

# Kontinuierliche Wahrscheinlichkeitstheorie: Kurze Einführung im Rahmen der Vorlesung „Diskrete Mathematik“ (WS 2009/2010)

Hans Ulrich Simon

6. Oktober 2010

## **Zusammenfassung**

Das hier bereitgestellte Material gehört nicht zum Prüfungsstoff der Vorlesung „Diskrete Mathematik“. Es handelt sich um eine knappe Einführung in die Kontinuierliche Wahrscheinlichkeitstheorie, die mit elementaren mathematischen Grundkenntnissen aus sich selbst heraus verständlich sein sollte. Es dient der Vorbereitung auf Folgevorlesungen, die evtl. Grundkenntnisse in diesem Bereich voraussetzen. In der letzten Vorlesungswoche zur „Diskreten Mathematik“ werden Auszüge aus dem hier bereitgestellten Material behandelt.

# 1 Kontinuierliche Wahrscheinlichkeitstheorie

Wie wir bereits wissen, besitzt ein diskreter Wahrscheinlichkeitsraum eine einfache Definition. Wir betrachten die Elemente  $\omega_1, \omega_2, \dots$  der Grundmenge  $\Omega$  als „Elementarereignisse“, denen wir Zahlen  $0 \leq \Pr(\omega_1), \Pr(\omega_2), \dots$  (die sich zur Gesamtsumme 1 aufaddieren) als Wahrscheinlichkeiten zuordnen. Jede Teilmenge  $E \subseteq \Omega$  repräsentiert dann ein „Ereignis“ der Wahrscheinlichkeit  $\Pr(E) := \sum_{\omega \in E} \Pr(\omega)$ .

Bei kontinuierlichen Wahrscheinlichkeitsräumen wird die Situation komplizierter. Wir betrachten zur Einführung das

**Beispiel 1.1 (Doc Holliday in Aktion)** *Doc Holliday feuert auf eine Zielscheibe. Da er ein guter Schütze ist, geht keiner der Schüsse völlig daneben. Andererseits hat er im Saloon zu viele Drinks zu sich genommen, so dass er die Scheibe jedes Mal ziemlich wahllos trifft. Es liegt nahe, die Zielscheibe als Kreis  $K \subseteq \mathbb{R}^2$  zu modellieren. Unter den getroffenen Annahmen (wahlloses Treffen der Zielscheibe) sollte die Wahrscheinlichkeit, eine Region  $R \subseteq K$  zu treffen, durch den Quotienten „Fläche von  $R$  dividiert durch Fläche von  $K$ “ gegeben sein. Offensichtlich hat dann jeder einzelne Punkt  $x \in K$  (als Objekt der Fläche 0) die Wahrscheinlichkeit 0, getroffen zu werden. Die einer Region  $R$  zugeordnete Wahrscheinlichkeit lässt sich also (im Unterschied zu diskreten Wahrscheinlichkeitsräumen) nicht aus den Wahrscheinlichkeiten der Elemente von  $R$  ableiten. Es gibt aber noch eine weitere Schwierigkeit: es wäre naiv anzunehmen, dass wir jeder Teilmenge  $R$  von  $K$  sinnvoll eine „Fläche“ zuordnen können.  $R$  könnte auf eine so bizarre Weise definiert sein (wozu Mathematiker in der Lage sind!), dass ihre Fläche nicht sinnvoll bemessen werden kann. In einem solchen Fall wäre es überehrgeizig,  $R$  eine Wahrscheinlichkeit zuordnen zu wollen.*

Wir werden den in diesem Beispiel zu Tage tretenden Tücken auf folgende Weise Rechnung tragen:

- Statt mit Elementarereignissen (Elemente der Grundmenge  $\Omega$ ) zu operieren, werden wir Elemente von  $\Omega$  zu größeren Ereignissen zusammenfassen (so dass einerseits interessante Ereignisse erfasst, andererseits aber „nicht messbare“ Mengen vermieden werden). Nur den als Ereignis ausgezeichneten Teilmengen von  $\Omega$  werden wir eine Wahrscheinlichkeit zuordnen.
- Die Teilmengen, welche als Ereignisse ausgezeichnet wurden, sollen eine „algebraische Struktur“ besitzen (mit Namen „ $\sigma$ -Algebra“). Die den Ereignissen zugeordneten Wahrscheinlichkeiten müssen bestimmte Bedingungen erfüllen (um ein sogenanntes „Wahrscheinlichkeitsmaß“ zu repräsentieren). Die Definition von  $\sigma$ -Algebra und Wahrscheinlichkeitsmaß wird sicherstellen, dass wir im kontinuierlichen Fall ähnlich elegant mit Ereignissen und Wahrscheinlichkeiten rechnen können, wie wir das im diskreten Fall gewohnt sind.
- Die Definition der Zufallsvariablen wird dem veränderten Konzept des Wahrscheinlichkeitsraumes Rechnung tragen. Wir werden aber Begriffe wie zum Beispiel Dichte(funktion), Verteilung(sfunktion), Erwartungswert und Varianz in einer ähnlichen

Weise definieren können wie im Falle der diskreten Wahrscheinlichkeitsräume. Der Hauptunterschied wird sein, dass an die Stelle der im diskreten Fall verwendeten „Summen“ oder „Reihen“ im kontinuierlichen Fall „Integrale“ treten.

Das weitere Kapitel ist folgendermaßen aufgebaut. In Abschnitt 1.1 definieren wir die zentralen Begriffe  $\sigma$ -Algebra, Wahrscheinlichkeitsmaß und Wahrscheinlichkeitsraum. Ein besonders wichtiges Beispiel einer  $\sigma$ -Algebra über Grundmenge  $\mathbb{R}^k$  ist durch die sogenannten Borelschen Mengen gegeben. In Abschnitt 1.2 definieren wir numerische Zufallsvariable sowie damit zusammenhängende Begriffe wie Dichte, Verteilung, Erwartungswert und Varianz. In Abschnitt 1.3 diskutieren wir ausgewählte Beispiele kontinuierlicher Verteilungen (uniforme Verteilung, Normalverteilung, Exponentialverteilung). Abschnitt 1.4 kümmert sich um das Zusammenspiel mehrerer (unter Umständen voneinander abhängender) Zufallsvariablen. Zum guten Schluss diskutieren wir in Abschnitt 1.5 den zentralen Grenzwertsatz der die besondere Bedeutung der Normalverteilung unterstreicht.

## 1.1 Kontinuierliche Wahrscheinlichkeitsräume

**Definition 1.2 (sigma-Algebra)** *Zu einer Menge  $\Omega$  bezeichnet  $\mathcal{P}(\Omega)$  die zugehörige Potenzmenge (Menge aller Teilmengen von  $\Omega$ ). Zu einer Teilmenge  $A \subseteq \Omega$  bezeichnet  $\bar{A} := \Omega \setminus A$  ihr Komplement. Ein System  $\mathcal{A} \subseteq \mathcal{P}(\Omega)$  von Teilmengen von  $\Omega$  heißt  $\sigma$ -Algebra (über Grundmenge  $\Omega$ ), wenn folgendes gilt:*

1.  $\Omega \in \mathcal{A}$ , d.h., die Grundmenge gehört zu  $\mathcal{A}$ .
2.  $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$ , d.h.,  $\mathcal{A}$  ist abgeschlossen unter Komplementbildung.
3.  $A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow A_1 \cup A_2 \cup A_3 \dots \in \mathcal{A}$ , d.h.,  $\mathcal{A}$  ist abgeschlossen unter Vereinigung von endlich bzw. abzählbar unendlich vielen Mengen.

Intuitiv können wir die Elemente von  $\mathcal{A}$  als die für unsere Betrachtung zugelassenen „Ereignisse“ anschauen. Die Definition einer  $\sigma$ -Algebra garantiert somit, dass zumindest das (sicher eintretende) „Universalereignis“  $\Omega$  zugelassen ist. Weiterhin ist zu jedem Ereignis  $A$  auch das Komplementärereignis  $\bar{A}$  zugelassen (welches genau dann eintritt, wenn  $A$  nicht eintritt); zu einer Sequenz  $A_1, A_2, \dots$  von zugelassenen Ereignissen ist auch deren Vereinigung als Ereignis zugelassen (welches eintritt, wenn mindestens eines der Ereignisse  $A_1, A_2, \dots$  eintritt).

