A Mathematische Grundlagen

In diesem Kapitel sind zur Erinnerung einige Definitionen und Ergebnisse (ohne Beweise) zusammengestellt, die man z. B. in einer Vorlesung über Wahrscheinlichkeitstheorie kennengelernt haben sollte, sowie einige weitere mathematische Kleinigkeiten.

A.1 Allgemeines

A.1 Definition Eine σ-*Algebra* $(\Omega, \mathbb{E})$ über einem *Ergebnisraum* Ω ist eine Menge $\mathbb{E} \subseteq 2^\Omega$ von *Ereignissen* $E \subseteq \Omega$ mit den Eigenschaften:

1. $\emptyset \in \mathbb{E}$
2. $E \in \mathbb{E} \implies \Omega \setminus E \in \mathbb{E}$
3. $(\forall i \in \mathbb{N} : E_i \in \mathbb{E}) \implies \bigcup_{i\in\mathbb{N}} E_i \in \mathbb{E}$

A.2 Wegen des Abschlusses unter Komplement enthält $\mathbb{E}$ stets auch Ω und ist abgeschlossen unter abzählbaren Durchschnitten.

Außerdem ist $\mathbb{E}$ auch abgeschlossen und *endlichen* Vereinigungen (man ergänze sie zu abzählbaren durch Hinzunahme von $\emptyset$) und endlichen Durchschnitten.

Es bezeichne im folgenden $\bar{\mathbb{R}} = \mathbb{R}_{\geqslant 0} \cup \{\infty\}$

A.3 Definition Ist $(\Omega, \mathbb{E})$ eine σ-Algebra, so heißt eine Abbildung $\mu : \mathbb{E} \to \bar{\mathbb{R}}$ ein *Maß auf* $\mathbb{E}$, wenn gelten:

- $\mu(\emptyset) = 0$ und
- „σ-Additivität“: für Folge (E_i) paarweise disjunkter $E_i \in \mathbb{E}$ gilt: $\mu\left(\bigcup_i E_i\right) = \sum_i \mu(E_i)$.

A.4 Definition Ein *Wahrscheinlichkeitsmaß* $\mathbf{P}(\cdot)$ auf einer σ-Algebra $(\Omega, \mathbb{E})$ ist ein Maß $\mathbf{P} : \mathbb{E} \to \bar{\mathbb{R}}$ mit der zusätzlichen Eigenschaft, dass $\mathbf{P}(\Omega) = 1$ ist.

Ein *Wahrscheinlichkeitsraum* $(\Omega, \mathbb{E}, \mathbf{P}(\cdot))$ ist eine σ-Algebra $(\Omega, \mathbb{E})$ mit einem darauf definierten Wahrscheinlichkeitsmaß $\mathbf{P}(\cdot)$. ◇

A.5 Definition Ein Wahrscheinlichkeitsraum heißt *diskret*, falls Ω höchstens abzählbar ist und für alle $\omega \in \Omega$ gilt: $\{\omega\} \in \mathbb{E}$. ◇

A.6 In einem diskreten Wahrscheinlichkeitsraum ist stets $\mathbb{E} = 2^\Omega$.

A.7 Beispiel. Für die Vorlesung sind Wahrscheinlichkeitsräume der folgenden Art sehr wichtig: Es sei R ein randomisierter Algorithmus und x eine Eingabe für R. Dann gibt es im allgemeinen mehrere verschiedene konkret mögliche Berechnungen von R für x. Das können wie zum Beispiel beim randomisierten Quicksort (siehe Kapitel 2) endlich viele sein.

Es können aber selbst für eine einzelne Eingabe auch abzählbar unendlich viele. Als einfaches Beispiel denke man an den (zugegebenermaßen reichlich langweiligen) randomisierten Algorithmus, der für jede natürliche Zahl x als Eingabe so lange Zufallsbits „würfelt", bis die x zuletzt produzierten Bits alle gleich 1 waren, und als Ausgabe z. B. die Gesamtzahl der benötigten Bits liefert.

A.8 Übung. Man gebe für die beiden eben genannten Beispiele Wahrscheinlichkeitsräume an, die jeweils allen möglichen Berechnungen für eine Eingabe x entsprechen.

