1. Wahrscheinlichkeit

1.1 Grundbegriffe

> **Definition:** Der *Ereignisraum* $\Omega \neq \emptyset$ ist die Menge aller möglichen Ergebnisse eines betrachteten Zufallsexperiment. Die Elemente $\omega \in \Omega$ heissen *Elementarereignisse* des Experiments.

> **Definition:** Die *Potenzmenge* von Ω, bezeichnet mit $\mathcal{P}(\Omega)$ oder 2^Ω, ist die Menge aller Teilmengen vom Ω. Ein *prinzipielles Ereignis* ist eine Teilmenge $A \subseteq \Omega$, also eine Kollektion von Elementarereignissen. Die Klasse aller (beobachtbaren) *Ereignisse* ist $\mathcal{F}$.

Generell sagen wir nun, dass das Ereignis A *eintritt*, falls das realisierte Elementarereignis ω in A liegt, d.h. $\omega \in A$.

> **Definition:** Ein *Wahrscheinlichkeitsmass* ist eine Abbildung $P : \mathcal{F} \to [0,\, 1]$, welche die nachfolgenden Axiome erfüllt. Für $A \in \mathcal{F}$ nennen wir $P[A] \in [0,\, 1]$ die *Wahrscheinlichkeit* (kurz ***WS***), dass A eintritt. Die geforderten Axiome sind:
>
> - **A0)** $P[A] \geq 0$ für alle Ereignisse $A \in \mathcal{F}$
> - **A1)** $P[\Omega] = 1$
> - **A2)** $P\left[\bigcup_{i=1}^\infty A_i\right] = \sum_{i=1}^\infty P[A_i]$, sofern die $A_i \in \mathcal{F}$ *paarweise disjunkt* sind. Wir schreiben dies auch kürzer als $P\left[\dot{\bigcup}_{i=1}^\infty A_i\right] = \sum_{i=1}^\infty P[A_i]$ die Notation $\dot{\bigcup}$ steht dabei für eine Vereinigung von paarweise disjunkten Mengen.

Aus den obigen Axiomen lassen sich einige grundlegende *Rechenregeln* ableiten:

1. $P[A^c] = 1 - P[A]$
2. $P[\emptyset] = 0$
3. Für $A \subseteq B$ gilt $P[A] \leq P[B]$
4. Für beliebige (nicht unbdeingt disjunkte) $A,\, B$ gilt die allgemeine *Additionsregel*

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

1.2 Diskrete Wahrscheinlichkeit

Ist $\Omega = \{\omega_1,\, \omega_2, \ldots,\, \omega_N\}$ endlich mit $|\Omega| = N$ und $\mathcal{F} = 2^\Omega$, und sind $\omega_1,\, \omega_2, \ldots,\, \omega_N$ alle gleich wahrscheinlich, also $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$, so heisst Ω ein *Laplace-Raum* und P die *diskrete Gleichverteilung* auf Ω. Für beliebige $A \subseteq \Omega$ ist dann

$$P[A] = \frac{\text{Anzahl der Elementarereignisse in } A}{\text{Anzahl der Elementarereignisse in } \Omega} = \frac{|A|}{|\Omega|}.$$

1.3 Bedingte Wahrscheinlichkeit

Definition: Seien A, B Ereignisse und $P[A] > 0$. Die *bedingte Wahrscheinlichkeit* von B unter der Bedingung. dass A eintritt (kurz: gegeben A) wird definiert durch

$$P[B|A] := \frac{P[B \cap A]}{P[A]}.$$

Direkt aus der Definition der bedingten Wahrscheinlichkeit erhält man die sogenannte *Multiplikationsregel:* Für beliebige Ereignisse A, B ist

$$P[A \cap B] = P[B|A] \cdot P[A].$$

Satz 1.1 (Satz der totalen Wahrscheinlichkeit): Sei $A_1, A_2, \ldots, A_n$ eine Zerlegung von Ω (in paarweise disjunkte Ereignisse). Für beliebige Ereignisse B gilt dann $P[B] = \sum_{i=1}^n P[B|A_i] \cdots P[A_i]$.

