2 Wahrscheinlichkeitstheorie

2.1 Grundlagen der Wahrscheinlichkeitstheorie

2.1.1 Ergebnisse und Ereignisse

- Ausgangspunkt der Wahrscheinlichkeitstheorie sind **Zufallsexperimente**.

  Definition 2.1. *Unter einem* Zufallsexperiment *versteht man einen Versuch (im weitesten Sinne des Wortes), dessen Ausgang unter bestimmten wesentlichen und fixierten Bedingungen im Rahmen bestimmter Möglichkeiten ungewiss ist.*

  – **Notation**: Die einzelnen Versuchsausgänge werden häufig mit ω und als *Ergebnisse* oder *Elementarereignisse* bezeichnet. Die Menge aller möglichen Versuchsausgänge, m.a.W. aller Ergebnisse, wird mit Ω ($\neq \emptyset$) und *Ergebnisraum* oder *Basismenge* bezeichnet.

    Beispiel 2.1. *(i) Werfen einer Münze:* $\omega \in \{K, Z\}, \Omega = \{K, Z\}$.

    (ii) Werfen zweier (unterschiedlicher) Münzen: $\omega = (\omega_1, \omega_2), \omega_i \in \{K, Z\}$, $i = 1, 2$, $\Omega = \{(K, K), (K, Z), (Z, K), (Z, Z)\}$.

    (iii) n-maliges Werfen einer Münze: $\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{K, Z\}$, $i = 1, \dots, n$, $\Omega = \{(\omega_1, \dots, \omega_n) : \omega_i \in \{K, Z\}, i = 1, \dots, n\}$.

    (iv) Werfen eines Würfels: $\omega \in \{1, 2, \dots, 6\}, \Omega = \{1, 2, \dots, 6\}$.

    (v) Werfen zweier (unterscheidbarer) Würfel: $\omega = (i, j)$. *Sei i die Augenzahl des ersten Würfels und j die Augenzahl des zweiten Würfels. Dann erhalten wir* $\Omega = \{(i, j) : i, j \in \{1, 2, \dots, 6\}\}$.

    (vi) Sonntagsziehung im Lotto „6 aus 49“ (ohne Zusatzzahl): $\Omega = \{\omega = \{i_1, \dots, i_6\} : i_1, \dots, i_6 \in \{1, \dots, 49\}, i_j < i_k \,\forall\, j < k\}$.

    (vii) Bei der Geburt eines Kindes werden das Gewicht ω_1 in Gramm, die Größe ω_2 in Zentimetern und das Geschlecht $\omega_3 \in \{M, W\}$ erhoben: $\omega = (\omega_1, \omega_2, \omega_3)$, $\Omega = (0, \infty) \times (0, \infty) \times \{M, W\}$.

    (viii) Niederschlagsmenge ω am Marienplatz am 19.08.2008, gemessen in Millimeter pro Quadratmeter: $\Omega = [0, \infty)$.

    (ix) Schadenshöhe ω bei einem PKW-Unfall, die der Versicherer in Euro zu zahlen hat: $\Omega = [0, \infty)$.

    (x) Anzahl ω aller polizeilich gemeldeter Kfz-Unfälle an einem bestimmten Tag auf der Ludwigstraße: $\Omega = \mathbb{N}_0$.

  – **Bemerkungen**:

    (i) Der Ergebnisraum Ω ist nicht eindeutig festgelegt. Einzige Bedingung: Nach Ausführung des Zufallsexperimentes muss genau ein Elementarereignis ω aus Ω als Versuchsergebnis feststehen. Insbesondere ist es nicht notwendig, dass alle $\omega \in \Omega$ auch tatsächlich auftreten können, d.h. der Ergebnisraum Ω kann größer gewählt werden als unbedingt notwendig.

    (ii) Vor der Durchführung des Zufallsexperimentes ist der tatsächliche Ausgang ungewiss, d.h. es ist unbekannt, welches Ergebnis $\omega \in \Omega$ auftreten wird. Nach der Durchführung des Zufallsexperimentes steht der aufgetretene Ausgang ω fest, damit ist die Ungewissheit verschwunden. Man sagt, **das Zufallsexperiment wurde *realisiert***. Daher wird das nach der Durchführung erschienene Elementarereignis ω auch als *Realisation des Zufallsexperimentes* bezeichnet. Wird das Zufallsexperiment n-mal durchgeführt, so ergibt sich eine *Folge von Realisierungen* $\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(n)}$, die wir als *Stichprobe* bezeichnen.

- Häufig sind wir nicht daran interessiert, welchen konkreten Ausgang ω ein Zufallsexperiment hat, sondern ob das Versuchsergebnis in der einen oder anderen Teilmenge aller Versuchsausgänge liegt.

Definition 2.2. *Alle die Teilmengen $A \subseteq \Omega$, für die nach den Versuchsbedingungen eine der Aussagen „der Versuchsausgang ω gehört zu A ($\omega \in A$)“ oder „der Versuchsausgang gehört nicht zu A ($\omega \notin A$)“ möglich ist, heißen* Ereignisse.

Beispiel 2.2. *(Fortsetzung von Beispiel 2.1)*

(i) $A :=$ *„Es erscheint Wappen“.*

(v) $A :=$ *„Die Summe der Augenzahlen ist gerade“.*

(vi) $A :=$ *„Bei der Ziehung erscheint mindestens ein Zahlenzwilling“,* $B :=$ *„Der abgegebene Tipschein enthält drei Richtige“.*

(viii) $A :=$ *„Es regnet mehr als 10[mm]“.*

(ix) $A :=$ *„Der Schaden ist größer als 100 000 Euro“.*

– **Fazit**: Ereignisse lassen sich durch mengentheoretische Verknüpfungen aus Ergebnissen bilden. Ebenso lassen sich aus Ereignissen neue Ereignisse bilden.
– Zum Ergebnis $\omega_i \in \Omega$ ist $\{\omega_i\} \subset \Omega$ ein Ereignis (Elementarereignis).
– Die Menge aller Ereignisse eines Zufallsexperimentes mit Ergebnisraum Ω entspricht der Menge aller Teilmengen von Ω und damit der *Potenzmenge* $\mathfrak{P}(\Omega)$, also $\mathfrak{P}(\Omega) := \{A : A \subseteq \Omega\}$.
– Ausgehend von einem gewissen Mengensystem $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$, dessen Elemente Ereignisse sind, kann man neue Ereignisse durch Konstruktionen von Aussagen mit den logischen Funktoren „oder ($\vee$)“, „und ($\wedge$)“ und „nicht ($\neg$)“ bilden, womit sich in der Sprache der Mengentheorie die im folgenden behandelten Operationen darstellen lassen.
– Dazu betrachten wir die Ereignisse $A, B \subset \Omega$ und Familien von Teilmengen von Ω. Eine Familie $(A_i)_{i \in I}$ von Teilmengen von Ω ist eine Abbildung der Indexmenge I in $\mathfrak{P}(\Omega)$, die jedem $i \in I$ eine Menge $A_i \in \mathfrak{P}(\Omega)$ als Bild zuordnet. Im Fall $I = \mathbb{N}$ ist $(A_i)_{i \in \mathbb{N}}$ gleich der (abzählbar unendlichen) Folge der Mengen $A_1, A_2, \ldots$, und für $I = \{1, 2, \ldots, n\}$ $(n \in \mathbb{N})$ ist $(A_i)_{i \in I}$ gleich dem geordneten n-Tupel $(A_1, A_2, \ldots, A_n)$.

Gleichheit:	$A = B \quad :\Leftrightarrow \quad \forall \omega \in \Omega : \quad \omega \in A \Leftrightarrow \omega \in B$
Teilmenge:	$A \subseteq B \quad :\Leftrightarrow \quad \forall \omega \in \Omega : \quad \omega \in A \Rightarrow \omega \in B$
Schnitt:	$A \cap B := \{\omega \in \Omega \mid (\omega \in A) \wedge (\omega \in B)\}$
	$\bigcap_{i \in I} A_i := \{\omega \in \Omega \mid \forall i \in I : \omega \in A_i\}$
Vereinigung:	$A \cup B := \{\omega \in \Omega \mid (\omega \in A) \vee (\omega \in B)\}$
	$\bigcup_{i \in I} A_i := \{\omega \in \Omega \mid \exists i \in I : \omega \in A_i\}$
	$\dot{\bigcup_{i \in I}} A_i := \{\omega \in \Omega \mid \exists i \in I : \omega \in A_i, A_i \cap A_j = \emptyset \text{ für } i \neq j\}$
Differenz:	$A \setminus B := \{\omega \in \Omega \mid (\omega \in A) \wedge (\omega \notin B)\} = A \cap B^c$
Komplement:	$\bar{A} := \{\omega \in \Omega \mid \omega \notin A\}$ (alternative Notation: A^c) $A = (A^c)^c$
Symmetrische Differenz:	$A \triangle B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$
Mächtigkeit:	$\lvert A \rvert :=$ Anzahl Elemente von A
Kardinalität:	$\lvert \mathbb{N} \rvert = \aleph_0$
	$\lvert \mathfrak{P}(\Omega) \rvert = 2^{\lvert \Omega \rvert}$
Kartesisches Produkt:	$A \times B := \{(a, b) \mid a \in A \wedge b \in B\}$
	$\prod_{i \in I} A_i := \{(a_1, a_2, \ldots, a_{\lvert I \rvert}) \mid a_i \in A_i \forall i \in I\}$
	$A^k = \prod_{i=1,\ldots,k} A = \{(a_1, \ldots, a_k) \mid a_i \in A, i = 1, \ldots, k\}$

– **Folgen von Mengen**: Wir nennen eine Folge $(A_n)_{n \geq 1}$ von Teilmengen von Ω monoton wachsend oder kurz *wachsend*, falls $A_n \subset A_{n+1} \; \forall \, n \in \mathbb{N}$ und monoton fallend oder kurz *fallend*, falls $A_n \supset A_{n+1} \; \forall \, n \in \mathbb{N}$. Eine Folge heißt *monoton*, wenn sie wachsend oder fallend ist.

  - $*$ Konvergiert $(A_n)_{n\geq 1}$ wachsend gegen A, so schreiben wir kurz $A_n \uparrow A$, also
$$A_n \uparrow A :\Leftrightarrow A_1 \subset A_2 \subset A_3 \subset \dots \quad \text{und} \quad \bigcup_{i=1}^{\infty} A_i = A.$$
  - $*$ Konvergiert $(A_n)_{n\geq 1}$ fallend gegen A, so schreiben wir kurz $A_n \downarrow A$, also
$$A_n \downarrow A :\Leftrightarrow A_1 \supset A_2 \supset A_3 \supset \dots \quad \text{und} \quad \bigcap_{i=1}^{\infty} A_i = A.$$
  - $*$ $\liminf A_n := \bigcup_{n=1}^{\infty} \bigcap_{m\geq n} A_m$, Menge aller Elemente aus Ω, die in fast allen A_n liegen.
    $\limsup A_n := \bigcap_{n=1}^{\infty} \bigcup_{m\geq n} A_m$, Menge aller Elemente aus Ω, die in unendlich vielen A_n liegen.
    Bemerkung: „fast alle“ … bis auf endlich viele
  - $*$ Die Folge A_n heißt konvergent, falls $\liminf A_n = \limsup A_n$.
- Interpretationen:

$A \subset \Omega$	…	„A tritt ein“, „erscheint“, „wird realisiert“
$\bar{A} \subset \Omega$	…	”A tritt nicht ein”
$A_1 \cup A_2$	…	”A_1 oder A_2 treten ein”
$\bigcup_{i\in I} A_i$	…	”mindestens eines der A_i tritt ein”
$A_1 \cap A_2$	…	”A_1 und A_2 treten ein”
$\bigcap_{i\in I} A_i$	…	”alle A_i treten ein”
$A_1 \cap A_2 = \emptyset$	…	”A_1 und A_2 treten nicht gleichzeitig ein”, ”sind unvereinbar”
$A \triangle B$	…	”Entweder A_1 oder A_2 tritt ein (aber nicht beide zusammen!)”
$A_1 = A_2$	…	”A_1 und A_2 beschreiben das gleiche Ereignis”
Ω	…	”Das sichere Ereignis”
$\bar{\Omega} = \emptyset, \{\,\}$	…	”Das unmögliche Ereignis”

Gesetzmäßigkeiten

$A, B, C \subset \Omega$.

Reflexivität:	$A \subseteq A$
Asymmetrie:	$A \subseteq B$ und $B \subseteq A \Rightarrow A = B$
Transitivität:	$A \subseteq B$ und $B \subseteq C \Rightarrow A \subseteq C$
	$\Rightarrow \mathfrak{P}(\Omega)$ ist bezüglich $\subseteq$ partiell geordnet.
Kommutativgesetz:	$A \cup B = B \cup A$
	$A \cap B = B \cap A$
Assoziativgesetz:	$(A \cup B) \cup C = A \cup (B \cup C)$
	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributivgesetz:	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
	$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
De Morgansche Regeln:	$\overline{A \cup B} = \bar{A} \cap \bar{B}$
	$\overline{A \cap B} = \bar{A} \cup \bar{B}$
Mächtigkeiten:	
Gleichmächtigkeit:	$\lvert A\rvert = \lvert B\rvert :\Longleftrightarrow \exists f : A \to B$ bijektiv
Addition von Mächtigkeiten:	$A \cap B = \emptyset \Longleftrightarrow \lvert A\rvert + \lvert B\rvert = \lvert A \cup B\rvert$

- Betrachtung von Maßen: Gegeben ist eine (nichtleere) Grundmenge Ω, und man möchte einer hinreichend großen Klasse von Teilmengen A von Ω eine Maßzahl $\mu(A)$ zuordnen, allgemeiner: es soll auf Ω eine Mengenfunktion $\mu : \mathfrak{E} \to \mathbb{R} \cup \{-\infty, \infty\}$ definiert werden, wobei $\mathfrak{E}$ eine Klasse von Teilmengen von Ω ist.

- Anforderungen an ein Maß (Volumendefinition):

$$\mu : A \subset \mathbb{R}^n \longrightarrow \mu(A) \in [0, \infty)$$

  (i) Ist $A \subset B$, so gilt $\mu(A) \leq \mu(B)$ (Monotonie),
  (ii) μ ist translationsinvariant, d.h. für $x_0 \in \mathbb{R}^n$ gilt $\mu(A + x_0) = \mu(A)$,
  (iii) Sind $A_1, A_2, \ldots$ abzählbar viele disjunkte Teilmengen des $\mathbb{R}^n$, so gilt

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad (\sigma - \text{Additivität}).$$

- Idee: Umfangreichste Klasse von Teilmengen ist die Potenzmenge $\mathfrak{P}(\Omega) := \{A | A \subseteq \Omega\}$, d.h. die Menge aller Teilmengen von Ω.

- Problem: Es existiert keine Funktion $\mu : \mathfrak{P}(\mathbb{R}^n) \longrightarrow [0, \infty)$, mit $\mathfrak{P}$ die Potenzmenge von $\mathbb{R}^n$, mit den o.g. Eigenschaften. (Satz von Vitali: Das Maßproblem ist unlösbar.)

- Lösung: Man beschränkt sich auf eine Menge von Teilmengen von $\mathfrak{P}(\Omega)$, die „hinreichend reichhaltig" (= alle interessierenden Ereignisse sind enthalten) ist, sog. $\sigma-$Algebren.

  Definition 2.3. *Ein nichtleeres Mengensystem $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$ heißt **Algebra** über Ω, falls gilt:*

  (A1) $A \in \mathfrak{A} \Longrightarrow \overline{A} \in \mathfrak{A}$
  (A2) $A, B \in \mathfrak{A} \Longrightarrow A \cup B \in \mathfrak{A}$

  Beispiel 2.3. $\Omega = \{a, b, c, d\}, \mathfrak{A}_1 = \{\emptyset, \{a\}, \{b, c, d\}, \{a, b, c, d\}\}$ *ist eine Algebra,*
  $\mathfrak{A}_2 = \{\emptyset, \{a\}, \{b, c\}, \{a, b, c, d\}\}$ *hingegen ist keine Algebra.*

  Lemma 2.4. *Sei $\mathfrak{A}$ eine Algebra und $A_1, A_2, \ldots, A_n \in \mathfrak{A}$. Dann gilt*

  (i) $\emptyset, \Omega \in \mathfrak{A}$,
  (ii) $A_1 \cap A_2 \in \mathfrak{A}$,
  (iii) $A_1 \backslash A_2 \in \mathfrak{A}$,
  (iv) $\bigcup_{i=1}^{n} A_i \in \mathfrak{A}, \quad \bigcap_{i=1}^{n} A_i \in \mathfrak{A}$.

