Randverteilungen durch Aufsummieren erhalten haben (vgl. Folie 79). Bei stetigen gemeinsamen Verteilungen erhalten wir nun analog die Randdichten durch Ausintegrieren der gemeinsamen Dichte nach den anderen Variablen (das gilt genauso auch im Fall $n > 2$).

Damit können wir nun auch die Gleichung $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ für den Fall zeigen, dass die gemeinsame Verteilung $\mathbb{P}_{(X,Y)}$ stetig ist: Mit $g(x,y) = x+y$ erhalten wir aus Bemerkung 10.5 b)

$$\mathbb{E}[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y_1+y_2)\, f_{(X,Y)}(y_1,y_2)\, dy_1\, dy_2$$

$$= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} y_1\, f_{(X,Y)}(y_1,y_2)\, dy_1\, dy_2 + \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} y_2\, f_{(X,Y)}(y_1,y_2)\, dy_1\, dy_2$$

$$= \int\limits_{-\infty}^{\infty} y_1 \left(\int\limits_{-\infty}^{\infty} f_{(X,Y)}(y_1,y_2)\, dy_2 \right) dy_1 + \int\limits_{-\infty}^{\infty} y_2 \left(\int\limits_{-\infty}^{\infty} f_{(X,Y)}(y_1,y_2)\, dy_1 \right) dy_2$$

$$= \int_{-\infty}^{\infty} y_1\, f_X(y_1)\, dy_1 + \int_{-\infty}^{\infty} y_2\, f_Y(y_2)\, dy_2 = \mathbb{E}[X] + \mathbb{E}[Y].$$

Die Definition der Unabhängigkeit von Zufallsvariablen (vgl. Definition 7.5) gilt genau so auch im stetigen Fall; man muss hier lediglich die Mengen $A_i \in \mathcal{P}(X_i(\Omega))$ durch integrierbare Mengen $B_i \subseteq \mathbb{R}$ ersetzen. Es genügt, Intervalle $B_i = [a_i, b_i]$ zu betrachten.

Satz 10.6 (Unabhängigkeit und Produktdichte)

Seien $X_1, \dots, X_n$ reelle Zufallsvariablen mit stetiger gemeinsamer Verteilung $\mathbb{P}_{(X_1,\dots,X_n)}$ und zugehöriger Dichte $f_{(X_1,\dots,X_n)}$. Dann sind $X_1, \dots, X_n$ genau dann unabhängig, wenn die gemeinsame Dichte gleich dem Produkt der Randdichten ist, d.h. wenn gilt

$$f_{(X_1,\dots,X_n)}(y_1, \dots, y_n) = \prod_{i=1}^{n} f_{X_i}(y_i).$$

Beweis: Sind $X_1, \dots, X_n$ unabhängig, gilt

$$\begin{aligned}
\mathbb{P}(a_1 \le X_1 \le b_1, \dots, a_n \le X_n \le b_n) &= \mathbb{P}_{(X_1,\dots,X_n)}([a_1, b_1] \times \dots \times [a_n, b_n]) \\
&\underset{\text{Unabh.}}{=} \prod_{i=1}^{n} \mathbb{P}_{X_i}([a_i, b_i]) = \prod_{i=1}^{n} \int_{a_i}^{b_i} f_{X_i}(y_i)\, dy_i \\
&= \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_{X_1}(y_1) \cdot \ldots \cdot f_{X_n}(y_n)\, dy_1 \dots dy_n,
\end{aligned}$$

also ist $f_{(X_1,\dots,X_n)}(y_1, \dots, y_n) = f_{X_1}(y_1) \cdot \ldots \cdot f_{X_n}(y_n)$.
Die Umkehrung (Produktdichte $\Rightarrow$ Unabhängigkeit) folgt analog. $\square$

Bemerkung 10.7

Aus Satz 10.6 folgt unmittelbar, dass die Produktformel (Satz 8.12) auch im Fall stetig gemeinsam verteilter, unabhängiger Zufallsvariablen gilt:

$$
\begin{aligned}
\mathbb{E}[XY] &\underset{10.5\text{ b)}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2\, f_{(X,Y)}(y_1, y_2)\, dy_1 dy_2 \\
&\underset{\text{Unabh., }10.6}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2\, f_X(y_1) f_Y(y_2)\, dy_1 dy_2 \\
&= \int_{-\infty}^{\infty} y_2\, f_Y(y_2) \underbrace{\left(\int_{-\infty}^{\infty} y_1\, f_X(y_1)\, dy_1\right)}_{=\mathbb{E}[X]} dy_2 = \mathbb{E}[X] \cdot \int_{-\infty}^{\infty} y_2\, f_Y(y_2) dy_2 \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y].
\end{aligned}
$$

Da die Definition der Kovarianz $\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ genau so auch im stetigen Fall gilt, bleibt die Aussage „Unabhängigkeit impliziert Unkorreliertheit“ auch hier korrekt.
Aus Bemerkung 10.5 b) und der Linearität des Integrals folgt, dass auch Korollar 8.15 und Satz 8.18 im stetigen Fall unverändert gültig bleiben.

Faltung von Verteilungen

Definition 11.1 (Faltung)

Sind X und Y zwei unabhängige, reelle Zufallsvariablen mit Verteilungen $\mathbb{P}_X$ und $\mathbb{P}_Y$, dann heißt die Verteilung $\mathbb{P}_{X+Y}$ ihrer Summe die **Faltung von $\mathbb{P}_X$ und $\mathbb{P}_Y$**.

Sind X und Y unabhängige, diskrete Zufallsvariablen mit Wertebereichen $X(\Omega) = \{x_1, x_2, \ldots\}$ und $Y(\Omega) = \{y_1, y_2, \ldots\}$, so ist auch der Wertebereich $(X+Y)(\Omega) = \{z_1, z_2, \ldots\}$ ihrer Summe höchstens abzählbar, und es gilt

$$\begin{aligned}\mathbb{P}_{X+Y}(\{z_k\}) = \mathbb{P}(X+Y=z_k) &= \sum_{i\geq 1} \mathbb{P}(X = x_i, Y = z_k - x_i) \\ &\underset{\text{Unabh.}}{=} \sum_{i\geq 1} \mathbb{P}(X = x_i)\mathbb{P}(Y = z_k - x_i) = \sum_{i\geq 1} \mathbb{P}_X(\{x_i\})\mathbb{P}_Y(\{z_k - x_i\}).\end{aligned}$$

Beispiel 11.2

a) **Faltung zweier Laplace-Verteilungen:** Seien X und Y unabhängig und identisch Laplace-verteilt auf $\{1, \ldots, n\}$ mit $1 < n \in \mathbb{N}$, d.h. $\mathbb{P}_X(\{j\}) = \mathbb{P}_Y(\{j\}) = \frac{1}{n}$, $j \in \{1, \ldots, n\}$.

Beispiel 11.2 (Forts.)

Dann gilt für $2 \leq k \leq 2n$

$$\begin{aligned}
\mathbb{P}_{X+Y}(\{k\}) &= \sum_{i=1}^{n} \mathbb{P}_X(\{i\})\mathbb{P}_Y(\{k-i\}) \\
&= \frac{1}{n^2} \#\{(i,j) \in \{1,\ldots,n\}^2 \mid i+j=k\} \\
&= \frac{n-|k-n-1|}{n^2}
\end{aligned}$$

Für $n = 6$ entspricht dies der Verteilung der Augensumme zweier unabhängiger Würfelwürfe, die wir bereits in Beispiel 7.3 a) kennengelernt haben.

b) **Faltung zweier Poissonverteilungen:** Seien X und Y unabhängig mit $X \sim \text{Pois}(\lambda_1)$ und $Y \sim \text{Pois}(\lambda_2)$, dann ergibt sich die Verteilung ihrer Summe unter Beachtung von $X(\Omega) = Y(\Omega) = \mathbb{N}_0$ für $k \geq 0$ zu

$$\begin{aligned}\mathbb{P}_{X+Y}(\{k\}) &= \sum_{i=0}^{k} \mathbb{P}_X(\{i\})\mathbb{P}_Y(\{k-i\}) = \sum_{i=0}^{k} \frac{\lambda_1^i}{i!}\, e^{-\lambda_1} \cdot \frac{\lambda_2^{k-i}}{(k-i)!}\, e^{-\lambda_2} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_1^i \lambda_2^{k-i} \underset{\text{Korollar 4.4}}{=} \frac{(\lambda_1+\lambda_2)^k}{k!}\, e^{-(\lambda_1+\lambda_2)} \\ &= P_{\lambda_1+\lambda_2}(\{k\}),\end{aligned}$$

d.h. die Summe zweier unabhängiger Poisson-verteilter Zufallsvariablen zu den Parametern λ_1 und λ_2 ist ebenfalls Poisson-verteilt zum Parameter $\lambda_1 + \lambda_2$.

Die Klasse der Poisson-Verteilungen ist also *faltungsstabil*.