1.4 Unabhängigkeit von Ereignissen

Definition: Zwei Ereignisse A, B heissen *stochastisch unabhängig*, falls

$$P[A \cap B] = P[A] \cdot P[B].$$

Ist $P[A] = 0$ oder $P[B] = 0$, so sind A und B immer unabhängig. Für $P[A] \neq 0$ gilt

$$A, B \text{ unabhängig } \Leftrightarrow P[B|A] = P[B],$$

und symmetrisch gilt für $P[B] \neq 0$

$$A, B \text{ unabhängig } \Leftrightarrow P[A|B] = P[A].$$

Definition: Die Ereignisse $A_1, A_2, \ldots, A_n$ heissen *stochastisch unabhängig*, wenn für jede endliche Teilfamilie die Produktformel gilt, d.h. für $m \in \mathbb{N}$ und $\{k_1, \ldots, k_m\} \subseteq \{1, \ldots, n\}$ gilt immer

$$P\left[\bigcap_{i=1}^m A_{k_i}\right] = \prod_{i=1}^m P A_{k_i}.$$

2. Diskrete Zufallsvariablen und Verteilungen

2.1 Grundbegriffe

Definition: Eine *diskrete Zufallsvariable* (**ZV**) auf Ω ist eine Funktion $X : \Omega \to \mathbb{R}$. Mit Ω ist natürlich auch der Wertebereich $\mathcal{W}(X) = \{x_1, x_2, \dots\}$ endlich oder abzählbar. Die *Verteilungsfunktion* (***VF***) von X ist die Abbildung $F_X : \mathbb{R} \to [0, 1]$, die definiert ist durch

$$t \to F_X(t) := P[X \leq t] := P[\{\omega : X(\omega) \leq t\}].$$

Die *Gewichtsfunktion* oder *diskrete Dichte* von X ist die Funktion $p_X : \mathcal{W}(X) \to [0, 1]$, die durch

$$p_X(x_k) := P[X = x_k] = P[\{\omega : X(\omega) = x_k\}] \text{ für } k = 1, 2, \dots$$

definiert ist.

2.2 Erwartungswerte

Definition: Sei X eine diskrete Zufallsvariable mit Gewichtsfunktion p_X. Dann definieren wir den *Erwartungswert* von X als

$$E[X] := \sum_{x_k \in \mathcal{W}(X)} x_k p_X(x_k),$$

sofern die Reihe absolut konvergiert. Ansonsten existiert der Erwartungswert nicht.
Zudem gilt:

$$E[X] = \sum_{\omega_i \in \Omega} X(\omega_i) P[\{\omega_i\}] = \sum_{\omega_i \in \Omega} p_i X(\omega_i).$$

Satz 2.1: Sei X eine diskrete Zufallsvariable mit Gewichtsfunktion p_X. und sei $Y = g(X)$ für eine Funktion $g : \mathbb{R} \to \mathbb{R}$. Dann ist

$$E[Y] = E[g(X)] = \sum_{x_k \in \mathcal{W}(X)} g(x_k) p_X(x_k),$$

sofern die Reihe absolut konvergiert.

Satz 2.2: Seien X und Y diskrete Zufallsvariablen, für die jeweils der Erwartungswert existiert. Dann gilt:

1. *Monotonie:* Ist $X \leq Y$, d.h. $X(\omega) \leq Y(\omega)$ für alle ω, so gilt auch $E[X] \leq E[Y]$.
2. *Linearität:* Für beliebige $a, b \in \mathbb{R}$ gilt $E[aX + b] = aE[X] + b$.
3. Falls X nur Werte in $\mathbb{N}_0$ annimmt, so gilt

$$E[X] = \sum_{j=1}^{\infty} P[X \geq j] = \sum_{l=0}^{\infty} P[X > l].$$

Definition: Sei X eine diskrete Zufallsvariable. Ist $E[X^2] < \infty$, so heisst

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

die *Varianz* von X, und $\sqrt{\mathrm{Var}[X]}$ heisst die *Standardabweichung* von X. Manchmal schreibt man auch $\phi(X) = \mathrm{sd}(X) := \sqrt{\mathrm{Var}[X]}$.

Die Varianz beschreibt die durchschnittliche quadratische Abweichung der Zufallsvariablen X von ihrem Erwartungswert $m_X = E[X]$.