  Beweis. (i) Weil $\mathfrak{A}$ per Definition nichtleer ist, existiert ein $A \in \mathfrak{A}$ mit $A \subset \Omega$. Daher gilt wegen (A1) $\overline{A} \in \mathfrak{A}$ und wegen (A2), dass $A \cup \overline{A} = \Omega \in \mathfrak{A}$ und damit wegen (A1) auch $\overline{\Omega} = \emptyset \in \mathfrak{A}$.
  (ii) $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathfrak{A}$ wegen (A2).
  (iii) $A_1 \backslash A_2 = A_1 \cap A_2^c = (A_1^c \cup A_2)^c \in \mathfrak{A}$
  (iv) Beweis per Induktion.

  □

- Um Grenzwerte bilden zu können, ist es erforderlich, dass das Mengensystem $\mathfrak{A}$ nicht nur abgeschlossen ist bezüglich Vereinigung bzw. Durchschnitt von *endlich* vielen Mengen, sondern auch bezüglich Vereinigung bzw. Durchschnitt von *abzählbar unendlich* vielen Mengen (notwendig u.a. für überabzählbar unendliche Ergebnisräume, z.B. wenn $\Omega = [0, 1]$). Dies führt zum Begriff der **σ-Algebra**.

  Definition 2.5. *Eine Algebra $\mathfrak{A} \subseteq \mathfrak{P}(\Omega)$ heißt σ-**Algebra** über Ω, falls zusätzlich gilt:*

  (A3) $A_1, A_2, \ldots \in \mathfrak{A} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}$.

- Bemerkungen:

- Das Paar $(\Omega, \mathfrak{A})$ heißt *Meßraum*, falls $\mathfrak{A}$ eine σ-Algebra ist. Vorteil: Auf einem Meßraum kann ein Maß (in unserem Fall ein Wahrscheinlichkeitsmaß) definiert werden.
- Sei $\Omega = \mathbb{N}$ und sei $\mathfrak{A}$ das Mengensystem derjenigen Teilmengen A von $\mathbb{N}$, so dass entweder A oder A^c nur endlich viele Elemente hat. Das Mengensystem $\mathfrak{A}$ ist eine Algebra, jedoch keine σ-Algebra.
- Für jedes Ω ist die Potenzmenge $\mathfrak{P}(\Omega)$ stets eine σ-Algebra.
- Falls Ω endlich oder abzählbar unendlich ist, kann $\mathfrak{A} = \mathfrak{P}(\Omega)$ gewählt werden. Falls Ω überabzählbar unendlich ist (z.B. $\Omega = \mathbb{R}$ oder $\Omega = [0,1]$), dann muss eine *kleinere* σ-Algebra betrachtet werden (nicht $\mathfrak{P}(\Omega)$!)

Definition 2.6. *Ist $A \subset \Omega$, so heißt*

$$\sigma(A) = \{\emptyset, A, A^c, \Omega\}$$

die von A erzeugte σ-Algebra *und ist die kleinste σ-Algebra, die A enthält.*

Satz 2.7. *Sei $I \neq \emptyset$ eine beliebige Indexmenge und $\mathfrak{A}_i$ eine σ-Algebra über Ω für alle $i \in I$. Dann ist auch*

$$\mathfrak{A} := \bigcap_{i \in I} \mathfrak{A}_i$$

eine σ-Algebra über Ω.

Definition 2.8. *Sei $\mathcal{E} \subset \mathfrak{P}(\Omega)$ ein Mengensystem und Σ die Menge aller σ-Algebren über Ω, die $\mathcal{E}$ enthalten. Dann wird die σ-Algebra*

$$\sigma(\mathcal{E}) := \bigcap_{\mathfrak{A} \in \Sigma} \mathfrak{A}$$

als die von $\mathcal{E}$ erzeugte σ-Algebra $\sigma(\mathcal{E})$ *bezeichnet. Gilt umgekehrt für eine σ-Algebra $\mathfrak{A}$*

$$\sigma(\mathcal{E}) = \mathfrak{A},$$

so heißt $\mathcal{E}$ Erzeuger von $\mathfrak{A}$.

Beispiel 2.4. – *Sei $A \subset \Omega$, $\mathcal{E} = \{A\}$ (ein Mengensystem bestehend aus einer Menge). Dann ist*

$$\sigma(\mathcal{E}) = \sigma(A) = \{\emptyset, A, \bar{A}, \Omega\}.$$

- *Sei $\Omega = \{1, 2, \ldots, 7\}$ und $\mathcal{E} = \{\{1,2\}, \{6\}\}$. Dann ist*

$$\sigma(\mathcal{E}) = \{\emptyset, \{1,2\}, \underbrace{\{3,4,5,6,7\}}_{\overline{\{1,2\}}}, \{6\}, \underbrace{\{1,2,3,4,5,7\}}_{\overline{\{6\}}}, \underbrace{\{1,2,6\}}_{\{1,2\}\cup\{6\}}, \underbrace{\{3,4,5,7\}}_{\overline{\{1,2\}\cup\{6\}}}, \Omega\}.$$

- *Sei $\mathcal{A}$ eine σ-Algebra über Ω. Dann ist $\sigma(\mathcal{A}) = \mathcal{A}$.*

Beispiel 2.5. *Ist $\Omega = \mathbb{R}$, so wird oft die sogenannte* Borel-σ-Algebra $\mathfrak{B}$ *oder $\mathfrak{B}(\mathbb{R})$ betrachtet. Sie ist definiert als die kleinste σ-Algebra von Teilmengen von $\mathbb{R}$, die alle offenen Intervalle (a,b) enthält, wobei $-\infty < a < b < \infty$, d.h.*

$$\mathfrak{B}(\mathbb{R}) = \sigma\left(\{(a,b) : -\infty < a < b < \infty\}\right).$$

Insbesondere enthält $\mathfrak{B}(\mathbb{R})$ auch alle halboffenen bzw. abgeschlossenen Intervalle, denn es gilt

$$\begin{aligned}
(a,b] &= \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right) \in \mathfrak{B}(\mathbb{R}), \\
[a,b) &= \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b\right) \in \mathfrak{B}(\mathbb{R}), \\
[a,b] &= \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right) \in \mathfrak{B}(\mathbb{R}).
\end{aligned}$$

Für jede abzählbare Teilmenge $C = \{x_1, x_2, \ldots\}$ von $\mathbb{R}$ gilt $C \in \mathfrak{B}(\mathbb{R})$, denn für jedes $x \in \mathbb{R}$ gilt

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x + \frac{1}{n} \right) \in \mathfrak{B}(\mathbb{R})$$

und damit auch $C = \bigcup_{i=1}^{\infty} \{x_i\} \in \mathfrak{B}(\mathbb{R})$.

2.1.2 Axiomatik der Wahrscheinlichkeitstheorie

- **Kolmogorowsche Axiome der Wahrscheinlichkeitstheorie**:

  Als Wahrscheinlichkeitsverteilung $P(\cdot)$ auf einer σ-Algebra $\mathfrak{A}$ von Teilmengen einer nichtleeren Menge Ω bezeichnet man jede Abbildung P mit

  (K1) $P(\Omega) = 1$

  (K2) Für jede abzählbar unendliche Folge $\{A_i\}_{i=1}^{\infty}$ aus $\mathfrak{A}$ mit $A_i \cap A_j = \emptyset$, für $i \neq j$, gilt

  $$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

- Sind Ω eine nichtleere Menge, $\mathfrak{A}$ eine σ-Algebra von Teilmengen von Ω und P eine Abbildung von $\mathfrak{A}$ in $[0, 1]$ mit den Eigenschaften (K1)-(K2), so heißt das Tripel $(\Omega, \mathfrak{A}, P)$ **Wahrscheinlichkeitsraum**.

- Folgerungen:

  Satz 2.9. *Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und seien $A_1, A_2, \ldots \in \mathfrak{A}$ beliebige Ereignisse. Dann gilt*

  (i) $P(\overline{A}) = 1 - P(A)$

  (ii) $P(A \cap B) = P(A) - P(A \cap \overline{B})$

  (iii) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

  (iv) $P(A \cup B) \leq P(A) + P(B)$

  (v) $A_1 \subset A_2 \Rightarrow P(A_1) \leq P(A_2)$.

  Beweis. (i) Es gilt $A \cup \overline{A} = \Omega$, A und $\overline{A}$ sind disjunkt. Daraus folgt

  $$\begin{aligned} P(\Omega) &= P(A \cup \overline{A}) \overset{(K2)}{=} P(A) + P(\overline{A}) \\ \Longrightarrow P(\overline{A}) &= P(\Omega) - P(A) \\ &= 1 - P(A) \end{aligned}$$

  (ii) Idee: $A = (A \cap B) \cup (A \cap \overline{B})$. $A \cap B$ und $A \cap \overline{B}$ sind disjunkt. Daher folgt aus dem 2. Axiom:

  $$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \overline{B}) \\ \Longrightarrow P(A \cap B) &= P(A) - P(A \cap \overline{B}) \\ \text{bzw.} \Longrightarrow P(A \cap \overline{B}) &= P(A) - P(A \cap B) \end{aligned}$$

  (iii) Idee:

  $$\begin{aligned} A \cup B &= ((A \cap B) \cup (A \cap \overline{B})) \cup ((B \cap A) \cup (B \cap \overline{A})) \\ &= (A \cap B) \cup (A \cap \overline{B}) \cup (B \cap \overline{A}) \\ \Longrightarrow P(A \cup B) &= P(A \cap B) + P(A \cap \overline{B}) + P(B \cap \overline{A}) \\ &= P(A \cap B) + P(A) - P(A \cap B) + P(B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

(iv) folgt unmittelbar aus (iii).

(v) $P(A_2) = P(A_1 \cup (A_2 \backslash A_1)) = P(A_1) + \underbrace{P(A_2 \backslash A_1)}_{\geq 0}$ und damit $P(A_1) \leq P(A_2)$.

□

Aus dem Beweis bzw. den Aussagen von Satz 2.9 ergibt sich sofort, dass

(i) $P(\emptyset) = 0$,

(ii) $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ für jede endliche Folge $A_1, \ldots, A_n \in \mathfrak{A}$ paarweise disjunkter Mengen,

(iii) $P(A_2 \backslash A_1) = P(A_2) - P(A_1)$,

(iv) $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$ für jede beliebige Folge $A_1, \ldots, A_n \in \mathfrak{A}$.

In Verallgemeinerung der 3. Teilaussage von Satz 2.9 ergibt sich außerdem die folgende *Siebformel*

Satz 2.10. *Für jedes $n = 1, 2, \ldots$ und jede Folge $A_1, \ldots, A_n \in \mathfrak{A}$ gilt*

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 \leq k_2 \leq \ldots \leq k_i \leq n} P(A_{k_1} \cap A_{k_2} \cap \ldots \cap A_{k_i}). \tag{2.1}$$

Darüber hinaus kann man mit Hilfe von Satz 2.9 zeigen, dass Wahrscheinlichkeitsmaße stetig sind bezüglich der monotonen Konvergenz von Mengen.

Folgerung 2.11. *Sei $A_1, A_2, \ldots \in \mathfrak{A}$. Dann gilt*

$$P\left(\bigcup_{i=1}^\infty A_i\right) = \lim_{i \to \infty} P(A_i), \quad \textit{falls} \quad A_1 \subset A_2 \subset \ldots \tag{2.2}$$

bzw.

$$P\left(\bigcap_{i=1}^\infty A_i\right) = \lim_{i \to \infty} P(A_i), \quad \textit{falls} \quad A_1 \supset A_2 \supset \ldots. \tag{2.3}$$

Die Subadditivität von Wahrscheinlichkeitsmaßen, die in Aussage (iv) von Satz 2.9 betrachtet wurde, gilt auch für Folgen von unendlich vielen Ereignissen.

Satz 2.12. *Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A_1, A_2, \ldots \in \mathfrak{A}$ eine Folge beliebiger Ereignisse. Dann gilt*

$$P(\bigcup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty P(A_i) \tag{2.4}$$

Als Folgerung ergibt sich das erste Lemma von Borel-Cantelli:

Folgerung 2.13. *Sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum und $A_1, A_2, \ldots \in \mathfrak{A}$ eine Folge beliebiger Ereignisse. Dann gilt*

$$P(\limsup A_n) = 0, \tag{2.5}$$

falls $\sum_{i=1}^\infty P(A_i) < \infty$.

2.1.3 Laplace-Experimente und Grundzüge der Kombinatorik

- Als Laplace-Experiment bezeichnet man einen zufälligen Versuch mit endlich vielen Ausgängen, die alle die gleiche Wahrscheinlichkeit besitzen. Als mathematisches Modell für ein Laplace-Experiment wählen wir den Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ mit

$$\Omega = \{1, 2, \ldots, n\}, \quad n < \infty \quad \Longrightarrow \quad |\Omega| = n,$$

sowie $\mathfrak{A} = \mathfrak{P}(\Omega)$ und $P(\{\omega_i\}) =: p$, $\forall \omega_i \in \Omega$. n heißt Parameter des Laplace-Experimentes.

Satz 2.14. *Bei einem Laplace-Experiment gilt für jede Teilmenge A des Ergebnisraumes Ω, dass*

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

Beweis. Es gilt

$$P(\{\omega\}) = p = \frac{1}{n}.$$

Begründung:

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^{n} \{\omega_i\}\right) \overset{(K2)}{=} \sum_{i=1}^{n} P(\{\omega_i\}) = n \cdot p.$$

Damit gilt für jede Teilmenge A von Ω

$$P(A) = P\left(\bigcup_{i:\omega_i \in A} \{\omega_i\}\right) = \sum_{i:\omega_i \in A} P(\{\omega_i\}) = |A|p = \frac{|A|}{n} = \frac{|A|}{|\Omega|}.$$

□

- Folgerung: Für die Berechnung der Wahrscheinlichkeiten bei Laplace-Experimenten ist es ausreichend, die Mächtigkeiten der Mengen A und Ω zu bestimmen. Dafür sind die folgenden Grundzüge der Kombinatorik hilfreich.

Anzahl möglicher Stichproben vom Umfang N aus Grundgesamtheit vom Umfang N:

- *Permutation*: $P(N) = N!$.

  Beispiel: Anordnung von 10 Büchern in Regal: $P(10) = 10!$.

- *Permutation bei r Gruppen*:

$$P^W(N|g_1, \ldots, g_r) = \frac{N!}{g_1! \cdot \ldots \cdot g_r!}$$

  mit r Gruppen mit jeweils gleichen Elementen. Es muss gelten $g_1 + \ldots + g_r = N$.

  Beispiel: Anordnung von 3 Statistik-Büchern und 7 Mathematik-Büchern: $P^W(10|3, 7) = \frac{10!}{3!7!}$.

Anzahl möglicher Stichproben vom Umfang n aus Grundgesamtheit vom Umfang N:

- *Kombination ohne Wiederholung*: Modell ohne Zurücklegen ohne Berücksichtigung der Reihenfolge $K(N, n) = \binom{N}{n}$.

  Beispiel: Vierköpfiger Wahlvorstand aus 30 Teilnehmern (ohne Zuordnung von Funktionen): $K(30, 4) = \binom{30}{4}$.

- *Kombination mit Wiederholung*: Modell mit Zurücklegen ohne Berücksichtigung der Reihenfolge $K^W(N, n) = \binom{N+n-1}{n}$.

  Beispiel: Anzahl verschiedener Würfe mit 2 Würfeln: $K^W(6, 2) = \binom{7}{2}$.

- *Variation ohne Wiederholung*: Modell ohne Zurücklegen mit Berücksichtigung der Reihenfolge $V(N,n) = \frac{N!}{(N-n)!}$.

  Beispiel: Vierköpfiger Wahlvorstand aus 30 Teilnehmern mit Zuordnung von Funktionen: $V(30,4) = \frac{30!}{26!}$.

- *Variation mit Wiederholung*: Modell mit Zurücklegen mit Berücksichtigung der Reihenfolge $V^W(N,n) = N^n$.

  Beispiel: Fußball-Toto für 12 Spiele: $V^W(3,12) = 3^{12}$.

2.1.4 Bedingte Wahrscheinlichkeiten

- Bedingte Wahrscheinlichkeiten sind folgendermaßen definiert:

  Definition 2.15. *Es seien $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum, $A, B \in \mathfrak{A}$ und $P(B) > 0$. Dann heißt*

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

  die ***bedingte Wahrscheinlichkeit*** *von A unter der Bedingung, daß das Ereignis B eintritt.*

- Sei Ω eine nichtleere Menge. Eine **Partition** $\{S_1, \ldots, S_k\}$ von Ω (der Ordnung k) ist eine Zerlegung $\Omega = \cup_{i=1}^k S_i$ in paarweise disjunkte, nichtleere Teilmengen S_i von Ω.