Satz 11.3 (Faltungsformel)

Seien X, Y unabhängige, reelle Zufallsvariablen mit stetigen Verteilungen $\mathbb{P}_X, \mathbb{P}_Y$ und zugehörigen Dichten f_X und f_Y, dann ist auch die Verteilung $\mathbb{P}_{X+Y}$ ihrer Summe stetig und besitzt die Dichte

$$f_{X+Y}(y) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(y-x)\, dx.$$

Beispiel 11.4 (Faltung von Normalverteilungen)

Sind $X \sim N(\mu_1, \sigma_1^2)$ und $Y \sim N(\mu_2, \sigma_2^2)$ zwei unabhängige, normalverteilte Zufallsvariablen, so ist auch ihre Summe normalverteilt mit $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$, d.h. auch die Klasse der Normalverteilungen ist faltungsstabil.

Beweis: Es genügt, die Behauptung für $\mu_1 = \mu_2 = 0$ zu zeigen, denn seien $\bar{X} \sim N(0, \sigma_1^2)$ und $\bar{Y} \sim N(0, \sigma_2^2)$, so gilt $\bar{X} + \mu_1 \sim N(\mu_1, \sigma_1^2)$, $\bar{Y} + \mu_2 \sim N(\mu_2, \sigma_2^2)$ und damit

$$(\bar{X} + \mu_1) + (\bar{Y} + \mu_2) = \underbrace{(\bar{X} + \bar{Y})}_{\overset{!}{\sim} N(0, \sigma_1^2 + \sigma_2^2)} + (\mu_1 + \mu_2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Nach der Faltungsformel gilt

$$\begin{aligned} f_{\bar{X}+\bar{Y}}(y) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}}\, e^{-\frac{x^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}}\, e^{-\frac{(y-x)^2}{2\sigma_2^2}}\, dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{(\sigma_1^2+\sigma_2^2)x^2}{\sigma_1^2\sigma_2^2} - \frac{2xy}{\sigma_2^2} + \frac{y^2}{\sigma_2^2}\right)}\, dx. \end{aligned}$$

Setze nun $z := \frac{\sqrt{\sigma_1^2+\sigma_2^2}x}{\sigma_1\sigma_2} - \frac{y\sigma_1}{\sigma_2\sqrt{\sigma_1^2+\sigma_2^2}}$, dann ist

$$\begin{aligned} z^2 + \frac{y^2}{\sigma_1^2+\sigma_2^2} &= \frac{(\sigma_1^2+\sigma_2^2)x^2}{\sigma_1^2\sigma_2^2} - \frac{2xy}{\sigma_2^2} + \frac{y^2\sigma_1^2}{\sigma_2^2(\sigma_1^2+\sigma_2^2)} + \frac{y^2\sigma_2^2}{\sigma_2^2(\sigma_1^2+\sigma_2^2)} \\ &= \frac{(\sigma_1^2+\sigma_2^2)x^2}{\sigma_1^2\sigma_2^2} - \frac{2xy}{\sigma_2^2} + \frac{y^2}{\sigma_2^2} \end{aligned}$$

und $dx = \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\,dz$, also ist

$$\begin{aligned} f_{\bar{X}+\bar{Y}}(y) &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(z^2+\frac{y^2}{\sigma_1^2+\sigma_2^2}\right)}\,\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\,dz \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\,\frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}}\,e^{-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}}\,\underbrace{\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}\,dz}_{=\sqrt{2\pi}\ \text{(Normierung von } N(0,1)!)} \\ &= \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\,e^{-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}} \end{aligned}$$

was der Dichte von $N(0,\sigma_1^2+\sigma_2^2)$ entspricht. $\square$

Bemerkung 11.5

Aus Aufgabe 4 von Blatt 8 kann man folgern, dass auch die Klasse der Gammaverteilungen (mit festem Parameter b) stabil unter Faltungen ist.

Dies gilt aber bei weitem nicht für alle Klassen stetiger Verteilungen: Seien z.B. X und Y unabhängig und identisch gleichverteilt auf $[0,1]$, dann ist die Dichte der Verteilung ihrer Summe nach der Faltungsformel

$$f_{X+Y}(y) = \int_{-\infty}^{\infty} \mathbb{1}_{[0,1]}(x) \cdot \mathbb{1}_{[0,1]}(y-x)\, dx = \int_0^1 \mathbb{1}_{[0,1]}(y-x)\, dx$$

Da $y - x \geq 0$ genau dann, wenn $x \leq y$, und $y - x \leq 1$ genau dann, wenn $x \geq y - 1$, folgt

$$f_{X+Y}(y) = \int_{\max(y-1,0)}^{\min(y,1)} 1\, dx = \begin{cases} \int_0^y 1\, dx = y, & 0 \leq y \leq 1, \\ \int_{y-1}^1 1\, dx = 1 - (y-1) = 2 - y, & 1 \leq y \leq 2, \\ 0, & \text{sonst.} \end{cases}$$

Also ist $X + Y$ *nicht* gleichverteilt.

Satz 12.1 (Markov- und Tschebyscheff-Ungleichungen)

Sei X eine reelle Zufallsvariable (mit stetiger oder diskreter Verteilung) und $g : \mathbb{R}_+ \to \mathbb{R}_+$ eine monoton wachsende Funktion mit $g(x) > 0$ für $x > 0$, für die $\mathbb{E}[g(|X|)]$ existiert. Dann gilt

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[g(|X|)]}{g(\varepsilon)} \quad \forall\, \varepsilon > 0. \qquad \text{(Markov-Ungleichung)}$$

Ist $\mathbb{E}[X^2] < \infty$, dann gilt auch

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}[X]}{\varepsilon^2} \quad \forall\, \varepsilon > 0. \quad \text{(Tschebyscheff-Ungleichung)}$$

Beweis: Wegen der Monotonie von g ist $g(|x|) \geq g(\varepsilon)$ und damit $\frac{g(|x|)}{g(\varepsilon)} \geq 1$ für $|x| \geq \varepsilon > 0$. Damit folgt

$$\mathbb{P}(|X| \geq \varepsilon) \underset{8.3\text{ c)},\ 10.2}{=} \mathbb{E}[\mathbb{1}_{\{|X| \geq \varepsilon\}}] \underset{8.3\text{ b)},\ 10.2}{\leq} \mathbb{E}\left[\frac{g(|X|)}{g(\varepsilon)} \mathbb{1}_{\{|X| \geq \varepsilon\}}\right] \leq \mathbb{E}\left[\frac{g(|X|)}{g(\varepsilon)}\right] = \frac{\mathbb{E}[g(|X|)]}{g(\varepsilon)}.$$

Die Tschebyscheff-Ungleichung folgt direkt aus der Markov-Ungleichung durch Einsetzen von $\bar{X} := X - \mathbb{E}[X]$ und $g(x) = x^2$. $\square$

Satz 12.2 (Schwaches Gesetz großer Zahlen)

Sei $(X_i)_{i\geq 1}$ eine Folge reeller, identisch verteilter und paarweise unkorrelierter Zufallsvariablen (mit stetiger oder diskreter Verteilung), die $\mathbb{E}[X_i^2] = \mathbb{E}[X_1^2] < \infty$ erfüllen. Dann gilt für alle $\varepsilon > 0$

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

Beweis: Wegen Satz 8.3 a) und der identischen Verteilung der X_i gilt
$\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_1] = \mathbb{E}[X_1]$.
Wegen der paarweisen Unkorreliertheit gilt nach Satz 8.5 c) und der Bienaymé-Gleichung (Korollar 8.15) ferner
$\mathrm{Var}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_1] = \frac{1}{n}\mathrm{Var}[X_1]$.
Damit folgt aus der Tschebyscheff-Ungleichung für jedes $\varepsilon > 0$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]}{\varepsilon^2} = \frac{1}{n\varepsilon^2}\mathrm{Var}[X_1] \xrightarrow[n\to\infty]{} 0. \ \square$$

Definition 12.3 (Stochastische Konvergenz)

Eine Folge $(X_n)_{n\geq 1}$ *reeller Zufallsvariablen* **konvergiert stochastisch** (*oder auch:* **nach Wahrscheinlichkeit**) *gegen eine reelle Zufallsvariable* X*, falls für jedes* $\varepsilon > 0$ *gilt;* $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$.

Schreibweisen: $\mathbb{P}\text{-}\lim_{n\to\infty} X_n = X$ *oder auch* $X_n \xrightarrow{\mathbb{P}} X$*.*

Mit anderen Worten besagt also das schwache Gesetz großer Zahlen, dass unter den Voraussetzungen von Satz 12.2 das arithmetische Mitel der Zufallsvariablen stochastisch gegen deren Erwartungswert konvergiert.

Bemerkung 12.4

a) Durch Übergang zu Komplementen bzw. Gegenwahrscheinlichkeiten lässt sich die Aussage des SGGZ alternativ auch wie folgt formulieren:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1]\right| < \varepsilon\right) = 1,$$

d.h. für große n ist das arithmetische Mittel $\frac{1}{n}\sum_{i=1}^{n} X_i$ der X_i mit sehr großer Wahrscheinlichkeit sehr nahe (Entfernung kleiner ε) am Erwartungswert.