A.9 Lemma. (Einschluss-Ausschluss-Prinzip) Sind $E_1, \ldots, E_k$ beliebige Ereignisse, dann gilt

$$\mathbf{P}(E_1 \cup E_2) = \mathbf{P}(E_1) + \mathbf{P}(E_2) - \mathbf{P}(E_1 \cap E_2)$$

und allgemeiner

$$\begin{aligned}\mathbf{P}\left(\bigcup_i E_i\right) \;=\; & \sum_i E_i - \sum_{i<j} \mathbf{P}\big(E_i \cap E_j\big) + \sum_{i<j<k} \mathbf{P}\big(E_i \cap E_j \cap E_k\big) - \cdots \\ & + (-1)^{l+1} \sum_{i_1<i_2<\cdots<i_l} \mathbf{P}\left(\bigcap_{r=1}^{l} E_{i_r}\right) + \cdots\end{aligned}$$

A.10 Definition Die *bedingte Wahrscheinlichkeit* von E_1 unter der Bedingung E_2 mit $\mathbf{P}(E_2) > 0$ ist $\mathbf{P}(E_1 \mid E_2) := \mathbf{P}(E_1 \cap E_2)\,/\mathbf{P}(E_2)$. Ist $\mathbf{P}(E_2) = 0$, so sei $\mathbf{P}(E_1 \mid E_2) := 0$. ◇

A.11 Satz. *Ist* $E_1, \ldots, E_k$ *eine Partitionierung von* Ω *und ist* $\mathbf{P}(E) > 0$*, dann gilt:*

$$\mathbf{P}(E) = \sum_{i=1}^{k} \mathbf{P}(E \mid E_i) \cdot \mathbf{P}(E_i)$$

A.12 Korollar. (Formel von Bayes) Ist $E_1, \ldots, E_k$ eine Partitionierung von Ω und ist $\mathbf{P}(E) > 0$, dann gilt:

$$\mathbf{P}(E_i \mid E) = \frac{\mathbf{P}(E_i \cap E)}{\mathbf{P}(E)} = \frac{\mathbf{P}(E \mid E_i)\,\mathbf{P}(E_i)}{\sum_{j=1}^{k} \mathbf{P}\big(E \mid E_j\big)\,\mathbf{P}\big(E_j\big)}\ .$$

A.13 Definition Zwei Ereignisse E_1 und E_2 heißen *(stochastisch) unabhängig*, falls gilt: $\mathbf{P}(E_1 \cap E_2) = \mathbf{P}(E_1) \cdot \mathbf{P}(E_2)$.

Allgemeiner heißt eine Menge $\{E_i \mid i \in I\}$ *unabhängig*, falls für alle $S \subseteq I$ gilt:

$$\mathbf{P}\left(\bigcap_{i \in S} E_i\right) = \prod_{i \in S} \mathbf{P}(E_i)\ .$$

Die Ereignisse heißen k-*unabhängig*, wenn obige Gleichung für alle S einer Größe kleiner gleich k gilt. ◇

A.2 Zufallsvariablen

A.14 Definition Ist $(X, \mathcal{O})$ ein topologischer Raum, dessen offene Mengen gerade die in $\mathcal{O}$ sind, dann ist die *Borelsche* σ*-Algebra* $\mathcal{B}(X)$ die kleinste σ-Algebra, die alle Mengen aus $\mathcal{O}$ enthält. ◇

Ist $X = \mathbb{R}$, so ist im folgenden mit $\mathcal{B} = \mathcal{B}(\mathbb{R})$ stets diejenige σ-Algebra gemeint, die entsteht, wenn man die durch die offenen Intervalle erzeugte Topologie zugrunde legt. Die Mengen in $\mathcal{B}$ werden auch einfach *Borel-Mengen* genannt.

A.15 Definition Es sei $(\Omega, \mathbb{E}, \mathbf{P})$ ein Wahrscheinlichkeitsraum.

Eine *Zufallsvariable* X ist eine Abbildung $X : \Omega \to \mathbb{R}$, so dass für alle Borelmengen $B \subseteq \mathbb{R}$ gilt: $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathbb{E}$. ◇

A.16 Übung. Es sei $(\mathbb{R}, \mathbb{E})$ eine σ-Algebra so, dass $\mathbb{E}$ jedes offene Intervall der Form $(-\infty, x) = \{x' \in \mathbb{R} \mid x' < x\}$ für $x \in \mathbb{R}$ enthält. Zeigen Sie, dass $\mathbb{E}$ dann auch die folgenden Intervalle enthält:

- für jedes $x \in \mathbb{R}$ die Intervalle $(-\infty, x]$, (x, ∞) und $[x, \infty)$,
- für jede $x, y \in \mathbb{R}$ die Intervalle (x, y), $(x, y]$, $[x, y)$ und $[x, y]$;
- für jedes $x \in \mathbb{R}$ das „Intervall“ $[x, x] = \{x\}$.