  Satz 2.16. ***Satz von der totalen Wahrscheinlichkeit***: *Für jedes Ereignis A gilt*

  $$P(A) = \sum_{i=1}^k P(A|S_i)P(S_i).$$

  Beweis.

  $$P(A) = P(A \cap \Omega) = P(A \cap (\cup_{i=1}^k S_i)) = P(\bigcup_{i=1}^k (A \cap S_i)) = \sum_{i=1}^k P(A \cap S_i) = \sum_{i=1}^k P(A|S_i)P(S_i).$$

  □

- **Spezialfall**: Wenn $S_1 = B$ und $S_2 = \overline{B}$ mit $0 \leq P(B) \leq 1$, so gilt:

  $$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B})$$

  Satz 2.17. ***Satz von Bayes***: *Es sei $\{S_1, \ldots, S_k\}$ eine Partition von Ω (der Ordnung k) und B ein Ereignis mit $P(B) > 0$. Es gilt für jedes $i = 1, \ldots, k$:*

  $$P(S_i|B) = \frac{P(B|S_i)P(S_i)}{\sum_{j=1}^k P(B|S_j)P(S_j)}.$$

  Beweis. Nach Definition der bedingten Wahrscheinlichkeit gilt

  $$P(S_i|B) = \frac{P(S_i \cap B)}{P(B)} = \frac{P(B|S_i)P(S_i)}{P(B)}.$$

  Auf den Nenner wird nun der Satz der totalen Wahrscheinlichkeit angewandt. □

- Es sei $(\Omega, \mathfrak{A}, P)$ ein Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \in \mathfrak{A}$ heißen voneinander stochastisch unabhängig (bezüglich der Wahrscheinlichkeitsverteilung P), falls

$$P(A \cap B) = P(A)P(B).$$

Im Fall $P(B) > 0$ bedeutet das $P(A|B) = P(A)$ und im Fall $P(A) > 0$ ergibt sich $P(B|A) = P(B)$. Die Unabhängigkeit ist eine symmetrische Eigenschaft bezüglich A und B.

Für eine Folge von Ereignissen $A_i \in \mathfrak{A}, i = 1, \ldots, n$ gilt

(i) Die A_i heißen voneinander paarweise stochastisch unabhängig unter der Wahrscheinlichkeitsverteilung P, wenn gilt

$$P(A_i \cap A_j) = P(A_i)P(A_j),\ i \neq j,\ i, j = 1, \ldots, n.$$

(ii) Die A_i heißen voneinander stochastisch unabhängig unter der Wahrscheinlichkeitsverteilung P, falls gilt

$$P(A_{i_1} \cap \ldots \cap A_{i_m}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_m}),$$

für je endlich viele verschiedene Indices $i_1, \ldots, i_m$.

2.2 Univariate Zufallsvariablen

2.2.1 Zufallsvariablen

- häufig: kein primäres Interesse an zugrundeliegenden Ergebnissen eines Zufallsexperiments mit Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$, sondern an einer daraus abgleiteten (quantitativen oder qualitativen) Kennzahl $X(\omega)$, d.h. wir betrachten die Abbildung $\omega \to X(\omega)$.

Beispiel 2.6. *Sei Ω die Menge von Eintragungen in einem Telefonbuch und ω der Familienname. Dann könnte man beispielsweise an $X(\omega)$ als der Anzahl der Buchstaben von ω interessiert sein. Alternativ könnte ω die Telefonnummer bezeichnen und $X(\omega)$ die Anzahl der Ziffer „1“ in ω.*

Beispiel 2.7. *Wir betrachten das zweimalige Würfeln mit Ergebnisraum $\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, 2, 3, 4, 5, 6\}, i = 1, 2\}$. Die Augensumme ergibt sich als Abbildung $\omega \to X(\omega) = \omega_1 + \omega_2$. Sei $A = \{\omega : X(\omega) = 10\} = \{\{4, 6\}, \{5, 5\}, \{6, 4\}\}$ bzw. allgemeiner $A = \{\omega : X(\omega) = k\}$ mit $k \in \{2, \ldots, 12\}$. Gesucht ist die Wahrscheinlichkeit $P(A)$. Dafür ist erforderlich, dass $A \in \mathfrak{A}$. Allgemein muss also gelten $\{\omega : \omega \in \Omega, X(\omega) = k\} \in \mathfrak{A}$ für jedes $k = 2, \ldots, 12$. In diesem Beispiel ist dies gleichbedeutend mit $\{\omega : \omega \in \Omega, X(\omega) \leq x\} \in \mathfrak{A}$ für jedes $x \in \mathbb{R}$.*

Dies führt zu folgender Begriffsbildung

Definition 2.18. *Sei $(\Omega, \mathfrak{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum. Die Abbildung $X : \Omega \mapsto E$ von Ω in eine Menge E (z.B. $E = \mathbb{R}, E = \mathbb{R}^n, E = \mathbb{N}_0$) heißt Zufallsvariable, falls*

$$\{\omega : \omega \in \Omega, X(\omega) \leq x\} \in \mathfrak{A} \quad \text{für alle } x \in E. \tag{2.6}$$

Bemerkungen:

(i) Die Regularitätsbedingung (2.6) wird Messbarkeit der Abbildung X bezüglich der σ-Algebra $\mathfrak{A}$ genannt.

(ii) In vielen Fällen interessiert nicht nur die Wahrscheinlichkeit, dass die Werte $X(\omega)$ der Zufallsvariablen X einen vorgegebenen Schwellenwert x nicht überschreiten, d.h. dass $X \in B = (-\infty, x]$, sondern dass X Werte in einer allgemeineren Teilmenge $B \subset E$ annehmen, wobei B beispielsweise die Vereinigung disjunkter Intervalle sein kann. Deshalb wird nicht nur im Ergebnisraum Ω, sondern auch im Bildraum E ein Mengensystem betrachtet, das bezüglich der Mengenoperationen $\cup, \cap$ und $\setminus$ abgeschlossen ist.

- Ist $E = \mathbb{R}$, so heißt X *univariate Zufallsvariable*, ist $E = \mathbb{R}^p, (p > 1)$, so heißt X *multivariate Zufallsvariable* (oder auch *Zufallsvektor*).
- Im folgenden beschränken wir uns auf den Fall, dass $E = \mathbb{R}$.

  Satz 2.19. *Die Abbildung $X : \Omega \to \mathbb{R}$ ist genau dann eine Zufallsvariable, wenn*
  $$\{\omega : \omega \in \Omega, X(\omega) \in B\} \in \mathfrak{A} \quad \textit{für alle } B \in \mathfrak{B}(\mathbb{R}). \tag{2.7}$$

- Jede Zufallsvariable besitzt eine Verteilung.

  Definition 2.20. *Sei $(\Omega, \mathfrak{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \mapsto \mathbb{R}$ eine beliebige (univariate) Zufallsvariable. Die* Verteilung *der Zufallsvariablen X ist die Mengenfunktion $P_X : \mathfrak{B}(\mathbb{R}) \to [0,1]$ mit*
  $$P_X(B) = P(\{\omega : \omega \in \Omega, X(\omega) \in B\}) \quad \textit{für alle } B \in \mathfrak{B}(\mathbb{R}). \tag{2.8}$$

  Bemerkung: Die in (2.8) definierte Mengenfunktion P_X ist ein Wahrscheinlichkeitsmaß auf dem Meßraum $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, denn P_X ist

  (i) normiert, weil $P_X(\mathbb{R}) = P(\Omega) = 1$,

  (ii) σ-additiv, weil für paarweise disjunkte $B_1, B_2, \ldots \in \mathfrak{B}(\mathbb{R})$ gilt
  $$P_X\left(\bigcup_{i=1}^{\infty} B_i\right) = P(X^{-1}(\bigcup_{i=1}^{\infty} B_i)) = P(\bigcup_{i=1}^{\infty} X^{-1}(B_i)) = \sum_{i=1}^{\infty} P(X^{-1}(B_i)) = \sum_{i=1}^{\infty} P_X(B_i).$$

  Die Abbildung $P \to P_X$ nennt man auch *Maßtransport* vom Meßraum $(\Omega, \mathfrak{A})$ in den Meßraum $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$.

  Die folgenden Kurzschreibweisen sind üblich

  (i) $P(X \in B) = P(\{\omega : \omega \in \Omega, X(\omega) \in B\})$ für alle $B \in \mathfrak{B}(\mathbb{R})$ und

  (ii) speziell $P(X \le x) = P(\{\omega : \omega \in \Omega, X(\omega) \le x\})$ für alle $x \in \mathbb{R}$.

- Verteilungsfunktion

  Definition 2.21. *Sei $(\Omega, \mathfrak{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$. Die Funktion $F_X : \mathbb{R} \to [0,1]$ mit*
  $$F_X(x) = P(X \le x), \quad \textit{für alle } x \in \mathbb{R}$$
  heißt Verteilungsfunktion *von X.*

  Eine Verteilungsfunktion besitzt folgende Eigenschaften

  Satz 2.22. *Sei $X : \Omega \to \mathbb{R}$ eine beliebige Zufallsvariable und $F_X : \mathbb{R} \to [0,1]$ ihre Verteilungsfunktion. Dann gilt*

  (i) Monotonie:
  $$F_X(x) \le F_X(x+h) \quad \forall x \in \mathbb{R} \textit{ und } h \ge 0,$$

  (ii) Asymptotisches Verhalten im Unendlichen:
  $$F_X(-\infty) := \lim_{x \to -\infty} F_X(x) = 0, \quad F_X(\infty) := \lim_{x \to \infty} F_X(x) = 1,$$

  (iii) Rechtsstetigkeit: $F_X(x)$ ist rechtsseitig stetig, d.h. für jede Folge $\{h_n\}$ mit $h_n \ge 0$ und $\lim_{n\to\infty} h_n = 0$ gilt
  $$\lim_{n \to \infty} F_X(x + h_n) = F_X(x), \quad \forall x \in \mathbb{R}.$$

Satz 2.23. *Sei $X : \Omega \to \mathbb{R}$ eine beliebige Zufallsvariable. Dann wird die Verteilung P_X von X durch die Verteilungsfunktion F_X eindeutig bestimmt.*

Bemerkungen:

(i) Mit Hilfe der Verteilungsfunktion F_X lassen sich auch die folgenden Wahrscheinlichkeiten ausdrücken

$$P(a \leq X \leq b), \quad P(a < X \leq b), \quad P(a < X < b), \quad P(a \leq X < b),$$

denn es gilt beispielsweise

$$P(a \leq X \leq b) = P(\{X \leq b\} \backslash \{X < a\}) = P(\{X \leq b\}) - P(\{X < a\}) = F_X(b) - \lim_{h \downarrow 0} F_X(a-h).$$

(ii) Im Allgemeinen gilt jedoch nicht $F_X(a) = \lim_{h \downarrow 0} F_X(a-h)$, sondern

$$F_X(a) = \lim_{h \downarrow 0} F_X(a-h) + P(X = a). \tag{2.9}$$

Begründung:

$$\begin{aligned} P(X = a) &= P(\bigcap_{n=1}^{\infty} \{a - \frac{1}{n} < X < a + \frac{1}{n}\}) \overset{\text{Satz 2.22 (iii)}}{=} P(\bigcap_{n=1}^{\infty} \{a - \frac{1}{n} < X \leq a\}) \\ &\overset{\text{Folgerung 2.11}}{=} \lim_{n \to \infty} P(a - \frac{1}{n} < X \leq a) \\ &= \lim_{n \to \infty} (P(X \leq a) - P(X \leq a - \frac{1}{n})) \\ &= F_X(a) - \lim_{n \to \infty} F_X(a - \frac{1}{n}) \end{aligned}$$

- Diskrete vs. stetige Zufallsvariablen.

Definition 2.24. *Die Zufallsvariable X (bzw. ihre Verteilung) heißt* diskret*, falls es eine abzählbare Teilmenge $C \in \mathbb{R}$ gibt, so dass $P(X \in C) = 1$.*

Begriff Stetigkeit: Eine Variable oder ein Merkmal X heißt *stetig*, falls zwischen zwei beliebigen Werten $a < b$ des Definitionsbereiches überabzählbar viele Zwischenwerte möglich sind.

Falls die Werte von X als Ergebnisse eines Zufallsvorgangs resultieren, wird X zu einer *stetigen Zufallsvariable.*

Definition 2.25. *Eine Zufallsvariable X heißt* ***stetig****, falls es eine integrierbare Funktion $f_X : \mathbb{R} \to [0, \infty)$ mit nichtnegativen Werten gibt, so daß für jedes Intervall $[a, b]$*

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

gilt. Die Funktion $f_X(x)$ heißt Dichtefunktion *(oder Wahrscheinlichkeitsdichte) von X.*

Eigenschaften der Dichtefunktion:

(i) Für stetige Zufallsvariablen X gilt

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

(ii) $P(X = x) = 0$. Daraus folgt, daß $f(x)$ keine Wahrscheinlichkeit ist. Die Dichten können daher auch Werte $f(x) > 1$ annehmen.

(iii) Nichtnegativität

(iv) Normierungseigenschaft:

$$\int_{-\infty}^{\infty} f(x)dx = 1,$$

d.h. die Gesamtfläche zwischen x-Achse und der Dichte $f(x)$ ist gleich 1.

Bemerkungen:

(i) In der Wahrscheinlichkeitstheorie wird vorausgesetzt, dass f_X eine Lebesgue-integrierbare Funktion ist. Das Integral wird im allgemeinen als Lebesgue-Integral aufgefasst.

(ii) Bei vielen Anwendungen ist f_X eine (zumindest stückweise) stetige Funktion. Das Integral in Definition 2.25 ist dann ein Riemann-Integral.

(iii) Falls X stetig ist und die Verteilungsfunktion F_X keine Sprünge aufweist, dann folgt aus (2.9), dass $P(X = x) = 0$ für alle $x \in \mathbb{R}$.

(iv) Die Verteilungsfunktion F_X einer stetigen Zufallsvariablen X ist im allgemeinen nicht überall differenzierbar. Und zwar ist F_X dort nicht differenzierbar, wo die Dichte f_X Sprungstellen besitzt.

(v) Die Verteilungsfunktion einer stetigen Zufallsvariable wird eindeutig durch ihre Dichte bestimmt.

- Unterscheidung von diskreten und stetigen Zufallsvariablen.
  - Diskrete Zufallsvariablen haben einen abzählbaren Wertebereich, z.B. wenn $\Omega = X^{-1}(C)$ mit $C \subset \mathbb{N} \subset \mathbb{R}$. Sei beispielsweise X die Augensumme beim zweimaligen Würfeln. Dann gilt $X : \Omega \to \{2, 3, \ldots, 12\}$.
  - Stetige Zufallsvariablen haben einen überabzählbaren Wertebereich, z.B. $[a, b], [a, \infty), (-\infty, 0]$ oder $\mathbb{R}$. Betrachten wir z.B. ein Roulettespiel mit drehbarem Zeiger und „kontinuierlicher" Skala, wobei X der Wert des Spiels sei (Winkel des Zeigers), dann erhalten wir $X : \Omega \to [0, 2\pi)$.
- Wahrscheinlichkeitsfunktion diskreter Zufallsvariablen

Definition 2.26. *Sei X eine diskrete Zufallsvariable, d.h. es gebe eine abzählbare Menge $C = \{x_1, x_2, \ldots\}$, so dass $P(X \in C) = 1$. Dann heißt die Folge $p_1, p_2, \ldots$ mit $p_k = P(X = k)$* Wahrscheinlichkeitsfunktion *von X.*

Bemerkungen:

(i) Für jede Wahrscheinlichkeitsfunktion $\{p_k\}$ gilt offenbar $p_k \geq 0$ für alle $k = 1, 2, \ldots$ und $\sum_{k=1}^{\infty} p_k = 1$.

(ii) Die Verteilung einer diskreten Zufallsvariablen X wird eindeutig durch die Wahrscheinlichkeitsfunktion $\{p_k\}$ bestimmt, denn es gilt für jedes $B \in \mathfrak{B}(\mathbb{R})$

$$P_X(B) = P_X(\bigcup_{i:x_i \in B} \{x_i\}) = \sum_{i:x_i \in B} P_X(\{x_i\}) = \sum_{i:x_i \in B} p_i.$$

(iii) Für jedes $x_k \in C$ heißt die Zahl $p_k = P(X = x_k)$ *Einzelwahrscheinlichkeit.*

2.2.2 Transformation von (univariaten) Zufallsvariablen

- Gegeben sei eine Zufallsvariable X mit Verteilungsfunktion F_X und Träger $\mathcal{X}$ und eine zweite Zufallsvariable $Y = g(X)$ mit Träger $\mathcal{Y}$ definiert über die Funktion

$$g(x) : \mathcal{X} \to \mathcal{Y},$$

so dass $P(Y \in A) = P(g(X) \in A)$.

- Frage: Können wir mit Hilfe von F_X und $g(X)$ auf die Verteilungsfunktion von Y schließen? Antwort: Ja!