Aus statistischer Sicht bedeutet das, dass sich der Erwartungswert einer Folge identisch verteilter, unkorrelierter Zufallsvariablen sehr gut durch deren arithmetisches Mittel schätzen lässt.

b) Sind die $(X_i)_{i\geq 1}$ eine Folge unkorrelierter Bernoulli-Variablen mit $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1 - p$ (z.B. mehrfacher Münzwurf oder allgemeiner unabhängige Wiederholungen eines Versuchs mit nur zwei möglichen Ausgängen), dann entspricht das arihmetische Mittel $\frac{1}{n}\sum_{i=1}^{n} X_i$ der relativen Anzahl der Erfolge in den ersten n Versuchen. Nach dem SGGZ gilt $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = p$, d.h. die relativen Erfolgshäufigkeiten konvergieren (stochastisch) gegen die zugrundeliegende Erfolgswahrscheinlichkeit p.

Allgemeiner gilt: Sind $(A_i)_{i\geq 1}$ unabhängige Ereignisse mit gleicher Eintrittswahrscheinlichkeit $\mathbb{P}(A_i) = \mathbb{P}(A_1) = p$, dann besagt das SGGZ wegen $\mathbb{E}[\mathbb{1}_{A_i}] = \mathbb{P}(A_i) = p$, dass $\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{A_i} \xrightarrow{\mathbb{P}} p$.

Anders ausgedrückt lässt sich also die Wahrscheinlichkeit von Ereignissen, deren Eintreten oder Nichteintreten unter gleichen und unabhängiggen Bedingungen beliebig oft beobachtbar ist, sehr genau abschätzen bzw. „messen“.

Ein anderer Typ von stochastischen Grenzwertsätzen behandelt Verteilungskonvergenz. Ein Beispiel dazu haben wir bereits früher kennengelernt:

Satz (Poissons Gesetz der kleinen Zahlen)

Falls $0 < p_n < 1$ *und* $\lim_{n\to\infty} np_n = \lambda > 0$*, dann gilt für jedes* $k = 0, 1, 2, \ldots$

$$\lim_{n\to\infty} b_{n,p_n}(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda} =: P_\lambda(\{k\}).$$

Beispiel 12.5

2% aller Fluggäste, die vorab Plätze reservieren, erscheinen in der Regel nicht zum Abflug. Die Fluggesellschaft weiß dies und verkauft daher 150 Flugtickets für 148 verfügbare Plätze in einem Flugzeug. Wie groß ist die Wahrscheinlichkeit, dass alle zum Abflug erscheinenden Fluggäste einen Platz im Flugzeug bekommen?

Exakte Lösung: Definiere X_i durch

$$X_i = \begin{cases} 1, & \text{falls Ticketinhaber Nummer } i \text{ nicht zum Abflug erscheint,} \\ 0, & \text{sonst.} \end{cases}$$

Dann gilt $\mathbb{P}(X_i = 1) = 0.02$ bzw. $\mathbb{P}(X_i = 0) = 0.98$. Unter der Annahme, dass die Fluggäste unabhängig voneinander zum Abflug erscheinen, ist dann die Anzahl $Y := \sum_{i=1}^{150} X_i$ der nicht erscheinenden Fluggäste binomialverteilt mit $n = 150$ und $p = 0.02$ ($Y \sim B(150, 0.02)$).
Die Wahrscheinlichkeit, dass alle erscheinenden Passagiere einen Platz im Flugzeug bekommen, ist dann

$$\begin{aligned}\mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y = 1) - \mathbb{P}(Y = 0) &= 1 - 150 \cdot 0.02 \cdot 0.98^{149} - 0.98^{150} \\ &\approx 0.804.\end{aligned}$$

Lösung mit Poisson-Approximation: Die Anzahl Y der nicht zum Abflug erscheinenden Passagiere kann näherungsweise als Poisson-verteilt angesehen werden zum Parameter $\lambda = np = 150 \cdot 0.02 = 3$, d.h. $Y \sim \text{Pois}(3)$. Dann ist

$$\begin{aligned}\mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y = 1) - \mathbb{P}(Y = 0) &= 1 - \frac{3^1}{1!} e^{-3} - \frac{3^0}{0!} e^{-3} \\ &= 1 - 4e^{-3} \approx 0.801,\end{aligned}$$

also im Vergleich zu obigem Resultat eine recht gute Näherung.

Im zentralen Grenzwertsatz wird die Konvergenz von Verteilungen standardisierter Summen unabhängiger Zufallsvariablen untersucht. Es zeigt sich, dass diese in vielen Fällen gegen die Standard-Normalverteilung konvergieren.

Satz 12.6 (Zentraler Grenzwertsatz von Lévy)

Seien $(X_i)_{i\geq 1}$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_1^2] < \infty$, $0 < \mathrm{Var}[X_1] = \sigma^2$, dann gilt für die standardisierten Summen $S_n^* = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_1])}{\sigma\sqrt{n}}$ und alle $-\infty \leq a < b \leq \infty$

$$\lim_{n\to\infty} \mathbb{P}(a \leq S_n^* \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = \Phi(b) - \Phi(a).$$

Bemerkung: S_n^* ist standardisiert, denn nach Satz 8.3 a) ist

$$\mathbb{E}[S_n^*] = \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_1])}{\sigma\sqrt{n}}\right] = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n \underbrace{\mathbb{E}[X_i - \mathbb{E}[X_i]]}_{=0} = 0.$$

Wegen der Unabhängigkeit der X_i gilt nach Korollar 8.15 und Satz 8.5 b),c), dass

$$\mathrm{Var}[S_n^*] = \mathrm{Var}\left[\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_1])}{\sigma\sqrt{n}}\right] = \frac{1}{\sigma^2 n}\sum_{i=1}^n \mathrm{Var}[X_i - \mathbb{E}[X_i]]$$

$$= \frac{1}{\sigma^2 n}\sum_{i=1}^n \mathrm{Var}[X_i] = \frac{1}{\sigma^2 n}\sum_{i=1}^n \mathrm{Var}[X_1] = \frac{n\sigma^2}{\sigma^2 n} = 1.$$

Ein Spezialfall des zentralen Grenzwertsatzes, in dem die Zufallsvariablen X_i unabhängige, identisch verteilte Bernoulli-Variablen sind, ist

Korollar 12.7 (Satz von deMoivre-Laplace)

Seien $(X_i)_{i\geq 1}$ unabhängige, identisch verteilte Bernoulli-Variablen mit $\mathbb{P}(X_1 = 1) = p = 1 - \mathbb{P}(X_1 = 0)$, $S_n^* = \frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}}$ und $-\infty \leq a < b \leq \infty$, dann gilt

$$\lim_{n\to\infty} \mathbb{P}(a \leq S_n^* \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx = \Phi(b) - \Phi(a).$$

Bemerkungen: Man beachte, dass beim zentralen Grenzwertsatz mit $\sqrt{n}$ im Nenner von S_n^* skaliert wird! Würde n statt $\sqrt{n}$ im Nenner von S_n^* stehen, so folgte aus dem schwachen Gesetz großer Zahlen, dass

$$\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_1])}{\sigma n} = \frac{1}{\sigma n} \sum_{i=1}^n X_i - \underbrace{\frac{\mathbb{E}[X_1]}{n\sigma}}_{\to 0} \xrightarrow{\mathbb{P}} \frac{\mathbb{E}[X_1]}{\sigma}.$$

In diesem Fall erhielt man also (stochastische) Konvergenz gegen eine Konstante und nicht gegen etwas Normalverteiltes.

Abgesehen von der Skalierung ist die Verteilung $\mathbb{P}_{S_n^*}$ wegen der Unabhängigkeit der X_i nichts anderes als die Faltung der Verteilungen der X_i. Mit anderen Worten besagt also der zentrale Grenzwertsatz, dass die standardisierte bzw. reskalierte Faltung identischer Verteilungen, die ein zweites Moment besitzen, stets (im obigen Sinne) gegen die Standard-Normalverteilung konvergiert.

Der zentrale Grenzwertsatz gilt oft auch dann noch, wenn die X_i nur unabhängig, aber nicht notwendigerweise mehr identisch verteilt sind. In diesem Fall müssen die Verteilungen $\mathbb{P}_{X_i}$ allerdings noch zusätzliche (technische) Bedingungen erfüllen und typischerweise auch höhere Momente $\mathbb{E}[X_i^r]$ mit $r > 2$ besitzen.