A.17 Vereinbarung. Wir schreiben statt $\mathbf{P}(\{\omega \in \Omega \mid X(\omega) \leqslant x\})$ kurz $\mathbf{P}(X \leqslant x)$ und analog $\mathbf{P}(X = x)$. Außerdem ist z. B. $\mathbf{P}(X \leqslant x \wedge Y \leqslant y)$ zu verstehen als $\mathbf{P}(\{\omega \in \Omega \mid X(\omega) \leqslant x\} \cap \{\omega \in \Omega \mid Y(\omega) \leqslant y\})$.

Wir gehen im folgenden stillschweigend davon aus, dass $\mathbf{P}(X \leqslant x)$ und $\mathbf{P}(X = x)$ stets existieren, sofern es nicht ohnehin klar ist, etwa wenn der Wahrscheinlichkeitsraum $(\Omega, \mathbb{E})$ diskret ist.

A.18 Beispiel. *Das* Beispiel für Zufallsvariablen in dieser Vorlesung schlechthin ist der Zeitbedarf eines randomisierten Algorithmus für eine konkrete Eingabe.

A.19 Übung. Man präzisiere die eben getroffene Aussage für die Wahrscheinlichkeitsräume aus Beispiel A.7.

Im Fall des randomisierten Quicksort mache man sich klar, dass der Erwartungswert für die Laufzeit nur von der Anzahl der Datenelemente, aber nicht von ihrer ursprünglichen Reihenfolge abhängt. Für den „Bit-Würfel-Algorithmus“ versuche man, den Erwartungswert für die Laufzeit in Abhängigkeit von der Anzahl der zu produzierenden 1-Bits zu bestimmen.

A.20 Definition Eine Zufallsvariable ist *diskret*, falls ihr Wertebereich endlich oder abzählbar unendlich ist.

Die *Indikatorvariable* für ein Ereignis E ist die diskrete Zufallsvariable X mit

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega \in E \\ 0 & \text{falls } \omega \notin E \end{cases}$$

A.21 Definition Die *Verteilungsfunktion* F_X einer Zufallsvariablen X ist die Abbildung

$$F_X : \mathbb{R} \to [0, 1] : x \mapsto \mathbf{P}(X \leqslant x) \ .$$

Die *Dichtefunktion* p_X einer Zufallsvariablen X ist die Abbildung

$$p_X : \mathbb{R} \to [0, 1] : x \mapsto \mathbf{P}(X = x) \ .$$

A.22 Definition Die *gemeinsame Verteilungsfunktion* $F_{X,Y}$ zweier Zufallsvariablen X und Y, die auf dem gleichen Ergebnisraum definiert sind, ist die Abbildung

$$F_{X,Y} : \mathbb{R} \times \mathbb{R} \to [0, 1] : (x, y) \mapsto \mathbf{P}(X \leqslant x \wedge Y \leqslant y) \ .$$

Die *gemeinsame Dichtefunktion* $p_{X,Y}$ von X und Y ist die Abbildung

$$p_{X,Y} : \mathbb{R} \times \mathbb{R} \to [0, 1] : (x, y) \mapsto \mathbf{P}(X = x \wedge Y = y) \ .$$

A.23 Definition Zwei Zufallsvariablen X und Y heißen *unabhängig*, wenn für alle $x, y \in \mathbb{R}$ gilt:

$$\mathbf{P}(X = x \wedge Y = y) = \mathbf{P}(X = x) \cdot \mathbf{P}(Y = y) \ .$$

Allgemeiner heißt eine Menge $\{X_i \mid i \in I\}$ von Zufallsvariablen *unabhängig*, falls für alle $S \subseteq I$ und alle Mengen $\{x_i \in \mathbb{R} \mid i \in I\}$ gilt:

$$\mathbf{P}\left(\bigwedge_{i \in S} X_i = x_i\right) = \prod_{i \in S} \mathbf{P}(X_i = x_i) \ .$$

Die Zufallsvariablen heißen k-*unabhängig*, wenn obige Gleichung für alle S einer Größe kleiner gleich k gilt. ◇