Lemma 2.27. *Sei X eine Zufallsvariable mit Verteilungsfunktion $F_X(x)$ und $Y = g(X)$, mit g streng monoton. Dann gilt*

(i) Ist g streng monoton steigend in $\mathcal{X}$, dann gilt $F_Y(y) = F_X(g^{-1}(y))$ für $y \in \mathcal{Y}$,

(ii) Ist g streng monoton fallend in $\mathcal{X}$, dann gilt $F_Y(y) = 1 - F_X(g^{-1}(y))$ für $y \in \mathcal{Y}$.

Beweis. Für eine streng monoton steigende Funktion g gilt, dass $u > v \Rightarrow g(u) > g(v)$. Außerdem besitzt eine streng monotone Transformation eine Umkehrfunktion, d.h. $y = g(x) \Leftrightarrow g^{-1}(y) = x$. Damit erhalten wir

$$
\begin{aligned}
F_Y(y) = P(Y \leq y) = P(g(X) \leq y) &= P(\{x \in \mathcal{X} : g(X) \leq y\}) \\
&= P(\{x \in \mathcal{X} : x \leq g^{-1}(y)\}) \\
&= \int_{\{x \in \mathcal{X} : x \leq g^{-1}(y)\}} f_X(x) dx \\
&= \int_{-\infty}^{g^{-1}(y)} f_X(x) dx \\
&= F_X(g^{-1}(y)).
\end{aligned}
$$

Für g streng monoton fallend gilt $u > v \Rightarrow g(u) < g(v) \Rightarrow g^{-1}(u) < g^{-1}(v)$. Damit erhalten wir

$$\{x \in \mathcal{X} : g(x) \leq y\} = \{x \in \mathcal{X} : g^{-1}(g(x)) \geq g^{-1}(y)\} = \{x \in \mathcal{X} : x \geq g^{-1}(y)\}.$$

Diese Darstellung des Ereignisses verwenden wir im Folgenden

$$F_Y(y) = \int_{\{x \in \mathcal{X} : x \geq g^{-1}(y)\}} f_X(x) dx = \int_{g^{-1}(y)}^{\infty} f_X(x) dx = 1 - F_X(g^{-1}(y)).$$

□

- Mit Hilfe von Lemma 2.27 können wir auch die **Dichtefunktion** der transformierten Zufallsvariable berechnen.

Satz 2.28. *Sei $X \sim f_X(x)$ und $Y = g(X)$ eine streng monotone Transformation. $f_X(x)$ sei stetig in $\mathcal{X}$ und $g^{-1}(y)$ besitze eine stetige Ableitung in $\mathcal{Y}$. Dann gilt*

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|, & y \in \mathcal{Y}, \\ 0, & \textit{sonst.} \end{cases}$$

Beweis. Mit Hilfe der Kettenregel erhalten wir

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y), & \text{falls } g(\cdot) \text{ streng monoton steigend ist,} \\ -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y), & \text{falls } g(\cdot) \text{ streng monoton fallend ist.} \end{cases}$$

Da $\frac{d}{dy} g^{-1}(y) > 0$, falls $g(\cdot)$ streng monoton steigend ist bzw. $\frac{d}{dy} g^{-1}(y) < 0$, falls $g(\cdot)$ streng monoton fallend ist, ist $f_Y(y)$ in beiden Fällen stets nicht-negativ. □

2.2.3 Parameter univariater Zufallsvariablen

- Modus: $x_{\text{mod}} = \arg\max_x f(x)$,
- Quantil: Jeder Wert x_p mit $0 < p < 1$, für den

$$P(X \geq x_p) \geq 1 - p \quad \text{und} \quad P(X \leq x_p) \geq p$$

  gilt, heißt p-Quantil einer diskreten Verteilung. Analog gilt bei stetigen Verteilungen: Jeder Wert x_p mit $F(x_p) = p$ heißt p-Quantil einer stetigen Verteilung. Spezialfall: 50%-Quantil heisst Median.

- Median: $x_{\text{med}} = x_{0.5}$

  Definition 2.29. *Der Erwartungswert einer Zufallsvariable $g(X)$ wird mit $E[g(X)]$ bezeichnet und ist definiert als*

$$E[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x)f(x)dx, & \textit{falls } X \textit{ stetig ist,} \\ \sum_{x:f_X(x)>0} g(x)f_X(x) = \sum_{x:f_X(x)>0} g(x)P(X=x), & \textit{falls } X \textit{ diskret ist.} \end{cases}$$

- Bemerkung: Die Definition ist auch mit Hilfe des Lebesgue-Integrals möglich. Gilt $E|g(X)| = \infty$, so heißt der Erwartungswert nicht existent.

  Beispiel 2.8. *Eine Cauchy-verteilte Zufallsvariable besitzt keinen Erwartungswert! Die Dichtefunktion lautet*

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

  $f_X(x)$ ist zwar eine Dichtefunktion, d.h. es gilt $\int_{-\infty}^{\infty} f_X(x)dx = 1$, aber es gilt auch $E|X| = \infty$. Beweis:

$$\begin{aligned} E|X| = \int_{-\infty}^{\infty} \frac{|x|}{\pi}\frac{1}{1+x^2}dx &= \underbrace{\int_{-\infty}^{0} -\frac{x}{\pi}\frac{1}{1+x^2}dx}_{=\int_0^\infty \frac{x}{\pi}\frac{1}{1+x^2}dx} + \int_0^{\infty} \frac{x}{\pi}\frac{1}{1+x^2}dx \\ &= \frac{2}{\pi}\int_0^{\infty}\frac{x}{1+x^2}dx. \end{aligned}$$

  Für $M > 0$ gilt

$$\int_0^M \frac{x}{1+x^2}dx = \left.\frac{\log(1+x^2)}{2}\right|_0^M = \frac{\log(1+M^2)}{2} - \frac{\log 1}{2} = \frac{\log(1+M^2)}{2}$$

  und damit

$$E|X| = \lim_{M\to\infty} \frac{2}{\pi}\int_0^M \frac{x}{1+x^2}dx = \frac{1}{\pi}\lim_{M\to\infty} \log(1+M^2) = \infty.$$

- Eigenschaften des Erwartungswertoperators

  Satz 2.30. *Sei X eine Zufallsvariable, a, b, c Konstanten und $E|g_1(X)| < \infty, E|g_2(X)| < \infty$. Dann gilt*

  (i) $E[ag_1(X) + bg_2(X) + c] = aE[g_1(X)] + bE[g_2(X)] + c$ *(Linearität),*
  (ii) Ist $g_1(x) \geq 0$ für alle x, so gilt $E[g_1(X)] \geq 0$,
  (iii) Ist $g_1(x) \geq g_2(x)$ für alle x, so gilt auch $E[g_1(X)] \geq E[g_2(X)]$,
  (iv) Ist $a \leq g_1(x) \leq b$ für alle x, so gilt auch $a \leq E[g_1(X)] \leq b$.

  Beweis. Die folgenden Beweise sind für den stetigen Fall angegeben. Für diskrete Zufallsvariablen erfolgen sie analog.

(i)

$$\begin{aligned}
E[ag_1(X) + bg_2(X) + c] &= \int_{-\infty}^{\infty} (ag_1(x) + bg_2(x) + c) f_X(x)dx \\
&= \int_{-\infty}^{\infty} ag_1(x) f_X(x)dx + \int_{-\infty}^{\infty} bg_2(x) f_X(x)dx + \int_{-\infty}^{\infty} cf_X(x)dx \\
&= a \int_{-\infty}^{\infty} g_1(x) f_X(x)dx + b \int_{-\infty}^{\infty} g_2(x) f_X(x)dx + c \underbrace{\int_{-\infty}^{\infty} f_X(x)dx}_{=1 \text{ (Normierungseigenschaft)}} \\
&= aE[g_1(X)] + bE[g_2(X)] + c.
\end{aligned}$$

(ii) Für $g_1(x) \geq 0$, für alle $x \in \mathbb{R}$ gilt $g_1(x) f_X(x) \geq 0$, da $f_X(x) \geq 0$, für alle $x \in \mathbb{R}$. Damit gilt auch, dass

$$\int_{-\infty}^{\infty} g_1(x) f_X(x)dx \geq 0$$

und damit $E[g_1(X)] \geq 0$.

(iii) Analog zum Beweis von (ii) folgt aus $g_1(x) \geq g_2(x)$, für alle $x \in \mathbb{R}$, dass $g_1(x) f_X(x) \geq g_2(x) f_X(x)$, da $f_X(x) \geq 0$, für alle $x \in \mathbb{R}$. Damit gilt auch, dass

$$\int_{-\infty}^{\infty} g_1(x) f_X(x)dx \geq \int_{-\infty}^{\infty} g_2(x) f_X(x)dx$$

und schließlich $E[g_1(X)] \geq E[g_1(X)]$.

(iv) Analog zu (ii) und (iii) zeigt man, dass aus $a \leq g_1(x) \leq b$, für alle $x \in \mathbb{R}$ gilt, dass $af_X(x) \leq g_1(x) f_X(x) \leq bf_X(x)$, da $f_X(x) \geq 0$, für alle $x \in \mathbb{R}$. Damit gilt auch, dass

$$a \int_{-\infty}^{\infty} f_X(x)dx \leq \int_{-\infty}^{\infty} g_1(x) f_X(x)dx \leq b \int_{-\infty}^{\infty} f_X(x)dx$$

und daher $a \leq E[g_1(X)] \leq b$.

□

- Bemerkung: Aussage (i) impliziert $E(aX + b) = aE(X) + b$ und $E(X + Y) = E(X) + E(Y)$.
- Lageregeln: Die Verteilung heißt
  (i) symmetrisch, unimodal, wenn $x_{\text{mod}} = x_{\text{med}} = E(X)$,
  (ii) rechtsschief, wenn $x_{\text{mod}} < x_{\text{med}} < E(X)$,
  (iii) linksschief, wenn $x_{\text{mod}} > x_{\text{med}} > E(X)$.

Definition 2.31. *Das n-te Moment* ($n \in \mathbb{N}$) *einer Zufallsvariable X ist definiert als*

$$\mu'_n := E(X^n).$$

Das n-te zentrale Moment ist definiert als

$$\mu_n = E[(X - \mu)^n], \quad \text{mit} \quad \mu = \mu'_1 = E(X).$$

- Bemerkung: Das zweite zentrale Moment heißt Varianz $Var(X) = E[(X - \mu)^2] = \sigma_X^2$. Die Standardabweichung σ_X ist die Wurzel aus der Varianz.
- Ziel: Beschreibung der Verteilung einer Zufallsvariablen durch einzelne “Kennzahlen“ (quasi Informationsverdichtung)

- Eigenschaften der Varianz

  Satz 2.32. *Es gilt*

  (i) $Var(X) = E[(X-\mu)^2] = E(X^2) - E(X)^2$ *bzw. verallgemeinert* $Var(X) = E[(X-c)^2] - (\mu-c)^2$ *(Verschiebungssatz),*

  (ii) Lineare Transformationen: Für $Y = aX + b$ *gilt*

  $$Var(Y) = Var(aX + b) = a^2 Var(X) \quad \textit{und} \quad \sigma_Y = |a|\sigma_X,$$

  Beweis. (i) $E[(X-\mu)^2] = E[X^2 - 2X\mu + \mu^2] = E[X^2] - 2E[X]\mu + \mu^2 = E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2$.

  (ii)

  $$\begin{aligned} Var(aX + b) &= E[\{(aX + b) - E(aX + b)\}^2] \\ &= E[\{aX - aE(X)\}^2] \\ &= E[a^2 \{X - E(X)\}^2] \\ &= a^2 E[\{X - E(X)\}^2] = a^2 Var(X). \end{aligned}$$

  □

- Bemerkung zu (ii): Für allgemeine Transformationen $Y = g(X)$ lässt sich die Varianz i.d.R. nur approximativ mit Hilfe der Delta-Methode bestimmen.

  Definition 2.33. *Das dritte zentrale (standardisierte) Moment*

  $$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{\sigma^3}$$

  heißt Schiefe (Skewness).

- Es gilt

  - $\gamma_1 > 0 \Longrightarrow$ Verteilung ist asymmetrisch zu positiven Argumenten,
  - $\gamma_1 < 0 \Longrightarrow$ Verteilung ist asymmetrisch zu negativen Argumenten,
  - $\gamma_1 = 0 \Longrightarrow$ Verteilung hat dieselbe Schiefe wie die Standardnormalverteilung.

  Definition 2.34. *Das vierte zentrale (standardisierte) Moment*

  $$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{E[(X-\mu)^4]}{\sigma^4}$$

  heißt Wölbung (Kurtosis).

- $\gamma_2 = 0$ entspricht der Wölbung der Standardnormalverteilung.

Momentenerzeugende Funktionen

- Idee erzeugender Funktionen: Speicherung der Information einer Folge $a = \{a_1, a_2, \ldots\}$ von reellen Zahlen in einer Funktion.

Beispiel 2.9. *Es sei* $G_a(s) = \sum_{i=0}^{\infty} a_i s^i$ *für alle* $s \in \mathbb{R}$*, für die die Summe konvergiert. Wir betrachten* $G_a^{(j)}(s) := \frac{d^j}{ds} G_a(s)$ *die* j*-te Ableitung. Es gilt*

$$\begin{aligned} G_a^{(1)}(s) &= \sum_{i=1}^{\infty} i \cdot a_i \cdot s^{i-1} \\ G_a^{(2)}(s) &= \sum_{i=2}^{\infty} i \cdot (i-1) \cdot a_i \cdot s^{i-2} \\ &\vdots \\ G_a^{(j)}(s) &= \sum_{i=j}^{\infty} i \cdot (i-1) \dots (i-j+1) \cdot a_i \cdot s^{i-j} \end{aligned}$$

Insbesondere gilt

$$\begin{aligned} G_a^{(j)}(0) &= \sum_{i=j}^{\infty} i \cdot (i-1) \dots (i-j+1) \cdot a_i \cdot 0^{i-j}, \quad 0^0 := 1 \\ &= \underbrace{j(j-1) \cdot \ldots \cdot 1}_{=j!} \cdot a_j + \sum_{i=j+1}^{\infty} i(i-1) \cdot \ldots (i-j+1) \cdot a_i \cdot 0^{i-j} \\ &= j! \cdot a_j \\ \Leftrightarrow a_j &= \frac{1}{j!} G_a^{(j)}(0). \end{aligned}$$

Folgerung: Man kann die Folge $a = \{a_1, a_2, \ldots\}$ *über die Ableitungen von* $G_a(s)$ *erzeugen.*

Beispiel 2.10. *exponential-erzeugende Funktion: Sei*

$$M_a(s) = \sum_{i=0}^{\infty} a_i \frac{s^i}{i!}$$

für alle $s \in \mathbb{R}$*, für die die Summe konvergiert. Dann gilt* $M_a^{(j)}(0) = a_j$*.*

- Frage: Können wir eine erzeugende Funktion definieren, so dass $M_X^{(j)}(0) = E[X^j]$? Antwort: Ja!

  Definition 2.35. *Sei* X *eine Zufallsvariable mit Verteilungsfunktion* F_X*. Die momentenerzeugende Funktion (MEF) von* X *ist definiert als*

  $$M_X(t) = E[e^{tX}],$$

  vorausgesetzt, dass der Erwartungswert für t *in einer Umgebung um* $t = 0$ *existiert (d.h.* $\exists h > 0 : \forall t \in [-h < t < h] \Rightarrow E[e^{tX}]$ *existiert).*

- Folgerung: Nach der Definition von MEFs und der Definition von Erwartungswerten erhalten wir

  $$M_X(t) = \begin{cases} \int_{-\infty}^{\infty} e^{tx} f_X(x) dx, & \text{falls } X \text{ stetig ist,} \\ \sum_{x:p(x)>0} e^{tx} P(X = x), & \text{falls } X \text{ diskret ist.} \end{cases}$$

  Bemerkung: Der stetige Fall entspricht übrigens der Laplace-Transformierten von f_X.