Anwendungsbeispiele des ZGWS

Beispiel 12.8 (Qualitätskontrolle)

In einer Firma werden Schrauben in Packungen zu je 1000 Stück abgepackt. Wie groß ist die Wahrscheinlichkeit, dass eine Packung nicht mehr als 1% fehlerhafter Schrauben enthält, wenn erfahrungsgemäß 1% der Gesamtproduktion fehlerhaft ist?

$$X_i = \begin{cases} 1, & \text{falls } i\text{-te Schraube in Packung fehlerhaft,} \\ 0, & \text{sonst.} \end{cases}$$

Dann gilt n.V. $\mathbb{P}(X_i = 1) = 0.01$, und $S_{1000} = \sum_{i=1}^{1000} X_i$ ist die Anzahl der defekten Schrauben in einer Packung. Unter der Annahme, dass defekte Schrauben unabhängig voneinander auftreten, ist $S_{1000} \sim B(1000, 0.01)$, und

$$\mathbb{P}(S_{1000} \leq 10) = \sum_{i=0}^{10} \binom{1000}{i} \cdot 0.01^i \cdot 0.99^{1000-i} \approx 0.5830.$$

Anwendung des ZGWS liefert wegen $\mathbb{E}[S_{1000}] = 1000 \cdot 0.01 = 10$ und $\mathrm{Var}[S_{1000}] = 1000 \cdot 0.01 \cdot 0.99 = 9.9$

$$\mathbb{P}(S_{1000} \le 10) = \mathbb{P}\left(\frac{S_{1000} - 10}{\sqrt{9.9}} \le \frac{10 - 10}{\sqrt{9.9}}\right) = \mathbb{P}(S^*_{1000} \le 0) \approx \Phi(0) = 0.5.$$

Hier ist offensichtlich die Approximation mit Hilfe des ZGWS noch recht ungenau. Eine Verbesserung erreicht man bei diskret verteilten Summen S_n (hier z.B. $S_n \sim B(n, p)$) oft durch die sog. **Stetigkeitskorrekur:**

$$\begin{aligned}\mathbb{P}(a \le S_n \le b) &= \mathbb{P}\left(\frac{a - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \le S_n^* \le \frac{b - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right) \\ &\approx \Phi\left(\frac{b + \mathbf{0.5} - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right) - \Phi\left(\frac{a - \mathbf{0.5} - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}}\right)\end{aligned}$$

In obigem Beispiel liefert diese

$$\mathbb{P}(S^*_{1000} \le 0) \approx \Phi\left(\frac{10 + 0.5 - 10}{\sqrt{9.9}}\right) = \Phi(0.1589) \approx 0.5631,$$

also eine deutlich bessere Abschätzung.

Den Sinn der Stetigkeitskorrektur zeigt folgende Überlegung: Ist $S_n \sim B(n,p)$ mit $0 < p < 1$, dann ist $\mathbb{P}(S_n = k) > 0$ für alle $0 \leq k \leq n$. Wegen

$$\mathbb{P}(S_n = k) = \mathbb{P}(k \leq S_n \leq k) = \mathbb{P}\left(\frac{k-np}{\sqrt{np(1-p)}} \leq S_n^* \leq \frac{k-np}{\sqrt{np(1-p)}}\right)$$

$$\underset{\text{ZGWS}}{\approx} \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right) = 0$$

erhielte man ohne Stetigkitskorrektur stets einen (u.U. ziemlich) falschen Wert. Mit Stetigekeitskorrektur ist dagegen

$$\mathbb{P}(S_n = k) \approx \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right) > 0.$$

Beachte: Stetige Verteilungen haben keine Punktmassen (d.h. $\mathbb{P}_X(\{y\}) = 0$ für alle $y \in \mathbb{R}$), während diskrete Verteilungen nur aus Punktmassen bestehen! Dies muss (zumindest für kleinere n) bei der Approximation von diskreten durch stetige Verteilungen berücksichtigt werden.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Zur Bestimmung von $\Phi(x)$ für $x < 0$ verwende die Symmetrie-Relation $\Phi(x) = 1 - \Phi(-x)$. Beispielsweise ist
$\mathbb{P}(S_n^* \leq -0.33) \approx \Phi(-0.33) = 1 - \Phi(0.33) = 1 - 0.6293 = 0.2707$,
$\mathbb{P}(S_n^* > -0.4) = 1 - \mathbb{P}(S_n^* \leq -0.4) \approx 1 - (1 - \Phi(0.4)) = \Phi(0.4) = 0.6554$.

Beispiel 12.9 (Wahlumfrage)

Wie viele Wähler muss man befragen, um das Wahlergebnis einer Partei mit einer Abweichung von höchstens 1% mit einer Sicherheit von 90% vorherzusagen?

$$X_i = \begin{cases} 1, & \text{falls } i\text{-te befragte Person die Partei wählen will,} \\ 0, & \text{sonst.} \end{cases}$$

Annahme: Die X_i sind unabhängig und identisch verteilt (repräsentative Umfrage).
Ist p_0 das tatsächliche (unbekannte) Wahlergebnis der Partei ein Prozent, gilt $S_n = \sum_{i=1}^n X_i \sim B(n, p_0)$, und das anhand der Umfrage geschätzte Wahlergebnis ist $\hat{p}_n = \frac{S_n}{n}$.
Dann lautet die Fragestellung: Wie groß ist n zu wählen, so dass

$$\mathbb{P}_{p_0}(|\hat{p}_n - p_0| \leq 0.01) \geq 0.9$$

für alle $0 < p_0 < 1$ gilt?

Aus den ZGWS folgt

$$\begin{aligned}
&\mathbb{P}_{p_0}(p_0 - 0.01 \le \hat{p}_n \le p_0 + 0.01) \\
&= \mathbb{P}_{p_0}\left(-\frac{\sqrt{n}\cdot 0.01}{\sqrt{p_0(1-p_0)}} \le S_n^* \le \frac{\sqrt{n}\cdot 0.01}{\sqrt{p_0(1-p_0)}}\right) \\
&\underset{\text{ZGWS}}{\approx} \Phi\left(\frac{\sqrt{n}\cdot 0.01}{\sqrt{p_0(1-p_0)}}\right) - \Phi\left(-\frac{\sqrt{n}\cdot 0.01}{\sqrt{p_0(1-p_0)}}\right) \\
&= 1 - 2\,\Phi\left(-\frac{\sqrt{n}\cdot 0.01}{\sqrt{p_0(1-p_0)}}\right) \overset{!}{\ge} 0.9
\end{aligned}$$

Auflösen der letzten Ungleichung nach n ergibt

$$0.05 \ge \Phi\left(-\frac{\sqrt{n}\cdot 0.01}{\sqrt{p_0(1-p_0)}}\right) \iff \Phi^{-1}(0.05)^2 \ge \frac{n\cdot 0.01^2}{p_0(1-p_0)},$$

d.h. es muss gelten

$$n \ge \frac{\Phi^{-1}(0.05)^2 \cdot p_0(1-p_0)}{0.01^2}.$$

Wegen $p_0(1-p_0) \leq \frac{1}{4}$ mit Gleichheit genau dann, wenn $p_0 = \frac{1}{2}$, folgt, dass eine für alle $0 < p_0 < 1$ gültige untere Schranke für n gegeben ist durch

$$n \geq \frac{\Phi^{-1}(0.05)^2}{4 \cdot 0.01^2} = \frac{(-1.645)^2}{0.0004} = 6765.062$$

Ferner ist $p_0(1-p_0)$ wachsend für $0 < p_0 \leq \frac{1}{2}$ und fallend für $\frac{1}{2} \leq p_0 < 1$. Wenn man aufgrund vergangener Erfahrungswerte recht sicher sein kann, dass der tatsächliche Wähleranteil p_0 nicht höher als z.B. 15% ist, gilt $\max_{p_0 \in (0,0.15]} p_0(1-p_0) = 0.15 \cdot 0.85$, und die untere Schranke für n verringert sich entsprechend zu

$$n \geq \frac{\Phi^{-1}(0.05)^2 \cdot 0.15 \cdot 0.85}{0.01^2} = \frac{(-1.645)^2 \cdot 0.1275}{0.0001} = 3450.182.$$

Man vergleiche diese Zahlen einmal mit der Anzahl der Befragten in den Sonntagsumfragen diverser Sender oder auch anderen Meinungsumfragen!

Beispiel 12.10 (Würfeln)

Wie groß ist (approximativ) die Wahrscheinlichkeit, dass bei wiederholtem Werfen mit einem fairen Würfel die 150. Sechs nach 1000 Würfen noch nicht aufgetreten ist?

Sei T_1 die Wartezeit bis zur ersten Sechs, T_2 die Wartezeit auf die zweite Sechs, *beginnend mit dem ersten Wurf nach* T_1, und entsprechend T_i die Wartezeit auf die i-te Sechs, beginnend ab dem ersten Wurf nach T_{i-1}, dann sind wegen der Unabhängigkeit der Würfelwürfe die Wartezeiten T_i unabhängig und identisch geometrisch verteilt zum Parameter $p = \frac{1}{6}$ ($T_i \sim \text{Geo}(\frac{1}{6})$). Nach Beispiel 8.6 d) ist $\mathbb{E}[T_i] = \frac{1}{p} = 6$ und $\text{Var}[T_i] = \frac{1}{p^2} - \frac{1}{p} = 36 - 6 = 30$.
Damit erhält man nach dem ZGWS für die gesuchte Wahrscheinlichkeit

$$\begin{aligned}\mathbb{P}(T_1 + T_2 + \cdots + T_{150} > 1000) &= \mathbb{P}\left(\frac{\sum_{i=1}^{150} T_i - 150 \cdot 6}{\sqrt{150} \cdot \sqrt{30}} > \frac{1000 - 150 \cdot 6}{\sqrt{150} \cdot \sqrt{30}}\right) \\ &= \mathbb{P}\left(S_n^* > \frac{100}{\sqrt{5} \cdot 30}\right) \approx 1 - \Phi\left(\frac{100}{\sqrt{5} \cdot 30}\right) = 1 - \Phi(1.4907) = 0.068.\end{aligned}$$

Bedingte Verteilungen von Zufallsvariablen

In Definition 6.1 haben wir die bedingte Wahrscheinlichkeit $\mathbb{P}(A|B)$ für zwei Ereignisse $A, B \subseteq \Omega$ definiert als $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Sind nun X_1, X_2 zwei reelle Zufallsvariablen, so erhält man mit $A = \{X_2 \in A_2\}$ und $B = \{X_1 \in A_1\}$

$$\mathbb{P}(X_2 \in A_2 \mid X_1 \in A_1) = \frac{\mathbb{P}(X_1 \in A_1,\, X_2 \in A_2)}{\mathbb{P}(X_1 \in A_1)} = \frac{\mathbb{P}_{(X_1, X_2)}(A_1 \times A_2)}{\mathbb{P}_{X_1}(A_1)},$$

wobei $\mathbb{P}(X_2 \in A_2 \mid X_1 \in A_1) =: 0$, falls $\mathbb{P}(X_1 \in A_1) = 0$.