A.24 Lemma. Zwei Zufallsvariablen X und Y sind genau dann unabhängig, wenn für alle $x, y \in \mathbb{R}$ gilt:

$$\mathbf{P}(X = x \mid Y = y) = \mathbf{P}(X = x) \ .$$

A.25 Definition Der *Erwartungswert* $\mathbf{E}[X]$ einer Zufallsvariablen X ist $\mathbf{E}[X] := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$, sofern diese Summe absolut konvergiert. ◇

Absolute Konvergenz bedeutet, dass sogar $\sum_{x \in \mathbb{R}} |x| \cdot p_X(x)$ konvergiert. In diesem Fall ist $\mathbf{E}[X]$ tatsächlich unabhängig von der Reihenfolge der Summanden in $\sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

A.26 Lemma. Für beliebige Zufallsvariablen $X_1, \ldots, X_k$ und beliebige lineare Funktionen h gilt:

$$\mathbf{E}[h(X_1, \ldots, X_k)] = h(\mathbf{E}[X_1], \ldots, \mathbf{E}[X_k]) \ .$$

A.27 Lemma. Für unabhängige Zufallsvariablen X und Y gilt:

$$\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y] \ .$$

A.28 Definition Für $k \in \mathbb{N}$ sind das k-*te Moment* m_X^k und das k-*te zentrale Moment* z_X^k definiert als

$$\begin{aligned} m_X^k &= \mathbf{E}\left[X^k\right] \\ z_X^k &= \mathbf{E}\left[(X - \mathbf{E}[X])^k\right] \ . \end{aligned}$$

Das erste Moment ist der Erwartungswert von X und wird manchmal mit μ bezeichnet. Das zweite zentrale Moment heißt auch *Varianz* und wird mit $\mathbf{var}[X]$ oder σ_X^2 bezeichnet. Die Größe σ_X heißt auch *Standardabweichung*. ◇

A.29 Lemma. $\mathbf{var}[X] = m_X^2 - \mu_X^2 = \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2$.

A.3 Kleinigkeiten

A.3.1 Exponentialfunktion

Die Exponentialfunktion e^x und der natürliche Logarithmus $\ln x$ sowie die Tatsache, dass $d/dx \ln x = 1/x$ werden als bekannt vorausgesetzt.

A.30 Lemma. Für .... ist

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

A.31 Beweis. Die folgende Argumentation findet man auf http://math.stackexchange.com/questions/358830/about-lim-left1-frac-xn-rightn.

Für $t \in]-1/2,\dots,1/2[$ definieren wir Funktionen

$$u(t) = t - \ln(1+t)$$
$$v(t) = t - t^2 - \ln(1+t)$$

Es ist

$$u'(t) = 1 - \frac{1}{1+t} = \frac{t}{1+t}$$
$$v'(t) = 1 - 2t - \frac{1}{1+t} = \frac{(1-2t)(1+t)-1}{1+t} = \frac{-t(2t+1)}{1+t}$$

Für $|t| < 1/2$ ist, wie man sieht, $u'(t) > 0$ und $v'(t) < 0$.

Also ist im betrachteten Intervall $u(t) \geqslant u(-1/2) = -1/2 - \ln(1/2) = -1/2 + \ln 2 > 0$. Und es ist $v(t) \leqslant v(-1/2) = -3/4 - \ln(1/2) = -3/4 + \ln 2 < 0$.

Also gilt für alle $|t| < 1/2$:

$$t - t^2 \leqslant \ln(1+t) \leqslant t\,.$$

Da für jedes $n \in \mathbb{N}_0$ und für $|z| < 1/2$ die Abbildung $z \mapsto e^{nz}$ monoton steigend ist, ergibt sich

$$e^{nt-nt^2} \leqslant (1+t)^n \leqslant e^{nt}\,.$$

Für jedes $x \in \mathbb{R}$ und jedes $n > 2|x|$ gilt für $t = x/n$, dass $|t| < 1/2$. Einsetzen ergibt

$$e^{x-x^2/n} \leqslant \left(1+\frac{x}{n}\right)^n \leqslant e^x\,.$$

Also ist

$$\lim_{n\to\infty} e^{x-x^2/n} \leqslant \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n \leqslant \lim_{n\to\infty} e^x$$

und folglich

$$\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x$$

■

Insbesondere ist also

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e \qquad \text{und} \qquad \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = \frac{1}{e}\,.$$