  Satz 2.36. *Es gilt*

  $$E[X^n] = M_X^{(n)}(0).$$

Beweis. Für $n = 1$ erhalten wir (im stetigen Fall)

$$\begin{aligned}\frac{d}{dt}M_X(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f_X(x)dx = \int_{-\infty}^{\infty} \frac{d}{dt} e^{tx} f_X(x)dx \\ &= \int_{-\infty}^{\infty} x e^{tx} f_X(x)dx \\ &= E[Xe^{tX}].\end{aligned}$$

Für $t = 0$ erhalten wir

$$\frac{d}{dt}M_X(t)\Big|_{t=0} = E[Xe^{tX}]\big|_{t=0} = E[X].$$

Den Fall $n > 1$ beweist man analog per Induktion. □

Beispiel 2.11. *MEF der Binomialverteilung. Es gilt*

$$\begin{aligned} M_X(t) = \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} &= \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x(1-p)^{n-x} \\ &\overset{(1)}{=} [pe^t + (1-p)]^n \\ \Longrightarrow \frac{d}{dt}M_X(t) &\overset{Kettenregel}{=} n[pe^t + (1-p)]^{n-1}pe^t \\ \Longrightarrow \frac{d}{dt}M_X(t)\Big|_{t=0} &= n[p + (1-p)]^{n-1}p = np, \end{aligned}$$

wobei wir in Relation (1) *den Binomischen Lehrsatz*

$$\sum_{x=0}^{n}\binom{n}{x}u^x v^{n-x} = (u+v)^n$$

mit $u = pe^t$ *und* $v = 1 - p$ *verwendet haben.*

2.2.4 Diskrete Verteilungen

Viele praktische Zufallsexperimente mit gleichwahrscheinlichen Elementarereignissen lassen sich am sogenannten Urnenmodell veranschaulichen. Im Urnenmodell besteht das Zufallsexperiment darin, aus einer Urne mit einer festgelegten Anzahl von Kugeln zweier Farben eine oder mehrere Kugeln nach einem bestimmten Plan zu ziehen. Im Folgenden werden die wichtigsten daraus abgeleiteten Verteilungen vorgestellt.

Die Bernoulli-Verteilung

- **Idee**: Betrachte eine binäre (dichotome) Zufallsvariable X mit

  $$X = \begin{cases} 1, & \text{falls } A \text{ eintritt,} \\ 0, & \text{sonst} \end{cases}$$

  also $\mathcal{T} = \{0, 1\}$.

- Es handelt sich hierbei um ein sogenanntes **Bernoulli-Experiment** (= man ist nur daran interessiert, ob ein bestimmtes Ereignis A eintritt oder nicht).

- Sei $P(A) := p$. Daraus folgt $P(A^c) = 1 - p$. Für die Wahrscheinlichkeitsverteilung von X ergibt sich damit

  $$P(X = 1) = p, \quad \text{und} \quad P(X = 0) = 1 - p.$$

  Diese Verteilung heißt **Bernoulli-Verteilung**. Diskrete Zufallsvariablen mit dieser Wahrscheinlichkeitsverteilung heißen Bernoulli-verteilt mit Parameter p, man schreibt auch $X \sim Ber(p)$.

- **Eigenschaften**:
  (i) $E(X) = p$
  (ii) $Var(X) = p(1-p)$

  Beweis:

  (i) $E(X) = \sum_{x:p(x)>0} p_i x_i = p \cdot 1 + (1-p) \cdot 0 = p$
  (ii)

$$\begin{aligned} Var(X) &= \sum_{x:p(x)>0} (x_i - p)^2 p_i = (1-p)^2 p + (0-p)^2(1-p) \\ &= p - 2p^2 + p^3 + p^2 - p^3 = p - p^2 = p(1-p) \end{aligned}$$

Die geometrische Verteilung

- **Idee**: Betrachte eine Folge unabhängiger Bernoulli-Experimente (sog. **Bernoulli-Kette**)
- **Frage**: Wieviele Versuche sind notwendig, bis zum ersten Mal A eintritt?
- Sei X die Anzahl der Versuche, bis zum ersten Mal A eintritt. Dann gilt

$$p_k = P(X = k) = (1-p)^{k-1} p, \ k = 1, 2, 3, \ldots .$$

  Da die Wahrscheinlichkeiten p_k eine geometrische Folge bilden (p_k und p_{k+1} unterscheiden sich nur durch den Faktor $1-p$), heißt die Verteilung geometrisch und die Zufallsvariable X geometrisch verteilt mit dem Parameter p. Man schreibt $X \sim G(p)$.

- **Eigenschaften**:
  (i) $E(X) = \frac{1}{p}$
  (ii) $Var(X) = \frac{1-p}{p^2}$
  (iii) $P(X \geq k) = (1-p)^{k-1}$
  (iv) Gedächtnislosigkeit: $P(X = n + k | X > n) = P(X = k)$

  Beweis:

(i)

$$
\begin{aligned}
E(X) &= \sum_{x:p(x)>0} x \cdot p(x) \\
&= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= p \sum_{k=1}^{\infty} \sum_{n=1}^{k} 1(1-p)^{k-1} \\
&= p \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} (1-p)^{k-1} \\
&= p \sum_{n=1}^{\infty} (1-p)^{n-1} \sum_{m=0}^{\infty} (1-p)^{m} \\
&= p \sum_{n=1}^{\infty} (1-p)^{n-1} \frac{1}{p} \\
&= \sum_{n=0}^{\infty} (1-p)^{n} \\
&= \frac{1}{p}
\end{aligned}
$$

(ii)

$$
\begin{aligned}
Var(X) &= E(X^2) - E(X)^2 \\
E(X^2) &= \sum_{x:p(x)>0} x^2 \cdot p(x) \\
&= \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p \\
&= \sum_{k=1}^{\infty} (k(k+1) - k)(1-p)^{k-1} p \\
&= \sum_{k=1}^{\infty} k(k+1)(1-p)^{k-1} p - \sum_{k=1}^{\infty} k(1-p)^{k-1} p \\
&= \sum_{k=1}^{\infty} k(k+1)(1-p)^{k-1} p - \frac{1}{p} \\
&\overset{(1)}{=} 2 \sum_{k=1}^{\infty} \sum_{n=1}^{k} n(1-p)^{k-1} p - \frac{1}{p} \\
&= 2p \sum_{k=1}^{\infty} \sum_{n=1}^{k} n(1-p)^{k-1} - \frac{1}{p} \\
&= 2p \sum_{n=1}^{\infty} n \sum_{k=n}^{\infty} (1-p)^{k-1} - \frac{1}{p} \\
&= 2p \sum_{n=1}^{\infty} n(1-p)^{n-1} \underbrace{\sum_{m=0}^{\infty} (1-p)^m}_{=1/p} - \frac{1}{p} \\
&= 2 \sum_{n=1}^{\infty} n(1-p)^{n-1} - \frac{1}{p} \\
&= 2 \sum_{n=1}^{\infty} (n-1)(1-p)^{n-1} + 2 \sum_{n=1}^{\infty} (1-p)^{n-1} - \frac{1}{p} \\
&= 2 \frac{1-p}{p^2} + 2 \frac{1}{p} - \frac{1}{p} \\
&= 2 \frac{1}{p^2} - \frac{1}{p} \\
\Longrightarrow Var(X) &= E(X^2) - E(X)^2 \\
&= 2 \frac{1}{p^2} - \frac{1}{p} - \left(\frac{1}{p}\right)^2 \\
&= \frac{1}{p^2} - \frac{1}{p} \\
&= \frac{1-p}{p^2}
\end{aligned}
$$

(1) ergibt sich aus $1 + 2 + ... + k = \frac{k(k+1)}{2}$.

(iii)

$$\begin{aligned}
P(X \geq k) &= \sum_{i=k}^{\infty}(1-p)^{i-1}p \\
&= p\sum_{i=k}^{\infty}(1-p)^{i-1} \\
&\overset{(1)}{=} p(1-p)^{k-1}\sum_{\ell=0}^{\infty}(1-p)^{\ell} \\
&= \frac{p(1-p)^{k-1}}{p} \\
&= (1-p)^{k-1}
\end{aligned}$$

(1) überlegt man sich folgendermaßen:

$$\begin{aligned}
\sum_{i=k}^{\infty}(1-p)^{i-1} &= (1-p)^{k-1} + (1-p)^{k} + (1-p)^{k+1} + \dots \\
&= (1-p)^{k-1}(1 + (1-p)^1 + (1-p)^2 + \dots \\
&= (1-p)^{k-1}\sum_{\ell=0}^{\infty}(1-p)^{\ell}
\end{aligned}$$

(iv)

$$\begin{aligned}
P(X = n+k|X > n) &= \frac{P(\{X = n+k\} \cap \{X > n\})}{P(\{X > n\})} \\
&= \frac{P(\{X = n+k\})}{P(\{X > n\})} \\
&= \frac{(1-p)^{n+k-1}p}{P(X \geq n) - P(X = n)} \\
&= \frac{(1-p)^{n+k-1}p}{(1-p)^{n-1} - (1-p)^{n-1}p} \\
&= \frac{(1-p)^{n+k-1}p}{(1-p)^{n-1}(1-p)} \\
&= \frac{(1-p)^{n+k-1}p}{(1-p)^{n}} \\
&= (1-p)^{k-1}p \\
&= P(X = k)
\end{aligned}$$

Die Binomialverteilung

- Wir betrachten eine Bernoulli-Kette von n-mal (unabhängig von einander) wiederholten Bernoulli-Experimenten mit $p = P(A)$ für ein uns interessierendes Ereignis A.
- **Frage**: In wievielen Versuchen tritt A ein?
- **Lösung**: Es handelt sich hierbei um Ziehen mit Zurücklegen. Sei X die Anzahl der eingetretenen Ereignisse A. Es ergibt sich dann

$$P(X = k) = p^k(1-p)^{n-k}\binom{n}{k}, \quad k = 0, 1, \dots, n.$$

Begründung: Der erste Faktor auf der rechten Seite gibt die Wahrscheinlichkeit dafür an, k-mal das Ereignis A zu erhalten. Wenn insgesamt n Versuche durchgeführt werden und nur k-mal Ereignis A eintritt, so muss in den restlichen $n-k$ Versuchen das Komplementärereignis A^c eintreten. Die Wahrscheinlichkeit dafür ist im zweiten Faktor gegeben. Der Binomialkoeffizient gibt schließlich die Anzahl der Möglichkeiten an, die k eingetretenen Ereignisse A aus n Versuchen auszuwählen.

- **Definition**: Eine Zufallsvariable heißt binomialverteilt mit den Parametern n und p, kurz $X \sim B(n,p)$, wenn sie die Wahrscheinlichkeitsfunktion

$$P(X=k) = B(k|n,p) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0,1,\ldots,n, \\ 0, & \text{sonst} \end{cases}$$

besitzt.

- **Eigenschaften**:

(i) $E(X) = np$

(ii) $Var(X) = np(1-p)$

(iii) Symmetrieeigenschaft:
Für $X \sim B(n,p)$ und $Y = n - X$ gilt $Y \sim B(n, 1-p)$.

(iv) $B(1,p) \sim Ber(p)$.

Beweis:

(i) + (ii)

$$\begin{aligned}
E(X^k) &= \sum_{x:p(x)>0} p(x)x^k \\
&= \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} i^k \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \\
&= np \sum_{i=0}^{n-1} (i+1)^{k-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\
&= npE((Y+1)^{k-1}), \quad Y \sim B(n-1,p) \\
\Longrightarrow E(X) &= npE((Y+1)^0) = np \\
\Longrightarrow E(X^2) &= npE(Y+1) \\
&= np(E(Y)+1) \\
&= np((n-1)p+1) \\
\Longrightarrow Var(X) &= E(X^2) - E(X)^2 \\
&= np((n-1)p+1) - (np)^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}$$

(iii)

$$\begin{aligned}P(Y=k)=P(n-X=k) &= P(X=n-k)\\ &= \binom{n}{n-k}p^{n-k}(1-p)^k = \frac{n!}{k!(n-k)!}(1-p)^k p^{n-k} = B(k|n,1-p)\end{aligned}$$

(iv)

$$\begin{aligned}\text{Für } n=1 \text{ gilt}: \quad P(X=k)=B(k|1,p) &= \binom{1}{k}p^k(1-p)^{1-k} = p^k(1-p)^{1-k}, \quad \text{für} \quad k=0,1\\ &\Rightarrow P(X=1)=p, \quad P(X=0)=1-p\\ &\Rightarrow X\sim \mathrm{Ber}(p)\end{aligned}$$

Die negative Binomialverteilung

- **Ansatz**: Wir führen eine Folge unabhängiger Versuche durch (Bernoulli-Kette) mit Erfolgswahrscheinlichkeit p. Sei X die Anzahl von Versuchen, die nötig sind, bis genau r Erfolge eingetreten sind ($r \in \mathbb{N}$).
- **Idee**: Sei A : „In den ersten $n-1$ Versuchen sind genau $r-1$ Erfolge eingetreten“, und B : „Im n-ten Versuch tritt ein Erfolg ein“. Aufgrund des Vorliegens einer Bernoulli-Kette folgt, dass

  $$P(X=n)=P(A\cap B)=P(A)P(B).$$

  Für die Wahrscheinlichkeiten der beiden Ereignisse ergibt sich

  $$P(A)=B(r-1|n-1,p)=\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$$

  und

  $$P(B)=p.$$

  Daraus folgt

  $$P(X=n)=P(A)P(B)=\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}p=\binom{n-1}{r-1}p^r(1-p)^{n-r},\ n=r,r+1,\dots$$

- X heißt negativ binomialverteilt mit den Parametern r und p, $X\sim NB(r,p)$.
- **Eigenschaften**:
  (i) $E(X)=\frac{r}{p}$
  (ii) $Var(X)=\frac{r(1-p)}{p^2}$
  (iii) Verknüpfung mit Binomialverteilung:
  Sei $X\sim NB(r,p)$ und $Y\sim B(n,p)$. Dann gilt $P(X>n)=P(Y<r)$.
  (iv) Sei $X\sim NB(1,p)$, so gilt $X\sim G(p)$.

  Beweis:

(i) + (ii) Für den Erwartungswert gilt

$$E(X) \quad = \quad \sum_{n=r}^{\infty} n\binom{n-1}{r-1}p^r(1-p)^{n-r}.$$

Wir ersetzen X durch $X_1 + \ldots + X_r$, wobei X_1 die Anzahl der Versuche bis zum ersten Erfolg und X_i die Anzahl der Versuche nach dem $(i-1)$-ten Erfolg bis zum i-ten Erfolg, $i = 2, \ldots, r$ bezeichnet. Die $X_i, i = 1, \ldots, r$ sind paarweise unabhängig und alle geometrisch verteilt mit Parameter p, d.h. $X_i \sim G(p), i = 1, \ldots, r$. Daher ergibt sich für die Erwartungswerte bzw. Varianzen

$$\begin{aligned} E(X_i) &= \frac{1}{p}, \\ Var(X_i) &= \frac{1-p}{p^2}, i = 1, \ldots, r. \end{aligned}$$

Da sich X aus r identisch verteilten Zufallsvariablen zusammensetzt, erhalten wir aufgrund der Linearitätseigenschaft des Erwartungswertes

$$\begin{aligned} E(X) &= E(X_1 + \ldots + X_r) \\ &= E(X_1) + \ldots + E(X_r) \end{aligned}$$

und der äquivalenten Linearitätseigenschaft der Varianz unabhängiger Zufallsvariablen

$$\begin{aligned} E(X) &= r\frac{1}{p}, \\ Var(X) &= r\frac{1-p}{p^2}. \end{aligned}$$

(iii) Betrachte eine Bernoulli-Kette mit Trefferwahrscheinlichkeit p. $\widetilde{X}$ sei der Zeitpunkt des r-ten Treffers. Dann ist $\widetilde{X}$ negativ binomialverteilt mit den Parametern r und p, $\widetilde{X} \sim NB(r,p)$. $\widetilde{Y}$ sei die Anzahl der Treffer in den ersten n Versuchen. Dann ist $\{\widetilde{X} > n\} = \{\widetilde{Y} < r\}$ und deshalb $P(X > n) = P(\widetilde{X} > n) = P(\widetilde{Y} < r) = P(Y < r)$, also

$$\sum_{i=n+1}^{\infty} \binom{i-1}{r-1} p^r (1-p)^{n-r} = \sum_{k=0}^{r-1} \binom{n}{k} p^k (1-p)^{n-k}$$

(iv)

$$NB(k|1,p) = P(X=k) = \binom{k-1}{0} p(1-p)^{k-1} = \frac{(k-1)!}{0!(k-1)!} p(1-p)^{k-1} = (1-p)^{k-1}p = G(k|p)$$