Definition 13.1 (Bedingte Verteilung)

Die durch $A_2 \mapsto \mathbb{P}(X_2 \in A_2 \mid X_1 \in A_1)$ definierte Abbildung von $\mathcal{B} \to [0,1]$ heißt die **bedingte Verteilung von X_2 gegeben $\{X_1 \in A_1\}$**. *Ist X_1 diskret mit Wertebereich $X_1(\Omega) = \{x_1, x_2, \ldots\}$, dann heißt die Abbildung von $\mathcal{B} \times X_1(\Omega) \to [0,1]$, definiert durch*

$$(A_2, x_i) \mapsto \mathbb{P}(X_2 \in A_2 \mid X_1 = x_i),$$

die **bedingte Verteilung von X_2 gegeben X_1**.

Bemerkung 13.2

a) Sind beide Zufallsvariablen X_1, X_2 diskret mit Wertebereichen $X_1(\Omega) = \{x_1, x_2, \ldots\}$ und $X_2(\Omega) = \{y_1, y_2, \ldots\}$, wird die bedingte Verteilung von X_2 gegeben X_1 als Abbildung von $X_1(\Omega) \times X_2(\Omega) \to [0,1]$ definiert durch

$$(x_i, y_j) \to \mathbb{P}(X_2 = y_j \mid X_1 = x_i)$$

b) Ist X_2 diskret mit Wertebereich $X_2(\Omega) = \{y_1, y_2, \ldots\}$, so gilt

$$\mathbb{P}(X_2 \in A_2 \mid X_1 \in A_1) = \sum_{y_i \in A_2} \mathbb{P}(X_2 = y_i \mid X_1 \in A_1).$$

c) Die Zufallsvariablen X_1, X_2 sind genau dann unabhängig, wenn für alle $A_1 \in \mathcal{P}(X_1(\Omega))$, $A_2 \in \mathcal{P}(X_2(\Omega))$ (diskreter Fall) bzw. $A_1, A_2 \in \mathcal{B}$ (stetiger Fall) gilt $\mathbb{P}(X_2 \in A_2 \mid X_1 \in A_1) = \mathbb{P}(X_2 \in A_2)$.

d) Korollar 6.3 sowie Sätze 6.4 und 6.5 gelten genauso für bedingte Verteilungen.

Beispiel 13.3

Seien X_1 und X_2 die Ergebnisse zweier unabhängiger Würfe mit einem fairen Würfel und $M = \max(X_1, X_2)$. Dann gilt

$$\mathbb{P}(M = j \mid X_1 = j) = \frac{\mathbb{P}(X_1 = j, X_2 \leq j)}{\mathbb{P}(X_1 = j)} \underset{\text{Unabh.}}{=} = \frac{\mathbb{P}(X_1 = j)\mathbb{P}(X_2 \leq j)}{\mathbb{P}(X_1 = j)}$$
$$= \mathbb{P}(X_2 \leq j) = \frac{j}{6}.$$

Für $i < j$ folgt analog

$$\mathbb{P}(M = j \mid X_1 = i) = \frac{\mathbb{P}(X_1 = i, X_2 = j)}{\mathbb{P}(X_1 = i)} = \mathbb{P}(X_2 = j) = \frac{1}{6},$$

und für $i > j$ ist offensichtlich $\mathbb{P}(M = j \mid X_1 = i) = 0$. Insgesamt erhält man also die bedingte Verteilung von M gegeben X_1 durch

$$\mathbb{P}(M = j \mid X_1 = i) = \begin{cases} \frac{1}{6}, & j > i, \\ \frac{i}{6}, & j = i, \\ 0, & j < i, \end{cases} \quad (i,j) \in \{1, \ldots, 6\}^2.$$

Definition 13.4 (Markov-Kette)

a) *Sei* $(\Omega, \mathcal{A}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum,* E *eine höchstens abzählbare Menge und* $X_i : \Omega \to E$ *diskrete Zufallsvariablen für alle* $i \geq 0$*.*

Dann heißt die Familie $\mathcal{X} = \{X_0, X_1, X_2, \ldots\}$ *von Zufallsvariablen* **Markov-Kette***, falls für alle* $i \geq 1$ *gilt*

$$\mathbb{P}(X_i = e_i \mid X_0 = e_0, \ldots, X_{i-1} = e_{i-1}) = \mathbb{P}(X_i = e_i \mid X_{i-1} = e_{i-1}),$$

d.h. die Eintrittswahrscheinlichkeit eines Zustands e_i *der Markov-Kette zur Zeit* i *hängt nur vom unmittelbar vorhergehenden Zustand ab.*

Die Markov-Kette heißt endlich, falls $|E| < \infty$*.*

E *heißt* **Zustandsraum***.*

b) *Eine* E*-wertige Markov-Kette* $\mathcal{X} = (X_i)_{i \geq 0}$ *heißt* **(zeitlich) homogen***, falls für alle* $i \geq 1$ *gilt*

$$\mathbb{P}(X_i = e_j \mid X_{i-1} = e_i) = \mathbb{P}(X_1 = e_j \mid X_0 = e_i) =: P_{ij},$$

d.h. die Übergangswahrscheinlichkeiten von einem Zustand e_i *zur Zeit* $i-1$ *zu einem anderen Zustand* e_j *zur Zeit* i *sind für alle Zeitpunkte* $i \geq 1$ *identisch.*

Bemerkung 13.5

a) Die durch Definition 13.4 b) definierte Matrix $P = (P_{ij})_{i\geq 1, j\geq 1}$ heißt *Übergangsmatrix* der homogenen Markov-Kette $\mathcal{X}$. Für diese gilt

$$0 \leq P_{ij} \leq 1 \quad \forall\, i, j \geq 1 \qquad \text{und} \quad \sum_{j=1}^{|E|} P_{ij} = 1. \qquad (*)$$

Die zweite Eigenschaft (Zeilensummen sind stets 1) folgt aus

$$1 = \mathbb{P}(X_i \in E \mid X_{i-1} = e_i) = \sum_{j=1}^{|E|} \mathbb{P}(X_i = e_j \mid X_{i-1} = e_i) = \sum_{j=1}^{|E|} P_{ij}.$$

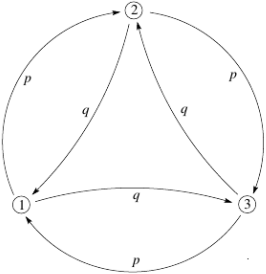
Matrizen mit den Eigenschaften $(*)$ heißen *stochastische Matrizen*.

b) Homogene Markov-Ketten $\mathcal{X}$ mit Übergangsmatrix P kann man durch ihren *Übergangsgraphen* (E, K, W) veranschaulichen: Dabei sind die Ecken des Graphen durch die verschiedenen Elemente des Zustandsraums E gegeben, die (gerichteten) Kanten durch $K = \{(e_i, e_j) \mid P_{ij} > 0\}$ und das Gewicht einer Kante (e_i, e_j) durch $w_{ij} = P_{ij}$, $W = \{w_{ij}, 1 \leq i, j \leq |E|\}$.

Beispiel 13.6 (Irrfahrt im Dreieck)

Betrachte eine homogene Markov-Kette $\mathcal{X}$ mit Zustandsraum $E = \{1, 2, 3\}$ und Übergangsmatrix

$$P = \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix}, \qquad p \in (0,1),\ q = 1 - p.$$

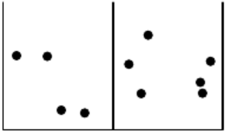


In jedem Zustand wandert man mit Wahrscheinlichkeit p im Uhrzeigersinn weiter und mit Wahrscheinlichkeit $1 - p$ im Gegenuhrzeigersinn.

Beispiel 13.7 (Ehrenfestsches Urnenmodell)

Gegeben ist eine Urne mit zwei Kammern, in denen insgesamt n Kugeln liegen. Es wird in jedem Zeitschritt zufällig eine Kugel ausgewählt, die dann aus ihrer bisherigen in die andere Kammer wechselt.