Lemma 2.3: Sei X eine diskrete Zufallsvariable mit $E[X^2] < \infty$, und sei $Y = aX + b$. Dann gilt:

1. $\mathrm{Var}[X] = E[X^2] - E^2[X]$
2. $\mathrm{Var}[Y] = \mathrm{Var}[aX + b] = a^2 \mathrm{Var}[X]$

3. Wichtige diskrete Verteilungen

3.1 Diskrete Gleichverteilung

Die *diskrete Gleichverteilung* auf einer endlichen Menge $\mathcal{W} = \{x_1, \ldots, x_N\}$ gehört zu einer Zufallsvariable X mit Wertebereich $\mathcal{W}$ und Gewichtsfunktion

$$p_X(x_k) = P[X = x_k] = \frac{1}{N} \text{ für } k = 1, \ldots, N.$$

Das typische ist hier also, dass alle individuellen Elementarereignisse mit der gleichen Wahrscheinlichkeit auftreten.

3.2 Unabhängige 0-1-Experimente

Um die nächsten vier Verteilungen zu beschreiben, betrachten wir eine Folge gleichartiger Experimente, die alle nur mit Erfolg oder Misserfolg enden können, und betrachten die Ereignisse
$A_i = \{\text{Erfolg beim } i\text{-ten Experiment.}\}$
Wir machen zwei Annahmen:

1. Die A_i sind unabhängig.
2. $P[A_i] = p$ für alle i.

Wir nennen dabei p den *Erfolgsparameter*. Mit $Y_i = I_{A_i}$ bezeichnen wir die *Indikatorfunktion* des Ereignisses A_i, also ist

$$Y_i(\omega) = \begin{cases} 1, & \text{für } \omega \in A_i \\ 0, & \text{für } \omega \notin A_i \end{cases}$$

und nimmt nur die Werte 0 und 1 an, mit $P[Y_i = 1] = P[A_i] = p$.
Wir codieren also die Folge der Ergebnisse als binäre Folge, indem wir 1 für Erfolg und 0 für Misserfolg schreiben. Unter den Annahmen (1) und (2) bilden die Zufallsvariablen Y_i eine *Folge unabhängiger 0-1-Experimente* mit Erfolgsparameter p.

3.3 Bernoulli-Verteilung

Machen wir ein einziges 0-1-Experiment und nennen wir das Ergebnis X, so hat X eine *Bernoulli-Verteilung* mit Parameter p. Es gilt also:

- Wertebereich $\mathcal{W}(X) = \{0, 1\}$
- Gewichtsfunktion $p_X(x) = p^x(1-p)^{1-x}$ für $x \in \{0, 1\} = \mathcal{W}(X)$
  - $p_X(1) = P[X = 1] = p$
  - $p_X(0) = P[X = 0] = 1 - p$

Wir schreiben kurz $X \sim Be(p)$. Anders gesagt ist X die *Anzahl der Erfolge bei einem einzelnen 0-1-Experiment* mit Erfolgsparameter p, das kann also nur die Werte 0 oder 1 liefern.

Des weiteren gilt für eine Bernoulli-Verteilung:

- Erwartungswert $E[X] = p$
- Varianz $\mathrm{Var}[X] = E[X^2] - E^2[X] = p(1-p)$

3.4 Binomialverteilung

Die *Binomialverteilung* mit Parametern n und p beschreibt die *Anzahl der Erfolge* bei n unabhängigen 0-1-Experimenten mit Erfolgsparameter p. In unserem allgemeinen Rahmen hat also die Zufallsvariable
$X = \sum_{i=1}^n I_{A_i} = \sum_{i=1}^n Y_i$
eine solche Verteilung. Es gilt also:

- Wertebereich $\mathcal{W}(X) = \{0, 1, 2, \ldots, n\}$
- Gewichtsfunktion $p_X(k) = P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$ für $k = 0, 1, 2, \ldots, n$

Wir schreiben kurz $X \sim Bin(n, p)$. Im Spezialfall $n = 1$ erhalten wir gerade die Bernoulli-Verteilung, d.h. $Bin(1, p) = Be(p)$.