Die hypergeometrische Verteilung

- Ausgangspunkt: Ziehen *ohne* Zurücklegen im Urnenmodell, Reihenfolge der gezogenen Kugeln spielt *keine* Rolle
- Insgesamt seien N Kugeln vorhanden, davon M schwarze und $N - M$ weiße.
- Die Kugeln seien durchnummeriert von 1 bis N, dabei tragen die schwarzen Kugeln die Nummern 1 bis M.
- Es erfolgt eine zufällige Auswahl von n Kugeln ($n \leq N$).
- Für den Träger gilt $\mathcal{T} = \{max(0, n-(N-M)), \ldots, min(n, M)\}$.
- Wenn die Reihenfolge der ausgewählten Kugeln keine Rolle spielt, so umfaßt Ω alle n-elementigen Teilmengen von $\{1, 2, \ldots, N\}$, so daß

$$\Omega = \{\omega = \{i_1, \ldots, i_n\} |\, i_k \in \{1, 2, \ldots, N\}, i_k < i_\ell, \forall\, k \neq \ell, k, \ell \in \{1, \ldots, n\}\}$$

$$\Longrightarrow |\Omega| = \binom{N}{n}$$

- Annahme: Jedes $\omega \in \Omega$ hat die gleiche Wahrscheinlichkeit, gezogen zu werden. Es handelt sich dann um ein Laplace-Experiment.
- Sei A_j : „Unter den n gezogenen Kugeln $\{i_1, \ldots, i_n\}$ befinden sich genau j schwarze Kugeln.“
- Die A_j sind paarweise disjunkt mit
$$\bigcup_{j \in \mathcal{T}} A_j = \Omega$$
$$\Longrightarrow P(A_j) = \frac{|A_j|}{|\Omega|},$$
mit
$$|A_j| = \binom{M}{j}\binom{N-M}{n-j}, \quad j \in \mathcal{T}.$$
Es gibt insgesamt $\binom{M}{j}$ Möglichkeiten zur Auswahl von j schwarzen Kugeln aus den M vorhandenen schwarzen Kugeln. Weiterhin gibt es $\binom{N-M}{n-j}$ Möglichkeiten, die $n-j$ weißen Kugeln aus den $N-M$ vorhandenen weißen Kugeln zu ziehen.
- Gilt für die Wahrscheinlichkeitsfunktion einer Zufallsvariablen X, dass
$$P(X = j) = \frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}}, \quad j \in \mathcal{T},$$
so heißt X hypergeometrisch verteilt mit den Parametern n, M und N, also $X \sim H(n, M, N)$.
- **Eigenschaften**:
  (i) $E(X) = n\frac{M}{N}$
  (ii) $Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$
  (iii) Konvergenz gegen die Binomialverteilung

  Beweis:

  (i) Es gilt
$$E(X) = \frac{\sum_{k=0}^{n} k\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$
Wir ersetzen X durch $X_1 + \ldots + X_M$, wobei
$$X_i = \begin{cases} 1, & \text{falls die } i\text{-te schwarze Kugel gezogen wird,} \\ 0, & \text{sonst,} \end{cases} \quad i = 1, \ldots, M$$
Aufgrund der Linearitätseigenschaft des Erwartungswertes gilt
$$\begin{aligned} E(X) &= E(X_1 + \ldots + X_M) \\ &= E(X_1) + \ldots + E(X_M). \end{aligned}$$
X_i ist nach seiner Definition Bernoulli-verteilt. Es gilt daher
$$E(X_i) = P(X_i = 1).$$
Für die Eintrittswahrscheinlichkeit ergibt sich
$$\begin{aligned} P(X_i = 1) &= \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} \\ &= \frac{n}{N}. \end{aligned}$$

Begründung: Wenn die i-te Kugel ohne Zurücklegen gezogen wird, dann gibt es für die übrigen $n-1$ Kugeln in der Stichprobe vom Umfang n genau $\binom{N-1}{n-1}$ Möglichkeiten, diese aus den übrigen $N-1$ Kugeln auszuwählen (ohne Berücksichtigung der Reihenfolge).

Da dies für jede der insgesamt M schwarzen Kugeln gilt, erhält man für den Erwartungswert von X

$$E(X) = M\frac{n}{N}.$$

(ii) Wir wählen denselben Ansatz wie in (i), indem wir X als Summe von M Zufallsvariablen darstellen. Damit ergibt sich

$$\begin{aligned} Var(X) &= Var(\sum_{i=1}^{M} X_i) \\ &= \sum_{i=1}^{M} Var(X_i) + \sum_{\substack{i=1 \\ i\neq j}}^{M} Cov(X_i, X_j), \end{aligned} \tag{2.10}$$

mit

$$Cov(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j).$$

Da X_i Bernoulli-verteilt ist, erhalten wir

$$Var(X_i) = \frac{n}{N}\left(1 - \frac{n}{N}\right) = \frac{n(N-n)}{N^2}. \tag{2.11}$$

Sei $Z_{ij} := X_iX_j$. Die Zufallsvariable Z_{ij} ist wieder Bernoulli-verteilt, da

$$Z_{ij} = \begin{cases} 1, & \text{falls die } i\text{-te und die } j\text{-te schwarze Kugel gezogen werden,} \\ 0, & \text{sonst,} \end{cases} \qquad i,j = 1,\ldots,M, i \neq j$$

Es gilt

$$\begin{aligned} E(Z_{ij}) &= P(Z_{ij} = 1) = P(\{X_i = 1\} \cap \{X_j = 1\}) = P(X_i = 1, X_j = 1) \\ &= \frac{\binom{2}{2}\binom{N-2}{n-2}}{\binom{N}{n}} \\ &= \frac{n(n-1)}{N(N-1)}. \end{aligned}$$

Damit erhalten wir

$$\begin{aligned} Cov(X_i, X_j) &= \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2} \\ &= \frac{n(n-N)}{N^2(N-1)}. \end{aligned} \tag{2.12}$$

Setzen wir (2.11) und (2.12) in (2.10) ein, so erhalten wir

$$\begin{aligned} Var(X) &= M\frac{n(N-n)}{N^2} + M(M-1)\frac{n(n-N)}{N^2(N-1)} \\ &= \frac{MnN - Mn^2}{N^2} + \frac{(M^2-M)(n^2-Nn)}{N^2(N-1)} \\ &= \frac{nM(N^2 - nN + Mn - MN)}{N^2(N-1)} \\ &= n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}. \end{aligned}$$

Bemerkung: Es sind insgesamt M identische Varianzen und $M(M-1)$ identische Kovarianzen zu berücksichtigen.

(iii) Für Gesamtumfang N und die Anzahl schwarzer Kugeln M gelte bei konstantem Stichprobenumfang n und $N, M \to \infty$, dass $\frac{M}{N} \to p \in [0,1]$. Für $0 \leq j \leq n$ ergibt sich dann

$$\begin{aligned}
P(X=j) &= \frac{\binom{M}{j}\binom{N-M}{n-j}}{\binom{N}{n}} \\
&= \frac{M!(N-M)!n!(N-n)!}{j!(M-j)!(n-j)!(N-M-(n-j))!N!} \\
&= \frac{n!}{j!(n-j)!} \frac{M(M-1)\cdot \ldots \cdot (M-j+1)}{1} \cdot \\
&\quad \cdot \frac{(N-M)(N-M-1)\cdot \ldots \cdot (N-M-(n-j)+1)}{N(N-1)\cdot \ldots \cdot (N-n+1)} \\
&= \frac{n!}{j!(n-j)!} \underbrace{\frac{M(M-1)\cdot \ldots \cdot (M-j+1)}{N(N-1)\cdot \ldots \cdot (N-j+1)}}_{\longrightarrow p^j} \cdot \\
&\quad \cdot \underbrace{\frac{(N-M)(N-M-1)\cdot \ldots \cdot (N-M-(n-j)+1)}{(N-j)(N-j-1)\cdot \ldots \cdot (N-j-(n-j)+1)}}_{\longrightarrow (1-p)^{n-j}} \\
&\to \binom{n}{j} p^j (1-p)^{n-j} = B(j|n,p).
\end{aligned}$$

Die Verteilung konvergiert also gegen die Binomialverteilung zu den Parametern $(n, p = M/N)$. Anschaulich: Bei großem Gesamtumfang N und großer Anzahl schwarzer Kugeln M spielt es asymptotisch keine Rolle, ob man bei der Stichprobenentnahme zurücklegt oder nicht.

Als Faustregel zur Approximation von $H(n, N, M)$ durch $B(n, p)$ fordert man, dass $p = M/N$, $n/N \leq 0.05$.

Die Poisson-Approximation

- Sei X binomialverteilt mit den Parametern n und p, also $X \sim B(n,p)$.
- **Frage**: Wie verhält sich die Verteilung von X für $n \to \infty$?
- **Idee**: Um Trivialität auszuschließen, soll für $n \to \infty$ die Wahrscheinlichkeit gegen null konvergieren, d.h $p_n \to 0$ für $n \to \infty$, so daß

  $$np_n = \lambda_n,\ \lambda_n \to \lambda > 0,\ n \in \mathbb{N}$$

  Damit ergibt sich

  $$\begin{aligned}
  P(X=k) = B(k|n,p_n) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
  &= \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \\
  &\overset{(1)}{=} \frac{n(n-1)\cdot \ldots \cdot (n-k+1)}{n^k} \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\
  &= \underbrace{\frac{n(n-1)\cdot \ldots \cdot (n-k+1)}{n^k}}_{\to 1} \underbrace{\frac{\lambda_n^k}{k!}}_{\to \frac{\lambda^k}{k!}} \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^{n}}_{\to e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^{-k}}_{\to 1} \\
  &\to \frac{\lambda^k}{k!} e^{-\lambda} \quad (n \to \infty),
  \end{aligned}$$

wobei wir in Relation (1) benutzt haben, daß

$$p_n^k = \frac{\lambda_n^k}{n^k}.$$

Damit können wir die Poissonverteilung definieren:

- **Definition**: Eine Zufallsvariable X mit Werten in $\mathbb{N}_0$ und Wahrscheinlichkeitsfunktion

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots$$

heißt *Poisson-verteilt* mit Parameter $\lambda > 0$, oder kürzer $X \sim Po(\lambda)$.

- Anschaulich: Falls in n Versuchen mit Erfolgswahrscheinlichkeit p_n gilt: $np_n \to \lambda$ $(n \to \infty)$, so ist für große n die Anzahl der Erfolge in n Versuchen approximativ Poisson-verteilt zum Parameter λ.

- Faustregel zur Approximation: $X \sim B(n,p) \overset{approx.}{\sim} Po(\lambda)$, falls $\lambda = np$, $n \geq 30$ und $p \leq 0.05$.

- Beispiele für approximativ Poisson-verteilte Zufallsvariablen:

  (i) Anzahl der Druckfehler auf einer Buch- oder Zeitungsseite

  (ii) Anzahl der Einhundertjährigen in einer Bevölkerung

  (iii) Anzahl der falsch gewählten Telefonnummern an einem Tag

  (iv) Anzahl der Kunden am Postschalter in einer Stunde

  (v) Anzahl von Teilchen, die in einer radioaktiven Substanz in einer festen Zeitperiode zerfallen

- Eigenschaften:
  (i) $E(X) = \lambda$
  (ii) $Var(X) = \lambda$

  Beweis:

  (i)

$$\begin{aligned}
E(X) = \sum_{x:p(x)>0} xp(x) &= \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda.
\end{aligned}$$

  (ii) Ansatz: $Var(X) = E(X^2) - E(X)^2$

$$\begin{aligned}
E(X^2) = \sum_{x:p(x)>0} x^2 p(x) &= \sum_{k=0}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=1}^{\infty} (k(k-1)+k) \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\
&= \sum_{k=2}^{\infty} \frac{\lambda^k}{(k-2)!} e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\
&= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda^2 e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} + \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
&= \lambda^2 + \lambda.
\end{aligned}$$

  $\Longrightarrow Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$

2.2.5 Stetige Zufallsvariablen

Stetige Gleichverteilung

Definition 2.37. *Eine stetige Zufallsvariable heißt* gleichverteilt *auf dem Intervall* $[a, b]$, *wenn sie eine Dichte*

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{für}\, a \leq x \leq b, \\ 0, & \text{sonst} \end{cases}$$

besitzt.

Bemerkung: Eine auf $[0, 1]$ gleichverteilte Zufallsvariable nennt man auch *standardgleichverteilt.*
Die Verteilungsfunktion einer gleichverteilten Zufallsvariable lautet

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

Bemerkung: Die Dichte ist an den Stellen a und b unstetig $\Longrightarrow$ Verteilungsfunktion ist dort nicht differenzierbar.

Satz 2.38. *Sei $X \sim U([a, b])$. Der Erwartungswert lautet*

$$E(X) = \frac{a+b}{2},$$

die Varianz lautet

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

Beweis. Für den Erwartungswert erhalten wir

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)dx = \int_a^b \frac{x}{b-a}dx = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b = \frac{1}{b-a}\frac{1}{2}(b^2-a^2) \\ &= \frac{1}{b-a}\frac{1}{2}(b-a)(b+a) = \frac{a+b}{2}. \end{aligned}$$

Für die Berechnung der Varianz verwenden wir den Verschiebungssatz $Var(X) = E(X^2) - E(X)^2$. Dafür müssen wir noch das 2. Moment berechnen

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3-a^3}{3(b-a)}.$$

Damit erhalten wir dann

$$Var(X) = E(X^2) - E(X)^2 = \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}.$$

□

Bemerkung: Die Varianz wächst quadratisch und die Standardabweichung $\sigma = (b-a)/\sqrt{12}$ linear mit der Länge des Intervalls.

Die Gamma-Verteilung

- Sei $\Gamma : (0, \infty) \to (0, \infty)$ die Gammafunktion mit

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt, \quad \alpha > 0. \tag{2.13}$$

- Durch partielle Integration ergibt sich, dass für jedes $\alpha > 0$

$$\begin{aligned} \Gamma(\alpha+1) &= \int_0^{\infty} t^{\alpha}e^{-t}dt = \underbrace{\left[-t^{\alpha}e^{-t}\right]_0^{\infty}}_{=0} + \alpha\int_0^{\infty} t^{\alpha-1}e^{-t}dt \\ &= \alpha\Gamma(\alpha). \end{aligned} \tag{2.14}$$

- Da

$$\Gamma(1) = \int_0^{\infty} e^{-t}dt = \left[-e^{-t}\right]_0^{\infty} = 1,$$

  ergibt sich aus (2.14), dass $\Gamma(\alpha+1) = \alpha!$ für jedes $\alpha \in \mathbb{N}$.

- Weitere wichtige Eigenschaft: $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.
- Es gilt
$$\Gamma(\alpha) = \underbrace{\int_0^\infty t^{\alpha-1}e^{-t}dt}_{>0 \text{ für alle } \alpha>0} \overset{:\Gamma(\alpha)}{\Longrightarrow} 1 = \int_0^\infty \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}dt.$$
Daraus folgt, dass
$$f(t) = \frac{t^{\alpha-1}e^{-t}}{\Gamma(\alpha)}$$
eine Dichtefunktion ist.

 Unter Verwendung der Reparametrisierung $X = \beta t, \beta > 0$ erhalten wir (mit Hilfe der Substitutionsmethode)

 Definition 2.39. *Eine Zufallsvariable X mit der Dichtefunktion*
$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}, \quad 0 < x < \infty, \alpha, \beta > 0$$
heißt gammaverteilt mit den Parametern α und β, oder kurz $X \sim \Gamma(\alpha,\beta)$.

 Satz 2.40. *Es gilt für $X \sim \Gamma(\alpha,\beta)$*

 (i) $E(X) = \alpha\beta$,

 (ii) $Var(X) = \alpha\beta^2$.

- Spezialfälle:

 (i) Sei $\alpha = p/2, p \in \mathbb{N}$ und $\beta = 2$, so erhalten wir
$$f(x|p) = \frac{1}{\Gamma\left(\frac{p}{2}\right)2^{p/2}}x^{p/2-1}e^{-x/2}, \quad 0 < x < \infty. \tag{2.15}$$
Eine Zufallsvariable mit der Dichte (2.15) heißt χ^2-*verteilt* mit p Freiheitsgraden, kurz $X \sim \chi^2_p$.

 (ii) Sei $\alpha = 1$, so erhalten wir
$$f(x|\beta) = \frac{1}{\beta}e^{-x/\beta}, \quad 0 < x < \infty. \tag{2.16}$$
Eine Zufallsvariable mit der Dichte (2.16) heißt *exponentialverteilt* mit (skaliertem) Parameter β. Alternativ:
$$\lambda := \frac{1}{\beta} \Longrightarrow f(x|\lambda) = \lambda e^{-\lambda x}, \quad \lambda > 0.$$
Man schreibt dann auch kurz $X \sim Exp(\lambda)$.

 (iii) Sei $X \sim Exp(\beta)$ und $Y = X^{1/\gamma}$, dann ist
$$f_Y(y|\gamma,\beta) = \frac{\gamma}{\beta}y^{\gamma-1}e^{-y^\gamma/\beta}, \quad 0 < y < \infty, \gamma, \beta > 0. \tag{2.17}$$
Eine Zufallsvariable mit der Dichte (2.17) heißt *Weibull-verteilt* mit den Parametern γ und β.