Sei X_i die Anzahl der Kugeln in der linken Kammer vor der $i+1$-ten Ziehung, dann ist die Familie $\mathcal{X} = (X_i)_{i \geq 0}$ eine endliche Markov-Kette mit Zustandsraum $E = \{0, 1, \ldots, n\}$, denn die Anzahl der Kugeln in einer Kammer kann sich in jedem Schritt nur entweder um 1 erhöhen oder um 1 erniedrigen, daher hängt X_{i+1} nur von X_i ab für alle $i \geq 0$. Es gilt



$$P_{ij} = \begin{cases} \frac{i}{n}, & j = i - 1, \\ \frac{n-i}{n}, & j = i + 1, \\ 0, & \text{sonst.} \end{cases}$$

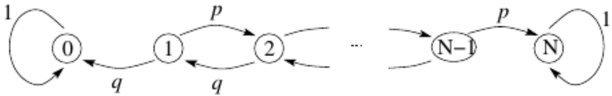
Im obigen Beispiel ist $n = 10$ und $X_i = 4$. Im nächsten Schritt wechselt mit Wahrscheinlichkei $\frac{4}{10}$ eine Kugel von der linken in die rechte Kammer ($X_{i+1} = 3$) und mit Wahrscheinlichkeit $\frac{6}{10}$ eine Kugel von der rechten in die linke Kammer ($X_{i+1} = 5$).

Beispiel 13.8 (Ruinproblem)

Zwei Spieler spielen gegeneinander. Das Startkapital von Spieler 1 sind n Euro, das von Spieler 2 beträgt $N - n$ Euro. In jeder Spielrunde gewinnt Spieler 1 mit Wahrscheinlichkeit $0 < p < 1$ einen Euro von Spieler 2 und verliert mit Wahrscheinlichkeit $q = 1 - p$ einen Euro an Spieler 2. Das Spiel endet, wenn einer der beiden Spieler kein Geld mehr hat.
Sei X_i das Vermögen von Spieler 1 nach dem i-ten Spiel, dann ist $\mathcal{X} = (X_i)_{i \geq 0}$ eine endliche Markov-Kette mit Zustandsraum $E = \{0, 1, 2, \ldots, N\}$ und Übergangsmatrix

$$P = \begin{pmatrix} 1 & 0 & & & & \\ q & 0 & p & & & \\ & q & 0 & p & & \\ & & \ddots & \ddots & \ddots & \\ & & & q & 0 & p \\ & & & & 0 & 1 \end{pmatrix}.$$

Der zugehörige Übergangsgraph sieht wie folgt aus:



Das Spiel wird nach genügend langer Zeit mit dem Sieg von Spieler 1 oder 2 enden, d.h. für ein genügend großes k wird entweder $X_k = 0$ (Spieler 1 verliert) oder $X_k = N$ (Spieler 1 gewinnt) sein.

Frage: Mit welcher Wahrscheinlichkeit gewinnt Spieler 1, d.h. was ist

$$p_n := \mathbb{P}(X_k = N \text{ für ein } k \geq N - n \mid X_0 = n)?$$

Es ist $p_N = 1$ und $p_0 = 0$. Nach Satz 6.4 gilt für alle $1 \leq n \leq N - 1$

$$\begin{aligned}
p_n &= \mathbb{P}(X_k = N \text{ für ein } k \geq N - n \mid X_0 = n) \\
&\underset{6.4}{=} \mathbb{P}(X_k = N \text{ für ein } k \geq N - n \mid X_0 = n, X_1 = n-1) \cdot \mathbb{P}(X_1 = n-1) \\
&\quad + \mathbb{P}(X_k = N \text{ für ein } k \geq N - n \mid X_0 = n, X_1 = n+1) \cdot \mathbb{P}(X_1 = n+1) \\
&= \mathbb{P}(X_k = N \text{ für ein } k \geq N - n \mid X_0 = n-1) \cdot q \\
&\quad + \mathbb{P}(X_k = N \text{ für ein } k \geq N - n \mid X_0 = n+1) \cdot p \\
&= q \cdot p_{n-1} + p \cdot p_{n+1}
\end{aligned}$$

Auflösen von $p_n = (p+q) \cdot p_n = q \cdot p_{n-1} + p \cdot p_{n+1}$ ergibt

$$q \cdot (p_n - p_{n-1}) = p \cdot (p_{n+1} - p_n) \qquad (+)$$

Es gilt stets (Teleskopsumme!) $\sum_{i=1}^{N}(p_i - p_{i-1}) = p_N - p_0 = 1$ und analog $\sum_{i=1}^{n}(p_i - p_{i-1}) = p_n$. Für $p = q = \frac{1}{2}$ folgt induktiv aus $(+)$, dass

$$p_N - p_{N-1} = p_{N-1} - p_{N-2} = \ldots = p_1 - p_0 = p_1.$$

Damit gilt für $p = q = \frac{1}{2}$

$$p_n = \frac{\sum_{i=1}^{n}(p_i - p_{i-1})}{\sum_{i=1}^{N}(p_i - p_{i-1})} = \frac{\sum_{i=1}^{n} p_1}{\sum_{i=1}^{N} p_1} = \frac{np_1}{Np_1} = \frac{n}{N}.$$

Falls $p \neq q$, setze $u := \frac{q}{p}$, dann folgt induktiv aus $(+)$

$$p_n - p_{n-1} = u \cdot (p_{n-1} - p_{n-2}) = u^2 \cdot (p_{n-2} - p_{n-3}) = \ldots = u^{n-1} \cdot (p_1 - p_0) = u^{n-1} \cdot p_1$$

für alle $1 \leq n \leq N$. Damit ist

$$1 = \sum_{i=1}^{N}(p_i - p_{i-1}) = p_1 \sum_{i=0}^{N-1} u^i = p_1 \cdot \frac{1-u^N}{1-u}, \quad \text{d.h. } p_1 = \frac{1-u}{1-u^N}.$$

Damit erhält man

$$p_n = \sum_{i=1}^{n} (p_i - p_{i-1}) = p_1 \sum_{i=0}^{n-1} u^i = \frac{1-u}{1-u^N} \cdot \frac{1-u^n}{1-u} = \frac{1-u^n}{1-u^N}.$$

Insgesamt ergibt sich für die Gewinnwahrscheinlichkeit von Spieler 1 bei einem Startkapital von n Euro und einem Gesamtkapital von N Euro

$$p_n = \begin{cases} \frac{n}{N}. & \text{falls } p = q = \frac{1}{2}, \\ \frac{1-\left(\frac{q}{p}\right)^n}{1-\left(\frac{q}{p}\right)^N}, & \text{sonst.} \end{cases}$$

Die zugehörigen Ruinwahrscheinlichkeiten

$$q_n = 1 - p_n = \mathbb{P}(X_k = 0 \text{ für ein } k \geq n \mid X_0 = n)$$

für Spieler 1 sind damit

$$q_n = \begin{cases} 1 - \frac{n}{N}. & \text{falls } p = q = \frac{1}{2}, \\ \frac{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)^N}, & \text{sonst.} \end{cases}$$

Satz 13.9

Seien $\mathcal{X} = (X_i)_{i\geq 0}$ eine homogene Markov-Kette mit Übergangsmatrix P und Zustandsraum $E = \{e_1, e_2, \ldots\}$, $p_n^i := \mathbb{P}(X_i = e_n)$ und $p^i := (p_1^i, p_2^i, \ldots)^\top$ die Wahrscheinlichkeitsverteilung von X_i für alle $i \geq 0$ (vgl. Satz 3.2). Dann gilt

$$p^i = p^0 \cdot P^i,$$

wobei die rechte Gleichungsseite als Matrixprodukt des Zeilenvektors p^0 mit der i-ten Potenz $P^i = \underbrace{P \cdot P \cdot \ldots \cdot P}_{i \text{ mal}}$ der Matrix P zu verstehen ist.

Dabei ist $p^0 = \big(\mathbb{P}(X_0 = e_n)\big)_{n\geq 1}$ die **Startverteilung** *der Markov-Kette.*

Beispiel 13.10

a) Für die Dreiecks-Irrfahrt aus Beispiel 13.6 ist

$$P = \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix}, \qquad P^2 = \begin{pmatrix} 2pq & q^2 & p^2 \\ p^2 & 2pq & q^2 \\ q^2 & p^2 & 2pq \end{pmatrix}$$

Angenommen, die Irrfahrt startet in 1, so dass

$$p^0 = \big(\mathbb{P}(X_0 = 1), \mathbb{P}(X_0 = 2), \mathbb{P}(X_0 = 3)\big) = (1, 0, 0),$$

dann ist

$$p^2 = \big(\mathbb{P}(X_2 = 1), \mathbb{P}(X_2 = 2), \mathbb{P}(X_2 = 3)\big) = p^0 \cdot P^2 = (2pq, q^2, p^2).$$

Das kann man sich alternativ auch am Übergangsgraphen klarmachen: Z.B. ist $X_2 = 1$, wenn man im ersten Schritt von 1 nach 2 geht und im zweiten wieder zurück, oder im ersten Schritt von 1 nach 3 geht und im zweiten wieder zurück. Beide Ereignisse haben Wahrscheinlichkeit pq, so dass $\mathbb{P}(X_2 = 1) = pq + pq = 2pq$ gelten muss.

b) Falls Spieler 1 im Ruinproblem 13.8 mit n Euro Startkapital beginnt, ist

$$p^0 = \big(\mathbb{P}(X_0 = 0), \ldots, \mathbb{P}(X_0 = N)\big) = (\underbrace{0, \ldots, 0}_{n-1 \text{ mal}}, 1, \underbrace{0, \ldots, 0}_{N-n \text{ mal}}).$$

Ferner ist

$$P = \begin{pmatrix} 1 & 0 & & & & & \\ q & 0 & p & & & & \\ & q & 0 & p & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & q & 0 & p \\ & & & & 0 & 1 \end{pmatrix} \quad \text{und } P^2 = \begin{pmatrix} 1 & 0 & & & & & & \\ q & pq & 0 & p^2 & & & & \\ q^2 & 0 & 2pq & 0 & p^2 & & & \\ 0 & q^2 & 0 & 2pq & 0 & p^2 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & q^2 & 0 & 2pq & 0 & p^2 \\ & & & & q^2 & 0 & pq & p \\ & & & & & & 0 & 1 \end{pmatrix}$$

Damit erhält man (für $n \geq 3$)

$$\begin{aligned} p^2 &= \big(\mathbb{P}(X_2 = 0), \ldots, \mathbb{P}(X_2 = N)\big) = p^0 \cdot P^2 \\ &= (\underbrace{0, \ldots, 0}_{n-3 \text{ mal}}, q^2, 0, 2pq, 0, p^2, \underbrace{0, \ldots, 0}_{N-n-2 \text{ mal}}) \end{aligned}$$

was ebenfalls intuitiv klar ist: Z.B. ist $\mathbb{P}(X_2 = n - 2) = q^2$, da man in zwei Schritten zweimal (jeweils mit Wahrscheinlichkeit q) verlieren muss, um vom Startkapital n auf $n - 2$ zu kommen.