Aus der Linearität des Erwartungswerts und der Summenformel für Varianz folgt für $X \sim Bin(n, p)$ sofort:

- Erwartungswert $E[X] = \sum_{i=1}^n E[Y_i] = np$
- Varianz $\mathrm{Var}[X] = \sum_{i=1}^n \mathrm{Var}[Y_i] = np(1-p)$

3.5 Geometrische Verteilung

Betrachten wir nun eine unendliche Folge von unabhängigen 0-1-Experimenten mit Erfolgsparameter p. Sei X die Wartezeit auf den ersten Erfolg, also
$X = \inf\{i \in \mathbb{N} : A_i \text{ tritt ein}\} = \inf\{i \in \mathbb{N} : Y_i = 1\}$,
d.h. X ist die Nummer des ersten erfolgreichen Experiments. Dann hat X eine *geometrische Verteilung* mit Parameter p. Es gilt:

- Wertebereich $\mathcal{W}(X) = \mathbb{N}$
- Gewichtsfunktion $p_X(k) = P[X = k] = p(1-p)^{k-1}$ für $k = 1,\ 2,\ 3, \ldots$

Wir schreiben kurz $X \sim Geom(p)$.

Des weiteren gilt:

- Erwartungswert $E[X] = \frac{1}{p}$
- Varianz $\text{Var}[X] = \frac{1-p}{p^2}$

3.6 Negativbinomiale Verteilung

Betrachten wir eine unendliche Folge von unabhängigen 0-1-Experimenten mit Erfolgsparameter p, so können wir für $r \in \mathbb{N}$ auch die Wartezeit auf den r-ten Erfolg betrachten. Dann lässt sich X schreiben als
$X = \inf\left\{k \in \mathbb{N} : \sum_{i=1}^{k} I_{A_i} = r\right\} = \inf\left\{k \in \mathbb{N} : \sum_{i=1}{}' kY_i = r\right\}$.
Dann hat X eine *negativbinomiale Verteilung* mit Parametern r und p und es gilt:

- Wertebereich $\mathcal{W}(X) = \{r,\ r+1,\ r+2, \ldots\}$
- Gewichtsfunktion $p_X(k) = P[X = k] = \binom{k-1}{r-1} p^r (1-p)^{k-r}$ für $k = r,\ r+1,\ r+2, \ldots$

Wir schreiben kurz $X \sim NB(r, p)$. Des weiteren gilt:

- Erwartungswert $E[X] = \sum_{i=1}^{r} R[X_i] = \frac{r}{p}$
- Varianz $\text{Var}[X] = \sum_{i=1}^{r} \text{Var}[X_i] = \frac{r(1-p)}{p^2}$

3.7 Hypergeometrische Verteilung

In einer Urne seien n Gegenstände, davon r vom Typ 1 und $n - r$ vom Typ 2. Man zieht ohne Zurücklegen m der Gegenstände. Die Zufallsvariable X beschreibe nun die Anzahl der Gegenstände vom Typ 1 in dieser Stichprobe vom Umfang m. Dann hat X eine *hypergeometrische Verteilung* mit Parametern $n \in \mathbb{N}$ und $m, r \in \{1,\ 2, \ldots,\ n\}$. Es gilt:

- Wertebereich $\mathcal{W}(X) = \{0,\ 1, \ldots,\ \min(m, r)\}$
- Gewichtsfunktion
  $$p_X(k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}} \text{ für } k \in \mathcal{W}(k).$$

3.8 Poisson-Verteilung

Die *Poisson-Verteilung* mit Parameter $\lambda \in (0, \infty)$ ist eine Verteilung auf der Menge $\mathbb{N}_0$ mit Gewichtsfunktion
$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \text{ für } k = 0, 1, 2, \ldots$
Ist eine Zufallsvariable X Poisson-verteilt mit Parameter λ, so schreiben wir dafür kurz $X \sim \mathcal{P}(\lambda)$.

Sei X_n für jedes n eine Zufallsvariable mit $X_n \sim Bin(n, p_n)$ und $np_n = \lambda$. Lassen wir $n \to \infty$ gehen, so geht $p_n \to 0$, und X_n beschreibt die Anzahl Erfolge bei sehr vielen Versuchen mit sehr kleiner Erfolgswahrscheinlichkeit, also bei seltenen Ereignissen. Es gilt:

- $\lim_{n\to\infty} P[X_n = k] = P[X = k]$

In diesem Sinn ist die Poisson-Verteilung ein Grenzwert von Binomialverteilungen bei geeigneter Skalierung der Parameter. Für eine Poisson-verteilte Zufallsvariable $X \sim \mathcal{P}(\lambda)$ gilt:

- Erwartungswert $E[X] = \lambda$
- Varianz $\mathrm{Var}[X] = \lambda$