Zur Exponentialverteilung

- **Definition**: Eine stetige Zufallsvariable X mit nichtnegativen Werten heißt *exponentialverteilt* mit dem Parameter $\lambda > 0$, kurz $X \sim Exp(\lambda)$, wenn sie die Dichte

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{für } x \geq 0, \\ 0, & \text{sonst} \end{cases}$$

  besitzt. Die zugehörige Verteilung heißt *Exponentialverteilung* mit Parameter λ.
- Die Exponentialverteilung ergibt sich als Grenzfall der geometrischen Verteilung.
- Die Verteilungsfunktion ergibt sich durch Integration zu

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{für } x \geq 0, \\ 0, & \text{sonst.} \end{cases}$$

- Eigenschaften:
  (i) $E(X) = \frac{1}{\lambda}$,
  (ii) $Var(X) = \frac{1}{\lambda^2}$,
  (iii) Gedächtnislosigkeit: Für $s, t \geq 0$ gilt

$$P(X > s + t | X > t) = P(X > s).$$

Die Beta-Verteilung

Definition 2.41. *Eine Zufallsvariable X heißt beta(α, β)-verteilt, falls sie die Dichtefunktion*

$$f(x|\alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \alpha, \beta > 0$$

besitzt, wobei $B(\alpha, \beta)$ die Beta-Funktion

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$$

bezeichnet.

- Relation Beta-Funktion und Gamma-Funktion:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}.$$

- Anwendung: Modellierung von Verhältnisdaten (mit Wertebereich zwischen null und eins).
- Eigenschaften
  (i) $E(X) = \frac{\alpha}{\alpha+\beta}$,
  (ii) $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Normalverteilung

- **Definition**: Eine Zufallsvariable X heißt *normalverteilt* mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, kurz $X \sim N(\mu, \sigma^2)$, wenn sie die Dichte

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \ x \in \mathbb{R}$$

  besitzt.

- Bemerkung: Gilt $\mu = 0$ und $\sigma^2 = 1$, so nennt man X *standardnormalverteilt*, kurz $X \sim N(0,1)$ mit der Dichte
$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right).$$

- Die Verteilungsfunktion ist definitionsgemäß gegeben durch
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt,$$
bzw. für die Standardnormalverteilung durch
$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt.$$

- Problem: Das Integral der Verteilungsfunktion läßt sich nicht analytisch berechnen und durch bekannte Funktionen in geschlossener Form schreiben. Daher muß $\Phi(x)$ durch spezielle numerische Verfahren berechnet werden. Die Werte von $\Phi(x)$ liegen tabelliert vor.

- Eigenschaften der Normalverteilung:
  - (i) $E(X) = \mu$,
  - (ii) $Var(X) = \sigma^2$,
  - (iii) Symmetrie zu μ, d.h. es gilt
  $$f(\mu - x) = f(\mu + x),\ x \in \mathbb{R},$$
  bzw. für standardnormalverteilte Zufallsvariablen gilt $\phi(-x) = \phi(x)$, und somit
  $$\Phi(-x) = 1 - \Phi(x).$$
  - (iv) **Standardisierung**: Ist X eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable, so ist die standardisierte Zufallsvariable
  $$Z = \frac{X - \mu}{\sigma}$$
  standardnormalverteilt, d.h. $Z \sim N(0,1)$. Damit ergibt sich
  $$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(z),\ \text{mit } z = \frac{x-\mu}{\sigma}.$$

- **Quantile**: Die Quantile z_p der Standardnormalverteilung sind durch die Gleichung
$$\Phi(z_p) = p,\ 0 < p < 1$$
bestimmt.

- Das p-Quantil z_p teilt die Fläche unter der Dichte $\phi(z)$ in eine Fläche mit Inhalt p links von z_p und eine Fläche mit Inhalt $1 - p$ rechts davon auf.

- Relation zwischen Quantil x_p einer $N(\mu, \sigma^2)$-verteilte Zufallsvariable X und dem der zugehörigen standardisierten Zufallsvariable Z:
$$z_p = \frac{x_p - \mu}{\sigma} \quad \text{bzw.} \quad x_p = \mu + \sigma z_p.$$

- **Zentrale Schwankungsintervalle, $k\sigma$-Bereiche**: Für $X \sim N(\mu, \sigma^2)$ gilt
$$P(\mu - z_{1-\alpha/2}\sigma \leq X \leq \mu + z_{1-\alpha/2}\sigma) = 1 - \alpha.$$

- Sei $k = z_{1-\alpha/2}$, so gilt
$$\begin{aligned} F(\mu + k\sigma) - F(\mu - k\sigma) &= \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\ &= \Phi(k) - \Phi(-k) = 2\Phi(k) - 1. \end{aligned}$$

2.2.6 Exponentialfamilien

- Konzept der Exponentialfamilien geht zurück auf die Untersuchung suffizienter Statistiken (Satz von Pitman-Koopman-Darmois)
- allgemeine Form: Eine Zufallsvariable Y gehört zur Exponentialfamilie von Verteilungen, wenn für ihre Dichte gilt
$$f(y;\theta) = \exp\{t(y)d(\theta) + e(y) + g(\theta)\}, \tag{2.18}$$
mit $\theta \in \Theta \subset \mathbb{R}$ und t, d, e und g sind bekannte und spezifische Funktionen in Abhängigkeit der Verteilung von y
- natürliche Form:
$$f(y;\theta,\tau) = \exp\left\{\frac{y\theta - b(\theta)}{a(\tau)} + c(y,\tau)\right\}, \tag{2.19}$$
wobei a, b, c bekannte Funktionen in Abhängigkeit von der Verteilung von y sind. τ ist ein Streuungsparameter.
- Beispiel: Für $Y \sim N(\mu, \sigma^2)$ ergibt sich die natürliche Form mit $\theta = \mu$, $\tau = \sigma^2, a(\tau) = \tau$
$$b(\theta) = \frac{\theta^2}{2}, \quad c(y,\tau) = -\frac{1}{2}\left(\frac{y^2}{\tau} + \log(2\pi\tau)\right).$$
Lemma 2.42. *Gehört Y zur natürlichen Exponentialfamilie (2.19), so gilt*
$$E(Y) = b'(\theta) \quad \textit{and} \quad Var(Y) = a(\tau)b''(\theta). \tag{2.20}$$

2.3 Multivariate Zufallsvariablen

Definition 2.43. *Sei* $(\Omega, \mathfrak{A}, P)$ *ein beliebiger Wahrscheinlichkeitsraum und sei* $X_1, \ldots, X_n$ *eine beliebige Folge von Zufallsvariablen* $X_i : \Omega \to \mathbb{R},\ i = 1, \ldots, n$.
Die Abbildung $\mathbf{X} = (X_1, \ldots, X_n)$ *von* Ω *nach* $\mathbb{R}^n$ *heißt dann* n-dimensionaler Zufallsvektor *(bzw.* n-variate Zufallsvariable*) mit den Komponenten* $X_1, \ldots, X_n$.
Die Verteilung *des Zufallsvektors* $\mathbf{X}$ *ist die Mengenfunktion* $P_X : \mathfrak{B}(\mathbb{R}^n) \to [0,1]$ *mit*
$$P_X(B) = P(\{\omega : \omega \in \Omega, \mathbf{X}(\omega) \in B\}) \quad \textit{für alle } B \in \mathfrak{B}(\mathbb{R}^n). \tag{2.21}$$
Die Funktion $F_X : \mathbb{R}^n \to [0,1]$ *mit*
$$F_X(\mathbf{x}) = F_X(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n) = P(\mathbf{X} \le \mathbf{x}) \tag{2.22}$$
heißt gemeinsame Verteilungsfunktion *des Zufallsvektors* $\mathbf{X} = (X_1, \ldots, X_n)$.

Im Folgenden beschränken wir uns auf bivariate Zufallsvariablen, d.h $Z = (X, Y) : \Omega \to \mathbb{R}^2$.

2.3.1 Gemeinsame Verteilungen und Randverteilungen

- Gemeinsame Verteilungsfunktion von (X, Y)
$$F_{X,Y}(a,b) = P(X \le a, Y \le b) = P(\{X \le a\} \cap \{Y \le b\}),\ a, b \in \mathbb{R}$$
- Zusammenhang von gemeinsamer Verteilungsfunktion und Verteilungsfunktionen von X und Y
$$F_X(a) = P(X \le a) = P(X \le a, Y < \infty) = \lim_{n\to\infty} F_{X,Y}(a,n)$$
Entsprechend für $F_Y(b) = \lim_{n\to\infty} F_{X,Y}(n,b)$. $F_X(a)$ und $F_Y(b)$ heißen **Randverteilungsfunktionen** von (X, Y).

- Der Zufallsvektor $Z = (X, Y)$ heißt *diskret*, falls es eine abzählbare Menge $C \subset \mathbb{R}^2$ gibt, so dass $P(Z \in C) = 1$.
  - Folgerung: Falls Z ein diskreter Zufallsvektor ist, dann sind auch seine Komponenten diskrete Zufallsvariablen.
  - Sei Z ein diskreter Zufallsvektor. Dann heißt $\{P(Z = z), z \in C\}$ *gemeinsame Wahrscheinlichkeitsfunktion* von Z.
  - Wir verwenden folgende Notation für die gemeinsame Wahrscheinlichkeitsfunktion

$$f_{X,Y}(x, y) = P(X = x, Y = y) = P(\{X = x\} \cap \{Y = y\}),\ x, y \in \mathbb{R}$$

  - Randverteilungen:

$$\begin{aligned} f_X(x) = P(X = x) &= P(\{X = x\} \cap \{Y < \infty\}) \\ &= P(\bigcup_{y: f_{X,Y}(x,y)>0} \{X = x, Y = y\}) \\ &\overset{(K2)}{=} \sum_{y: f_{X,Y}(x,y)>0} f_{X,Y}(x, y) \\ f_Y(y) = P(Y = y) &= \sum_{x: f_{X,Y}(x,y)>0} f_{X,Y}(x, y) \end{aligned}$$

Beispiel 2.12. *Zweimaliges Würfeln. Als Ergebnisraum wählen wir $\Omega = \{\omega = (\omega_1, \omega_2) : \omega_i \in \{1, \ldots, 6\}, i = 1, 2\}$. Sei $X : \Omega \to \{0, 1, 2\}$ bzw. $Y : \Omega \to \{0, 1, 2\}$ die Anzahl, mit der die Augenzahl „6“ bzw. die Augenzahl „1“ beim zweimaligen Würfeln erzielt wird. Dann gilt für die Wahrscheinlichkeiten $P(X = x, Y = y)$ bzw. für die Randwahrscheinlichkeiten $P(X = x)$ und $P(Y = y)$*

	$P(X = x, Y = y)$	y: 0	1	2	$P(X = x)$
	0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$	$\frac{25}{36}$
x	1	$\frac{8}{36}$	$\frac{2}{36}$	0	$\frac{10}{36}$
	2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
	$P(Y = y)$	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	1

Aus der Tabelle kann man auch die Einzelwahrscheinlichkeiten der Summe $X + Y$ erhalten. Beispielsweise gilt

$$P(X + Y = 1) = P(X = 1, Y = 0) + P(X = 0, Y = 1) = \frac{8}{36} + \frac{8}{36} = \frac{16}{36}.$$

- Der Zufallsvektor $Z = (X, Y)$ heißt *stetig*, falls es eine (Lebesgue-integrierbare) Funktion $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ gibt, so dass

$$F_{X,Y}(a, b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f_{X,Y}(x, y)dxdy, \quad \text{für alle } a, b \in \mathbb{R}.$$

Die Funktion $f_{X,Y}$ heißt *gemeinsame Dichte* von (X, Y).

  - Randdichte von X

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

  - Randverteilungsfunktion von X

$$F_X(a) = \int_{-\infty}^{a} f_X(x)dx$$

– Analog für Y

Beispiel 2.13. *Wir betrachten den Zufallsvektor $Z = (X, Y)$ mit der gemeinsamen Dichte*

$$f_{X,Y}(x,y) = \begin{cases} 4xy, & \text{falls } 0 \le x, y \le 1, \\ 0, & \text{sonst.} \end{cases}$$

Für die Randdichten erhalten wir

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = 4x \cdot \frac{1}{2} = 2x \quad \text{für alle } x \in [0,1]$$

und

$$f_Y(y) = \int_0^1 f_{X,Y}(x,y)dx = 2y \quad \text{für alle } y \in [0,1].$$

Wir wollen nun noch die Wahrscheinlichkeit $P(X \le 2Y)$ berechnen. Dafür erhalten wir

$$\begin{aligned} P(X \le 2Y) &= P_{X,Y}((x,y) : 0 \le x, y \le 1, x \le 2y) \\ &= \int_0^{1/2} \int_0^{2y} 4xydxdy + \int_{1/2}^1 \int_0^1 4xydxdy \\ &= \int_0^{1/2} 4y \left[\frac{x^2}{2}\right]_0^{2y} dy + \int_{1/2}^1 4y \left[\frac{x^2}{2}\right]_0^1 dy \\ &= 8\left[\frac{y^4}{4}\right]_0^{1/2} + 2\left[\frac{y^2}{2}\right]_{1/2}^1 = \frac{7}{8}. \end{aligned}$$

- Der Träger $\mathcal{T}$ von $Z = (X, Y)^\top$ ist äquivalent zum univariaten Fall definiert als

$$\mathcal{T} = \{z = (x,y)^\top \in \mathbb{R}^2 : f_{X,Y}(x,y) > 0\}.$$

- bedingte Verteilungen

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ F_{X|Y}(x|y) &= \frac{F_{X,Y}(x,y)}{F_Y(y)} \\ F_{Y|X}(y|x) &= \frac{F_{X,Y}(x,y)}{F_X(x)} \end{aligned}$$

Beispiel 2.14. *Bei dem Beispiel des zweimaligen Würfelns ergeben sich folgende bedingten Wahrscheinlichkeitsfunktion von X unter der Bedingung $\{Y = j\}$*

		i		
	$P(X = i \mid Y = j)$	*0*	*1*	*2*
	0	$\frac{16}{25}$	$\frac{8}{25}$	$\frac{1}{25}$
j	*1*	$\frac{8}{10}$	$\frac{2}{10}$	0
	2	1	0	0

Beispiel 2.15. *Wir betrachten den Zufallsvektor aus Beispiel 2.13. Für $y \in (0,1]$ gilt dann für die bedingte Dichte von X unter der Bedingung $\{Y = y\}$*

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} 2x, & \text{falls } x \in [0,1], \\ 0, & \text{sonst.} \end{cases}$$

Bemerkung: Die bedingte Dichte stimmt in diesem Beispiel mit der Randdichte $f_X(x)$ überein.

- Unabhängigkeit

$$\begin{aligned} F_{X,Y}(a,b) &= F_X(a)F_Y(b) \\ \text{äquivalent}: \ f_{X,Y}(a,b) &= f_X(a)f_Y(b) \end{aligned}$$

- bedingter Erwartungswert: bedingte Erwartung (diskret)

$$E(X|Y=y) = \sum_{x:f_{X,Y}(x,y)>0} x f_{X|Y}(x|y)$$

bedingte Erwartung (stetig)

$$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

- Ebenso kann man die bedingte Varianz $Var(X|Y=x)$ definieren als

$$Var(X|Y=y) = Var(X|y) = E(X^2|y) - (E(X|y))^2.$$

Satz 2.44. *Seien X und Y zwei Zufallsvariablen. Dann gilt*

(i) $E(X) = E(E(X|Y))$ *(iterierter Erwartungswert),*

(ii) $Var(X) = E(Var(X|Y)) + Var(E(X|Y))$.

Beweis. (i) Für $E(X) = \int x f_X(x)dx$ benötigen wir die Randverteilung $f_X(x) = \int f_{X,Y}(x,y)dy$. Damit erhalten wir

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X|Y}(x|y)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} \underbrace{\left[\int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx\right]}_{=E(X|Y=y)} f_Y(y)dy \end{aligned}$$

Damit erhalten wir

$$E(X) = \int_{-\infty}^{\infty} E[X|Y=y]f_Y(y)dy = E[E[X|Y]].$$

(ii) Siehe Casella/Berger, S. 167-168.

□

2.3.2 Hierarchische Modelle und gemischte Verteilungen

Beispiel für hierarchische Modelle:
Die Anzahl an Kunden, die ein Postamt an einem Tag besuchen, ist Poisson-verteilt zum Parameter λ. Jede Person, die die Post betritt sei mit Wahrscheinlichkeit $p\,(0 < p < 1)$ eine Frau und mit Wahrscheinlichkeit $1-p$ ein Mann. Sei X die Anzahl der Frauen und Y die Anzahl der Männer, die das Postamt an einem Tag besuchen. Wie lautet die gemeinsame Wahrscheinlichkeitsverteilung $P(X=x, Y=y)$?
Lösung:
X ist die tägliche Anzahl an Frauen, Y ist die tägliche Anzahl an Männern, die das Postamt betreten. Die Gesamtzahl der Kunden ist damit $X+Y$. Bekannt ist, dass die Gesamtzahl Poisson-verteilt ist mit Parameter λ, also

$$P(X+Y=n) = e^{-\lambda}\frac{\lambda^n}{n!},\ n \geq 0.$$

Gesucht ist die gemeinsame Wahrscheinlichkeitsverteilung $P(X = x, Y = y)$.