Bemerkung 13.11 (Gem. Verteilung homog. Markov-Ketten)

Ist $\mathcal{X} = (X_i)_{i \geq 0}$ eine homogene Markov-Kette mit Zustandsraum E und Übergangsmatrix P, so folgt aus der Multiplikationsformel (Korollar 6.3) und der Markov-Eigenschaft 13.4 a), dass

$$
\begin{aligned}
&\mathbb{P}(X_0 = e_{i_0}, X_1 = e_{i_1}, \ldots, X_n = e_{i_n}) \\
&\underset{6.3}{=} \mathbb{P}(X_0 = e_{i_0}) \cdot \mathbb{P}(X_1 = e_{i_1} \mid X_0 = e_{i_0}) \cdot \mathbb{P}(X_2 = e_{i_2} \mid X_0 = e_{i_0}, X_1 = e_{i_1}) \\
&\quad \cdot \ldots \cdot \mathbb{P}(X_n = e_{i_n} \mid X_0 = e_{i_0}, X_1 = e_{i_1}, \ldots, X_{n-1} = e_{i_{n-1}}) \\
&\underset{13.4\text{ a)}}{=} \mathbb{P}(X_0 = e_{i_0}) \cdot \mathbb{P}(X_1 = e_{i_1} \mid X_0 = e_{i_0}) \cdot \mathbb{P}(X_2 = e_{i_2} \mid X_1 = e_{i_1}) \\
&\quad \cdot \ldots \cdot \mathbb{P}(X_n = e_{i_n} \mid X_{n-1} = e_{i_{n-1}}) = p^0_{i_0} P_{i_0 i_1} P_{i_1 i_2} \ldots P_{i_{n-1} i_n}
\end{aligned}
$$

Definition 13.12 (Stationäre und reversible Verteilung)

Sei $\mathcal{X} = (X_i)_{i \geq 0}$ eine homogene Markov-Kette mit Übergangsmatrix P und $\mathbb{P}^E$ eine Wahrscheinlichkeitsverteilung auf dem Zustandsraum $E = \{e_1, e_2, \ldots\}$, die durch den Zeilenvektor $\pi = (\pi_1, \pi_2, \ldots) = (\mathbb{P}^E(\{e_1\}), \mathbb{P}^E(\{e_2\}), \ldots)$ gegeben sei.

a) *Gilt $\pi \cdot P = \pi$, so heißt π* **stationäre Verteilung von $\mathcal{X}$**.

b) *Gilt $\pi_k P_{kl} = \pi_l P_{lk}$ für alle $k, l \geq 1$, so heißt π* **reversible Verteilung**.

Bemerkung: Ist π reversibel und gelte $p^i = \pi$ für ein $i \geq 0$, d.h. $\mathbb{P}(X_i = e_n) = \pi_n$ für alle $n \geq 1$, so folgt

$$\begin{aligned}
\mathbb{P}(X_i = e_{j_i}, X_{i+1} = e_{j_{i+1}}, \ldots, X_{i+k} = e_{j_{i+k}}) \\
&\underset{13.11}{=} \pi_{j_i} P_{j_i j_{i+1}} \ldots P_{j_{i+k-1} j_{i+k}} \\
&\underset{\pi \text{ rev.}}{=} \pi_{j_{i+1}} P_{j_{i+1} j_{i+2}} P_{j_{i+2} j_{i+3}} \ldots P_{j_{i+k-1} j_{i+k}} P_{j_{i+1} j_i} \\
&= \cdots = \pi_{j_{i+k}} P_{j_{i+k} j_{i+k-1}} \ldots P_{j_{i+1} j_i} \\
&\underset{13.11}{=} \mathbb{P}(X_i = e_{j_{i+k}}, X_{i+1} = e_{j_{i+k-1}}, \ldots, X_{i+k} = e_{j_i}).
\end{aligned}$$

Das heißt, die Wahrscheinlichkeit, dass die Markov-Kette in k aufeinander folgenden Schritten die Zustände $e_{j_i}, \ldots e_{j_{i+k}}$ durchläuft, ist genau gleich der Wahrscheinlichkeit, die Zustände in umgekehrter (reversibler) Reihenfolge zu durchlaufen (das gilt für alle $k \geq 2$).

Das nachfolgende Korollar veranschaulicht die Bedeutung einer stationären Verteilung: Hat die (homogene) Markov-Kette zu einem Zeitpunkt $i \geq 0$ die stationäre Verteilung π, d.h. gilt $\pi = \mathbb{P}_{X_i} = p^i$, dann bleibt diese für alle nachfolgenden Zeitpunkte unveränderlich (stationär) bestehen, d.h. alle $X_i, X_{i+1}, X_{i+2}, \ldots$ sind identisch nach π verteilt.

Korollar 13.13

Sei $\mathcal{X} = (X_i)_{i \geq 0}$ eine homogene Markov-Kette mit Übergangsmatrix P und π eine Wahrscheinlichkeitsverteilung auf dem Zustandsraum E.

a) *Ist π stationär und $\mathbb{P}_{X_i} = p^i = \pi$, dann gilt $p^i = \mathbb{P}_{X_i} = \mathbb{P}_{X_{i+1}} = p^{i+1}$, d.h. X_i und X_{i+1} (und damit auch $X_{i+2}, X_{i+3}, \dots$) sind identisch nach π verteilt.*

b) *Ist π reversibel, dann ist π auch stationär.*

Beispiel 13.14

a) Für die Irrfahrt im Dreieck (Beispiel 13.6) ist $\pi = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ eine stationäre Verteilung, denn für $0 < p < 1$ und $q = 1 - p$ ist

$$\pi \cdot P = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right) \cdot \begin{pmatrix} 0 & p & q \\ q & 0 & p \\ p & q & 0 \end{pmatrix} = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right).$$

π ist aber (außer im Fall $p = \frac{1}{2}$) nicht reversibel, denn z.B. ist $\pi_1 P_{12} = \frac{1}{3} p \neq \frac{1}{3} q = \pi_2 P_{21}$. Das zeigt auch, dass Stationarität und Reversibilität nicht äquivalent sind, i.A. gilt nur die Richtung „π reversibel $\Longrightarrow$ π stationär“.

b) Im Ruinproblem (Beispiel 13.8) ist, wie man leicht nachrechnet, jede Verteilung der Form $\pi = (p', 0, \dots, 0, q')$ mit $0 \leq p' \leq 1$ und $q' = 1 - p'$ stationär. Das verdeutlich nochmals die schon aus dem Übergangsgraph ersichtliche Tatsache, dass die Zustände 0 und N nicht mehr verlassen werden, sofern sie einmal erreicht werden. Solche Zustände nennt man *absorbierende Zustände* oder *Fallen*.

c) Im Ehrenfestschen Urnenmodell (Beispiel 13.7) mit insgesamt n Kugeln ist die Binomialverteilung $B(n, \frac{1}{2})$ eine stationäre Verteilung. Zum Nachweis reicht es nach Korollar 13.13 b) zu zeigen, dass die $B(n, \frac{1}{2})$-Verteilung reversibel ist.

