3 Einführung in die Wahrscheinlichkeitsrechnung

3.1 Axiomatische Grundlagen

Obschon die Beschäftigung mit Wahrscheinlichkeit und ihren Gesetzmäßigkeiten mehr als 300 Jahre alt ist – das erste wahrscheinlichkeitstheoretische Gesetz, J. Bernoullis Gesetz der großen Zahlen, stammt aus dem Jahre 1705 und wurde posthum 1713 publiziert – war die Frage einer genauen Axiomatisierung der Wahrscheinlichkeitstheorie lange offen. Hilbert stellte bei seinem berühmt gewordenen Vortrag 1900 die Axiomatisierung der Wahrscheinlichkeitstheorie und Physik(!) als das sechste seiner 23 offenen Probleme. Dieses wurde 1933 von A. N. Kolmogorov gelöst. Sein Ansatz baut nicht auf die intuitive Vorstellung von Wahrscheinlichkeiten als Limes relativer Häufigkeiten auf, sondern stützt sich auf das (damals noch recht junge) Gebiet der Maßtheorie. Wir werden diese Axiomatisierung nicht ganz im Detail nachvollziehen (können), weil uns dazu der Begriff des Maßes bzw. des Maßintegrals fehlt. Einige Elemente seiner Theorie aber können wir leicht übertragen. Zunächst bezeichne Ω die Menge aller möglichen Versuchsausgänge eines Zufallsexperiments.

Beispiel 3.1 1. *Besteht das Zufallsexperiment aus dem einmaligen Werfen einer Münze, so ist*

$$\Omega = \{\textit{Kopf, Zahl}\}.$$

2. *Besteht das Experiment aus dem einmaligen Werfen eines Würfels, so ist*

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

In der Potenzmenge

$$\mathcal{P}\Omega := \{A \subseteq \Omega\}$$

wollen wir nun die Mengen auszeichnen, denen wir eine Wahrscheinlichkeit zuweisen wollen. Hierbei sind die folgenden Regeln naheliegend:

- Die Wahrscheinlichkeit von Ω kann man immer messen.
- Kann man die Wahrscheinlichkeit von $A \in \Omega$ messen, so auch die von $A^c := \Omega \backslash A$.
- Kann man die Wahrscheinlichkeit von $A_1, A_2, \ldots$ usw. messen, so auch die von $\bigcup_{i=1}^{\infty} A_i$.

Wir wollen ein System $\mathcal{A}$ von Teilmengen von Ω als *σ-Algebra* bezeichnen, wenn gilt:

- $\Omega \in \mathcal{A}$,
- $A \in \mathcal{A} \;\Rightarrow\; A^c \in \mathcal{A}$,

- $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Beispiel 3.2 *1. Über der Menge*

$$\Omega = \{\textit{Kopf, Zahl}\} \stackrel{\wedge}{=} \{0, 1\}$$

(wenn wir z. B. Kopf=1 und Zahl=0 setzen) gibt es genau zwei σ-Algebren:

$$\begin{aligned} \mathcal{A}_1 &= \{\emptyset, \Omega\} \quad \textit{und} \\ \mathcal{A}_2 &= \mathcal{P}\Omega = \{\emptyset, \{0\}, \{1\}, \Omega\}. \end{aligned}$$

Dies folgt, da jede σ-Algebra die Menge Ω enthalten muss und somit auch $\emptyset$. Weiter impliziert $\{1\} \in \mathcal{A}_2$ sofort auch $\{0\} \in \mathcal{A}_2$.

2. Über jeder beliebigen Menge Ω sind stets

$$\mathcal{A}_1 = \{\emptyset, \Omega\} \quad \textit{und} \quad \mathcal{A}_2 = \mathcal{P}\Omega$$

σ-Algebren. Der Grund dafür, dass man nicht immer $\mathcal{P}\Omega$ als σ-Algebra nimmt und damit eine Wahrscheinlichkeit auf Ω definiert (das wäre an sich sehr praktisch, denn so könnte man sicher sein, jeder Teilmenge von Ω eine Wahrscheinlichkeit zuweisen zu können) ist der, dass sich auf $\mathcal{P}\Omega$ nicht immer eine Wahrscheinlichkeit mit allen gewünschten Eigenschaften definieren lässt (beispielsweise kann man auf

$$\Omega = [0, 1] \quad \textit{und} \quad \mathcal{A} = \mathcal{P}\Omega$$

keine Wahrscheinlichkeit definieren, die einem Intervall seine Länge zuweist). Dies kann und soll uns aber in der Folge nicht weiter interessieren. Für endliche oder abzählbar unendliche Mengen können wir getrost stets die Potenzmenge $\mathcal{P}\Omega$ als σ-Algebra verwenden.

Bei der Definition von Wahrscheinlichkeit spielt schließlich die Intuition eine große Rolle. Ein "üblicher" Begriff von Wahrscheinlichkeit würde die Wahrscheinlichkeit eines Ereignisses A als Limes der relativen Häufigkeiten des Auftretens von A bei n unabhängigen Versuchen definieren, wenn n gegen unendlich geht. Diese Definition birgt einige Nachteile (beispielsweise werden die relativen Häufigkeiten des Auftretens von A typischerweise von den Versuchen abhängen, man kann z. B. ja sowohl K, K, Z, K als auch Z, Z, K, K als Ergebnis eines vierfachen Münzwurfs erhalten); allerdings wäre es sicherlich nützlich, wenn man relative Häufigkeiten als Wahrscheinlichkeiten auffassen könnte. Die folgende Definition ist also gewissermaßen den relativen Häufigkeiten abgeschaut:

Definition 3.3 *Es sei Ω eine Menge, $\Omega \neq \emptyset$, und $\mathcal{A}$ eine σ-Algebra über Ω. Eine Wahrscheinlichkeit $\mathbb{P}$ ist eine Funktion*

$$\mathbb{P} : \mathcal{A} \to [0, 1]$$

mit

- $\mathbb{P}(\Omega) = 1$.
- *Sind* $A_1, A_2, \ldots$ *paarweise disjunkt, d. h.* $A_i \cap A_j = \emptyset$ *für* $i \neq j$, *so gilt:*

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad (\sigma\text{-}\mathit{Additivität})$$

Diese Regeln implizieren insbesondere, dass

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A) \tag{3.1}$$

für alle $A \in \mathcal{A}$ gilt, also auch

$$\mathbb{P}(\emptyset) = 0. \tag{3.2}$$

Der Grund hierfür ist, dass A und A^c disjunkt sind und $A \cup A^c = \Omega$ gilt. Also ist

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c),$$

also $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$; setzt man für $A = \Omega$ ein, so ergibt sich (3.2).

Beispiel 3.4 *Den fairen Münzwurf über* $\Omega = \{0, 1\}$, $\mathcal{A} = \mathcal{P}\Omega$ *modelliert man mit*

$$\mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = \frac{1}{2}.$$

Den fairen Würfelwurf auf $\Omega = \{1, \ldots, 6\}$, $\mathcal{A} = \mathcal{P}\Omega$, *modelliert man mit*

$$\mathbb{P}(\{1\}) = \ldots = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

Im Beispiel 3.4 haben wir die Wahrscheinlichkeit $\mathbb{P}$ gerade dadurch festgelegt, dass wir für jedes $\omega \in \Omega$ angeben, was dessen Wahrscheinlichkeit sein soll. So lange Ω höchstens abzählbar ist, geht dies ganz allgemein.

Satz 3.5 *Ist* Ω *höchstens abzählbar,* $\mathcal{A} = \mathcal{P}\Omega$ *und* $(p_\omega)_{\omega \in \Omega}$ *eine Folge nicht negativer Zahlen mit*

$$\sum_{\omega \in \Omega} p_\omega = 1,$$

dann ist durch

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega \tag{3.3}$$

für $A \subseteq \Omega$ *eindeutig eine Wahrscheinlichkeit auf* $\mathcal{A}$ *festgelegt.*

Beweis: Die Eindeutigkeit ist klar, da man positive Summen beliebig umordnen darf. $\mathbb{P}$, wie in (3.3) definiert, ist auch eine Wahrscheinlichkeit, denn es ist $0 \leq \mathbb{P}(A) \leq 1$ für alle, $A \subset \Omega$,

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p_\omega = 1,$$

und sind $A_1, A_2, \ldots$ paarweise disjunkt, so ist

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{\omega \in \bigcup_{i=1}^{\infty} A_i} p_\omega = \sum_{i=1}^{\infty} \sum_{\omega \in A_i} p_\omega = \sum_{i=1}^{\infty} \mathbb{P}(A_i),$$

da $\omega \in \bigcup_{i=1}^{\infty} A_i$ impliziert, dass $\omega \in A_i$ für genau ein i gilt. $\square$

Wir werden in der Folge immer mit einem sogenannten *Wahrscheinlichkeitsraum* $(\Omega, \mathcal{A}, \mathbb{P})$ arbeiten (Ω eine Menge, $\mathcal{A}$ eine σ-Algebra über Ω und $\mathbb{P}$ eine Wahrscheinlichkeit über $\mathcal{A}$). Dabei nennen wir Ω den Grundraum oder auch das *sichere Ereignis*, $\omega \in \Omega$ heißt *Ergebnis*, $\{\omega\}$ *Elementarereignis*, $A \in \mathcal{A}$ heißt *Ereignis* und $\mathbb{P}$ heißt *Wahrscheinlichkeit.*

3.2 Endliche Wahrscheinlichkeitsräume und mehrstufige Zufallsexperimente

In diesem Kapitel wollen wir uns Zufallsexperimenten zuwenden, bei denen der Grundraum nicht nur abzählbar, sondern sogar endlich ist. Beispiele hierfür haben wir im vorhergehenden Kapitel mit dem fairen Münzwurf und dem fairen Würfeln schon kennengelernt. Diese beiden Beispiele (siehe Beispiel 3.4) haben noch eine weitere Eigenschaft, die im Falle endlich vieler möglicher Versuchsausgänge (d. h. Ω ist eine endliche Menge) oft anzutreffen ist: Alle Elementarereignisse $\omega \in \Omega$ haben dieselbe Wahrscheinlichkeit. Solche Experimente heißen auch Laplace-Experimente.

Definition 3.6 *Sei Ω eine endliche Menge und $\mathcal{A} = \mathcal{P}\Omega$. Eine Wahrscheinlichkeit $\mathbb{P}$ über $\mathcal{A}$ heißt Laplace-Wahrscheinlichkeit (und $(\Omega, \mathcal{A}, \mathbb{P})$ Laplace-Experiment), wenn gilt*

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|} \quad \textit{für alle} \quad \omega \in \Omega.$$

Hierbei bezeichnet $|\Omega|$ die Anzahl der Elemente in Ω. Für ein Laplace-Experiment gilt offenbar

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} \quad \textit{für alle} \quad A \subseteq \Omega.$$

Man nennt die Elemente in A auch die "günstigen Fälle", daher spricht man auch in Laplace-Experimenten davon, dass $\mathbb{P}(A)$ die Anzahl der günstigen Fälle geteilt durch die Anzahl der möglichen Fälle ist.