Angenommen es kommen n Kunden in das Postamt. Die Wahrscheinlichkeit, dass davon k Frauen sind ergibt sich mit Hilfe der Binomialverteilung

$$P(X = k|\, X + Y = n) = \binom{n}{k} p^k (1-p)^{n-k},\ 0 \leq k \leq n.$$

Wenn von n Kunden k Frauen sind, dann sind $n - k$ Kunden Männer. Daher gilt

$$P(X = k|\, X + Y = n) = P(X = k, Y = n - k|\, X + Y = n).$$

Mit Hilfe dieser bedingten Wahrscheinlichkeit können wir nun die gesuchte gemeinsame Wahrscheinlichkeitsverteilung bestimmen. Es gilt

$$\begin{aligned} P(X = k, Y = n - k|\, X + Y = n) &= \frac{P(\{X = k, Y = n - k\} \cap \{X + Y = n\})}{P(\{X + Y = n\})} \\ &= \frac{P(\{X = k, Y = n - k)}{P(\{X + Y = n\})} \end{aligned}$$

und allgemein für $i, j \in \mathbb{N}_0$:

$$\begin{aligned} P(X = i, Y = j) &= P(X = i, Y = j|\, X + Y = i + j) P(X + Y = i + j) \\ &= \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\ &= \frac{1}{i!j!} e^{-\lambda} \cdot (p\lambda)^i \cdot ((1-p)\lambda)^j \\ &= e^{-p\lambda} \frac{(p\lambda)^i}{i!} e^{-(1-p)\lambda} \frac{((1-p)\lambda)^j}{j!}. \end{aligned}$$

Die gemeinsame Verteilung von X und Y entspricht dem Produkt zweier Poisson-Verteilungen.

2.3.3 Kovarianz und Korrelation

Definition 2.45. *Seien X und Y zwei Zufallsvariablen mit gemeinsamer Dichte $f(x, y)$ und $g : E \to \mathbb{R}$. Dann ist der Erwartungswert von $g(X, Y)$ definiert als*

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy. \tag{2.23}$$

Definition 2.46. *Die Kovarianz zweier Zufallsvariablen X und Y ist definiert als*

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Der Korrelationskoeffizient ist definiert als

$$\rho = \rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Satz 2.47. *Für die Kovarianz zweier Zufallsvariablen X und Y gilt*

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Beweis. Sei $\mu_X = E(X)$ und $\mu_Y = E(Y)$. Dann erhalten wir

$$\begin{aligned} Cov(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\ &= E(XY - X\mu_Y - \mu_X Y + \mu_X \mu_Y) \\ &= E(XY) - E(X)\mu_Y - \mu_X E(Y) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y = E(XY) - \mu_X \mu_Y. \end{aligned}$$

□

Satz 2.48. *Es gelten folgende Aussagen:*

(i) Sind X und Y unabhängig, so gilt $Cov(X,Y) = 0$ und $\rho(X,Y) = 0$.

(ii) Sind X und Y zwei Zufallsvariablen und a und b zwei Konstanten, so gilt

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y).$$

(iii) Für zwei Zufallsvariablen X und Y gilt

(i) $-1 \leq \rho(X,Y) \leq 1$

(ii) $|\rho(X,Y)| = 1$ genau dann, wenn Konstanten $a \neq 0$ und b existieren, so dass $P(Y = aX + b) = 1$. Ist $\rho(X,Y) = 1$, so ist $a > 0$, für $\rho(X,Y) = -1$ gilt $a < 0$.

(iv) Erwartungswert und Varianz von Linearkombinationen: Sei

$$X = a_1 X_1 + \ldots a_n X_n,$$

so gilt

$$E(X) = a_1 E(X_1) + \ldots + a_n E(X_n)$$

und

$$Var(X) \quad = \quad \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{i<j} a_i a_j Cov(X_i, X_j).$$

Beweis. (i) Für unabhängige Zufallsvariablen X und Y gilt

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x,y) dx dy \\
&\overset{X,Y \text{ unabh.}}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= E(X)E(Y)
\end{aligned}$$

und damit $Cov(X,Y) = E(X)E(Y) - E(X)E(Y) = 0$.

(ii) Für den Erwartungswert erhalten wir $E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y$ und damit

$$\begin{aligned}
Var(aX + bY) &= E[((aX + bY) - (a\mu_X + b\mu_Y))^2] \\
&= E[(a(X - \mu_X) + b(Y - \mu_Y))^2] \\
&= E[a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)] \\
&= a^2 E[(X - \mu_X)^2] + b^2 E[(Y - \mu_Y)^2] + 2abE[(X - \mu_X)(Y - \mu_Y)] \\
&= a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y).
\end{aligned}$$

(iii) siehe Casella/Berger, S. 172-173.

(iv) folgt aus der Linearität des Erwartungswertes und per Induktion aus (ii).

□

Beispiel 2.16. *X und Y seien zwei Zufallsvariablen, deren gemeinsame Dichtefunktion folgende Form hat*

$$f_{X,Y}(x,y) = \begin{cases} c \cdot (x + y + xy) & , 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & , sonst \end{cases}$$

(a) *Wie groß muss c sein, damit* $f_{X,Y}(x,y)$ *eine Dichte ist?*

(b) *Wie lauten die Randdichten von* X *und* Y*?*

(c) *Wie lauten die bedingten Dichten?*

(d) *Berechnen Sie die Kovarianz zwischen* X *und* Y.

(e) *Bestimmen Sie die gemeinsame Verteilungsfunktion* $F_{X,Y}(x,y)$.

Aufgabe a)
Den Parameter c erhält man über die Normierungsbedingung

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy$$

durch

$$\begin{aligned}
1 &= \int_0^1\int_0^1 c\cdot(x+y+xy)dxdy \\
&= c\cdot\int_0^1\left[\frac{1}{2}x^2+xy+\frac{1}{2}x^2y\right]_0^1 dy \\
&= c\cdot\int_0^1\left(\frac{1}{2}+y+\frac{1}{2}y\right)dy \\
&= c\cdot\left[\frac{1}{2}y+\frac{1}{2}y^2+\frac{1}{4}y^2\right]_0^1 \\
&= c\cdot\frac{5}{4} \Longrightarrow c=0.8.
\end{aligned}$$

Wegen $c>0$ *und* $0\le x,y\le 1$ *gilt weiterhin* $f_{X,Y}(x,y)\ge 0$.

Aufgabe b)
Für die Randdichte von X *gilt für* $0\le x\le 1$

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\
&= \int_0^1 0.8\cdot(x+y+xy)dy \\
&= 0.8\cdot\left[xy+\frac{1}{2}y^2+\frac{1}{2}y^2x\right]_0^1 \\
&= 0.8\cdot\left(\frac{3}{2}x+\frac{1}{2}\right) = 0.4\cdot(3x+1).
\end{aligned}$$

Analog erhält man für Y *im Bereich* $0\le y\le 1$

$$\begin{aligned}
f_Y(y) &= \int_0^1 0.8\cdot(x+y+xy)dx \\
&= 0.8\cdot\left[\frac{1}{2}x^2+xy+\frac{1}{2}x^2y\right]_0^1 \\
&= 0.4\cdot(3y+1)\,.
\end{aligned}$$

Aufgabe c)
Für die bedingte Dichte von X unter der Bedingung $Y = y$ gilt für $0 \leq x, y \leq 1$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{0.8 \cdot (x + y + xy)}{0.4 \cdot (3y + 1)} = 2\frac{x + y + xy}{3y + 1}.$$

Für die bedingte Dichte von Y unter der Bedingung $X = x$ gilt für $0 \leq x, y \leq 1$

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{0.8 \cdot (x + y + xy)}{0.4 \cdot (3x + 1)} = 2\frac{x + y + xy}{3x + 1}.$$

Aufgabe d)
Für die Kovarianz gilt allgemein

$$Cov(X,Y) = E(XY) - E(X)E(Y).$$

Mit den zuvor errechneten Dichten erhält man

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 0.4x(3x+1)dx = 0.4\left[x^3 + \frac{1}{2}x^2\right]_0^1 = 0.6 \\
E(Y) &= \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 0.4y(3y+1)dy = 0.4\left[y^3 + \frac{1}{2}y^2\right]_0^1 = 0.6 \\
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y) dxdy = \int_0^1\int_0^1 xy \cdot 0.8 \cdot (x + y + xy)dxdy \\
&= 0.8 \cdot \int_0^1\int_0^1 (x^2y + xy^2 + x^2y^2)dxdy \\
&= 0.8 \cdot \int_0^1 \left[\frac{1}{3}x^3y + \frac{1}{2}x^2y^2 + \frac{1}{3}x^3y^2\right]_0^1 dy \\
&= 0.8\left[\frac{1}{6}y^2 + \frac{5}{18}y^3\right]_0^1 = 0.8 \cdot \frac{4}{9} \\
&= \frac{16}{45}
\end{aligned}$$

und daraus

$$Cov(X,Y) = \frac{16}{45} - \left(\frac{3}{5}\right)^2 = -\frac{1}{225}.$$

Aufgabe e)
Es gilt:

$$\begin{aligned}
F_{X,Y}(x,y) = \int_0^x\int_0^y 0.8(u + v + uv)dvdu &= \int_0^x \left[uv + \frac{1}{2}v^2 + \frac{1}{2}v^2u\right]_{v=0}^{v=y} du \\
&= 0.8\int_0^x \left(uy + \frac{1}{2}y^2 + \frac{1}{2}y^2u\right) du \\
&= 0.8\left[\frac{1}{2}yu^2 + \frac{1}{2}y^2u + \frac{1}{4}y^2u^2\right]_{u=0}^{u=x} \\
&= 0.8\left(\frac{1}{2}x^2y + \frac{1}{2}xy^2 + \frac{1}{4}x^2y^2\right) \\
&= 0.4xy\left(x + y + \frac{1}{2}xy\right).
\end{aligned}$$

Damit folgt:

$$F(x,y) = \begin{cases} 0, & \textit{für}\, x,y < 0, \\ 0.4xy(x+y+\frac{1}{2}xy), & \textit{für}\, x,y \in [0,1], \\ F_X(x), & \textit{für}\, x \in [0,1] \wedge y > 1, \\ F_Y(y), & \textit{für}\, y \in [0,1] \wedge x > 1, \\ 1, & \textit{für}\, x,y > 1. \end{cases}$$

2.4 Konvergenzkonzepte und Grenzwertsätze

2.4.1 Konvergenz in Wahrscheinlichkeit

Definition 2.49. *Eine Folge von Zufallsvariablen $X_1, X_2, \ldots$* konvergiert in Wahrscheinlichkeit *gegen eine Zufallsvariable X, falls für jedes $\epsilon > 0$ gilt*

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0 \quad \textit{bzw.} \quad \lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1. \tag{2.24}$$

Als Notation für Konvergenz in Wahrscheinlichkeit verwenden wir $X_n \xrightarrow{P} X$.

Satz 2.50. ***(Ungleichung von Tschebyschev)*** *Für eine Zufallsvariable X mit $E(X) = \mu$ und $Var(X) = \sigma^2$ gelten für beliebiges $c > 0$ folgende Ungleichungen:*

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2} \quad \textit{und} \quad P(|X-\mu| < c) \geq 1 - \frac{\sigma^2}{c^2}$$

Beweis. Wir beweisen die Ungleichung mit Hilfe der Markov-Ungleichung:

Lemma 2.51. ***(Markov-Ungleichung)*** *Sei X eine nicht-negative Zufallsvariable mit $E(X) < \infty$, so gilt für alle $a \in \mathbb{R}_+$*

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Beweis: (für stetige Zufallsvariablen)

$$\begin{aligned} E(X) &= \int_0^\infty x f(x) dx \\ &= \int_0^a x f(x) dx + \int_a^\infty x f(x) dx \\ &\geq \int_a^\infty x f(x) dx \\ &\geq \int_a^\infty a f(x) dx \\ &= a \int_a^\infty f(x) dx \\ &= a P(X \geq a). \end{aligned}$$

Da $(X-\mu)^2$ eine nicht-negative Zufallsvariable ist, können wir die Markov-Ungleichung mit $a = c^2$ direkt anwenden und erhalten

$$P((X-\mu)^2 \geq c^2) \leq \frac{E((X-\mu)^2)}{c^2}. \tag{2.25}$$

Da aber $(X-\mu)^2 \geq c^2 \Leftrightarrow |X-\mu| \geq c$, entspricht Gleichung (2.25)

$$P(|X-\mu| \geq c) \leq \frac{E((X-\mu)^2)}{c^2} = \frac{\sigma^2}{c^2}.$$

□

Satz 2.52. ***(Schwaches Gesetz der großen Zahlen)*** *Sei $X_1, X_2, \ldots$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2 < \infty$. Definiere $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Dann gilt für jedes $\epsilon > 0$*

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1, \tag{2.26}$$

d.h. $\bar{X}_n$ konvergiert in Wahrscheinlichkeit gegen μ.

Satz 2.53. *Eine Folge $X_1, X_2, \ldots$ von Zufallsvariablen konvergiere in Wahrscheinlichkeit gegen X, d.h. $X_n \stackrel{P}{\to} X$. Sei h eine stetige Funktion. Dann gilt $h(X_n) \stackrel{P}{\to} h(X)$.*

2.4.2 Fast sichere Konvergenz

Definition 2.54. *Eine Folge von Zufallsvariablen $X_1, X_n, \ldots$* konvergiert fast sicher *gegen eine Zufallsvariable X, wenn für jedes $\epsilon > 0$ gilt*

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1. \tag{2.27}$$

Als Notation für fast sichere Konvergenz verwenden wir $X_n \stackrel{f.s.}{\to} X$ (bzw. $X_n \stackrel{a.s.}{\to} X$).

Bemerkung: Das Konzept der fast sicheren Konvergenz ist stärker als das Konzept der Konvergenz in Wahrscheinlichkeit. Fast sichere Konvergenz ist vergleichbar mit punktweiser Konvergenz einer Folge von Funktionen. Fast sichere Konvergenz bedeutet, dass für alle Elemente des Trägers Konvergenz herrscht, gegebenenfalls ausgenommen eine Menge $A \in \mathfrak{B}(\mathbb{R})$ mit $P(A) = 0$.

Satz 2.55. ***(Starkes Gesetz der großen Zahlen)*** *Sei $X_1, X_2, \ldots$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2 < \infty$. Definiere $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Dann gilt für jedes $\epsilon > 0$*

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1, \tag{2.28}$$

d.h. $\bar{X}_n$ konvergiert fast sicher gegen μ.

2.4.3 Konvergenz in Verteilung

Definition 2.56. *Eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ konvergiert in Verteilung gegen eine Zufallsvariable X, falls gilt*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x). \tag{2.29}$$

Als Notation für Konvergenz in Verteilung verwenden wir $X_n \stackrel{D}{\to} X$.

Satz 2.57. *Gilt für eine Folge von Zufallsvariablen, dass $X_n \stackrel{P}{\to} X$, dann gilt auch $X_n \stackrel{D}{\to} X$.*

Satz 2.58. ***(Zentraler Grenzwertsatz)*** *Sei $X_1, X_2, \ldots$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen, deren momentenerzeugenden Funktionen (MEFs) in einer Umgebung von null existieren (d.h. $M_{X_i}(t)$ existiert für $|t| < h$, $h > 0$). Sei $E(X_i) = \mu < \infty$ und $Var(X_i) = \sigma^2 \in (0, \infty)$. Definiere $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Sei $F_n(x)$ die Verteilungsfunktion von $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Dann gilt für jedes $x, -\infty < x < \infty$*

$$\lim_{n\to\infty} F_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \tag{2.30}$$

d.h. $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ konvergiert gegen die Verteilungsfunktion der Standardnormalverteilung.

Bemerkung: Die Annahmen über die Existenz der MEFs werden eigentlich nicht benötigt. Daher gilt auch folgende Version des zentralen Grenzwertsatzes:

Satz 2.59. ***(stärkere Form des zentralen Grenzwertsatzes)*** *Sei $X_1, X_2, \ldots$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen. Sei $E(X_i) = \mu < \infty$ und $Var(X_i) = \sigma^2 \in (0, \infty)$. Definiere $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Sei $F_n(x)$ die Verteilungsfunktion von $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Dann gilt für jedes $x, -\infty < x < \infty$*

$$\lim_{n\to\infty} F_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy, \tag{2.31}$$

d.h. $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ konvergiert gegen die Verteilungsfunktion der Standardnormalverteilung.

Satz 2.60. ***(Satz von Slutsky)*** *Gilt $X_n \overset{D}{\to} X$ und $Y_n \overset{D}{\to} a$, mit $a \in \mathbb{R}$, dann gilt*

(i) $X_n + Y_n \overset{D}{\to} X + a$,

(ii) $X_n \cdot Y_n \overset{D}{\to} a \cdot X$.