Da für die Übergangsmatrix P im Ehrenfestmodell gilt, dass $P_{ij} \neq 0$ nur dann, wenn $j = i + 1$ oder $j = i - 1$, genügt es zu zeigen, dass $\pi_i P_{i\,i-1} = \pi_{i-1} P_{i-1\,i}$ für $1 \leq i \leq n$. Da $\pi_i = b_{n,\frac{1}{2}}(\{i\}) = \binom{n}{i} \frac{1}{2^n}$, ist

$$
\begin{aligned}
&\pi_i P_{i\,i-1} \\
&= \binom{n}{i} \frac{1}{2^n} \cdot \frac{i}{n} = \frac{1}{2^n} \frac{n!}{i!(n-i)!} \cdot \frac{i}{n} = \frac{1}{2^n} \frac{n!}{(i-1)!(n-(i-1))!} \cdot \frac{n-i+1}{n} \\
&= \binom{n}{i-1} \frac{1}{2^n} \cdot \frac{n-(i-1)}{n} = \pi_{i-1} P_{i-1\,i} \qquad \square
\end{aligned}
$$

Definition 13.15 (Irreduzibilität und Aperiodizität)

Sei $\mathcal{X} = (X_i)_{i\geq 0}$ eine homogene Markov-Kette mit Zustandsraum $E = \{e_1, e_2, \dots\}$.

a) *Der Zustand e_j* **kommuniziert** *mit dem Zustand e_k, falls es ein $i \geq 1$ gibt mit $\mathbb{P}(X_i = e_k \mid X_0 = e_j) > 0$, d.h. wenn es möglich ist, ausgehend von e_j (irgendwann) den Zustand e_k zu erreichen.*
Notation: $e_j \to e_k$. *Falls $e_j \to e_k$ und $e_k \to e_j$, schreiben wir $e_j \leftrightarrow e_k$.*

b) *Falls $e_j \leftrightarrow e_k$ für alle $e_j, e_k \in E$, heißt die Markov-Kette $\mathcal{X}$* **irreduzibel**, *andernfalls heißt $\mathcal{X}$* **reduzibel**.

c) *Ein Zustand e_k heißt* **aperiodisch**, *falls gilt*

$$d(k) := \mathrm{ggT}\{i \geq 1 : \mathbb{P}(X_i = e_k \mid X_0 = e_k) > 0\} = 1,$$

d.h. wenn die Längen aller möglichen Folgen oder Pfade, entlang derer man wieder zum Ausgangszusand e_k zurückkommen kann, teilerfremd sind.

Sind alle Zustände $e_k \in E$ aperiodisch, heißt $\mathcal{X}$ **aperiodisch**.

Bemerkung 13.16

a) Für eine homogene Markov-Kette mit Übergangsmatrix P ist die Wahrscheinlichkeit, in i Schritten aus dem Zustand e_j in den Zustand e_k zu gelangen, gegeben durch $(P^i)_{jk}$. Das kann man aus Bemerkung 13.11 ableiten, indem man $p_j^0 =: 1$ setzt (Start in e_j mit Wahrscheinlichkeit 1) und dann über alle möglichen Zwischenstationen $e_{j_1}, \ldots, e_{j_{i-1}}$ aufsummiert.

   Somit gilt $e_j \to e_k$ genau dann, wenn es ein $i \geq 1$ gibt, so dass $(P^i)_{jk} > 0$.

   Ebenso ist die Wahrscheinlichkeit, in i Schritten vom Zustand e_k ausgehend wieder dorthin zurückzukommen, gegeben durch $(P^i)_{kk}$. Daher ist $d(k) = \text{ggT}\{i \geq 1 \mid (P^i)_{kk} > 0\}$.

b) Ob zwei Zustände e_j und e_k kommunizieren, kann man auch leicht dem Übergangsgraphen ansehen: Es gilt $e_j \to e_k$ genau dann, wenn es einen (endlich langen) Weg entlang der gerichteten Kanten vom Knoten e_j zum Knoten e_k gibt. Die Markov-Kette ist irreduzibel, wenn es in ihrem Übergangsgraphen von jedem Knoten einen endlich langen Weg enlang der gerichteten Kanten zu jedem anderen Knoten gibt.

Beispiel 13.17

a) Für die Dreiecks-Irrfahrt aus Beispiel 13.6 ist

$$P^2 = \begin{pmatrix} 2pq & q^2 & p^2 \\ p^2 & 2pq & q^2 \\ q^2 & p^2 & 2pq \end{pmatrix}, \qquad P^3 = \begin{pmatrix} p^3+q^3 & 3p^2q & 3pq^2 \\ 3pq^2 & p^3+q^3 & 3p^2q \\ 3p^2q & 3pq^2 & p^3+q^3 \end{pmatrix}$$

Wegen $0 < p < 1$ gilt $P^2 > 0$ (d.h. jede Komponente der Matrix P^2 ist > 0). Das bedeutet nach Bemerkung 13.16 a), dass man von jedem Zustand ausgehend in zwei Schritten jeden anderen Zustand erreichen kann. Somit kommunizieren alle Zustände, und die Markov-Kette ist irreduzibel.

Wegen $P^2 > 0$ sind insbesondere auch die Diagonalenelemente $(P^2)_{kk} > 0$, d.h. man kann von jedem Zustand aus in zwei Schritten wieder dorthin zurückkommen. Da offensichtlich auch $P^3 > 0$ und somit speziell $(P^3)_{kk} > 0$, kann man alternativ auch in drei Schritten von jedem Zustand ausgehend wieder zu diesem zurückkommen. Da 2 und 3 teilerfremd sind, folgt $d(k) = 1$ für alle $k = 1, 2, 3$, d.h. die Dreiecks-Irrfahrt ist auch aperiodisch.

b) Die zum Ruinproblem (Beispiel 13.8) gehörige Markov-Kette ist reduzibel, denn die beiden absorbierenden Zustände 0 und N kommunizieren mit keinem der anderen Zustände (ist man dort angelangt, führt kein Weg mehr zurück).

c) Die Markov-Kette $\mathcal{X}$, die jeweils die Anzahl der Kugeln in der linken Kammer der Ehrenfestschen Urne angibt, ist irreduzibel, denn man kann für alle $0 \leq i, j \leq n$ von einem Zustand mit i Kugeln in der linken Kammer zu einem Zustand mit j Kugeln gelangen: Für $i > j$ muss man nur nacheinander $i - j$ Kugeln von der linken in die rechte Kammer legen, für $i < j$ umgekehrt $j - i$ Kugeln von der rechten in die linke Kammer. Für $i = j$ muss man nacheinander eine Kugel nach rechts und wieder zurück (oder umgekehrt) legen.

Da in jedem Schritt eine Kugel von links nach rechs (oder umgekehrt) wandert, kann man einen bestimmten Ausgangszustand offensichtlich nur in einer geraden Anzahl von Schritten erreichen. Daher ist diese Markov-Kette periodisch, alle Zustände haben die Periode $d(k) = 2$, $0 \leq k \leq n$.

Wir geben zum Abschluss noch (ohne Beweis) zwei wichtige Konvergenzsätze für Markov-Ketten an:

Satz 13.18 (Markov-Ketten-Konvergenzsatz)

a) *Sei $\mathcal{X} = (X_i)_{i\geq 0}$ eine endliche, homogene, irreduzible und aperiodische Markov-Kette. Dann hat $\mathcal{X}$ genau eine stationäre Verteilung π, und für alle $e_k \in E$ gilt $\lim_{i\to\infty} \mathbb{P}(X_i = e_k) = \pi_k$.*

b) *Ist $\mathcal{X} = (X_i)_{i\geq 0}$ eine endliche, homogene, irreduzible und periodische Markov-Kette mit Periode ℓ, d.h. $d(k) \equiv \ell$ für alle $1 \leq k \leq |E|$. Dann hat $\mathcal{X}$ genau eine stationäre Verteilung π, und es gilt für alle $e_k \in E$, dass $\lim_{i\to\infty} \mathbb{P}(X_{\ell i} = e_k) = \pi_k$.*

Satz 13.19

Sei $\mathcal{X}$ eine irreduzible, aperiodische Markov-Kette mit stationärer Verteilung π und $f : E \to \mathbb{R}$ eine Funktion mit $\mathbb{E}_\pi[|f|] = \sum_{k\geq 1} |f(e_k)|\pi_k < \infty$. Dann gilt mit Wahrscheinlichkeit 1, dass

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = \mathbb{E}_\pi[f] = \sum_{k\geq 1} f(e_k)\pi_k.$$

Bemerkung 13.20

a) Man beachte, dass die Konvergenz gegen die stationäre Verteilung in Satz 13.18 nicht vom Startwert bzw. der Verteilung p^0 von X_0 abhängt! Anschaulich gesprochen vergisst die Markov-Kette nach langer Zeit, von wo aus genau sie gestartet ist.

b) p^i und π geben die Wahrscheinlichkeiten an, dass sich die Markov-Kette zu einem festen Zeitpunkt in einem bestimmten Zustand befindet (z.B. $p^i_k = \mathbb{P}(X_i = e_k)$ oder $\pi_k = \mathbb{P}(X_i = e_k)$, falls $\mathbb{P}_{X_i} = \pi$). Klar davon zu unterscheiden sind die Übergangswahrscheinlichkeiten $\mathbb{P}(X_i = e_j \mid X_{i-1} = e_i) = P_{ij}$ aus der Übergangsmatrix P.

c) Für die zum Ruin-Problem gehörige Markov-Kette aus Beispiel 13.8 ist die stationäre Verteilung nicht eindeutig, jede Verteilung π mit $\pi_0 + \pi_N = 1$ stationär. Ferner haben wir gesehen, dass auch die Gewinnwahrscheinlichkeiten $\lim_{i\to\infty} \mathbb{P}(X_i = N \mid X_0 = n) = p_n$ vom Startwert abhängen. Beides steht nicht im Widerspruch zu Satz 13.18, da in diesem Fall die Markov-Kette reduzibel ist und damit eine der zentralen Voraussetzungen des Satzes verletzt ist.