Im Folgenden meinen wir mit „Vereinigung“ bzw. „Durchschnitt“ immer eine Vereinigung bzw. einen Durchschnitt von endlich oder abzählbar unendlich vielen Mengen (ohne dies stets explizit hervorzuheben).

**Lemma 1.3** *Eine  $\sigma$ -Algebra  $\mathcal{A}$  hat folgende weitere Eigenschaften:*

1.  $\emptyset \in \mathcal{A}$ .
2.  $A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow A_1 \cap A_2 \cap A_3 \dots \in \mathcal{A}$ , d.h.,  $\mathcal{A}$  ist abgeschlossen unter Durchschnitt.

3.  $A, B \in \mathcal{A} \Rightarrow A \setminus B \in \mathcal{A}$ , d.h.,  $\mathcal{A}$  ist abgeschlossen unter Differenzmengenbildung.

### Beweis

1. Wegen  $\Omega \in \mathcal{A}$  gilt auch  $\emptyset = \bar{\Omega} \in \mathcal{A}$ .
2. Mit den Mengen  $A_i$  liegen auch ihre Komplemente  $\bar{A}_i$  und daher auch  $\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3 \cdots$  sowie  $\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3 \cdots$  in  $\mathcal{A}$ . Die Abgeschlossenheit unter Durchschnitt folgt dann unmittelbar aus

$$A_1 \cap A_2 \cap A_3 \cdots = \overline{\overline{A_1 \cap A_2 \cap A_3 \cdots}} = \overline{\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3 \cdots} .$$

3. Mit  $A, B$  liegen auch  $A, \bar{B}$  und  $A \cap \bar{B}$  in  $\mathcal{A}$ . Der Abschluss unter Differenzmengenbildung folgt dann unmittelbar aus  $A \setminus B = A \cap \bar{B}$ .

**qed.**

Intuitiv bedeutet dieses Resultat, dass auch das (niemals eintretende) „leere Ereignis“  $\emptyset$  zugelassen ist. Zu einer Sequenz  $A_1, A_2, \dots$  von zugelassenen Ereignissen ist auch deren Durchschnitt als Ereignis zugelassen (welches eintritt, wenn alle Ereignisse  $A_1, A_2, \dots$  eintreten). Schließlich ist mit zwei zugelassenen Ereignissen  $A, B$  auch das Ereignis  $A \setminus B$  zugelassen (welches eintritt, wenn  $A$  aber nicht  $B$  eintritt).

**Folgerung und Definition 1.4 (Spur einer sigma-Algebra)** Für eine Teilmenge  $T \subseteq \Omega$  und eine  $\sigma$ -Algebra  $\mathcal{A}$  über Grundmenge  $\Omega$  ist das Mengensystem

$$\mathcal{A}_T := \{A \cap T \mid A \in \mathcal{A}\}$$

eine  $\sigma$ -Algebra über Grundmenge  $T$ .  $\mathcal{A}_T$  wird die Spur der  $\sigma$ -Algebra  $\mathcal{A}$  auf der Menge  $T$  genannt.

**Beweis** Wegen  $\Omega \in \mathcal{A}$  gilt  $T = \Omega \cap T \in \mathcal{A}_T$ . Mit  $A \cap T \in \mathcal{A}_T$  für  $A \in \mathcal{A}$  (und somit  $\bar{A} \in \mathcal{A}$ ) liegt auch  $T \setminus (A \cap T) = \bar{A} \cap T$  in  $\mathcal{A}_T$ . Demnach ist  $\mathcal{A}_T$  abgeschlossen unter Komplement. Wegen

$$(A_1 \cap T) \cup (A_2 \cap T) \cup (A_3 \cap T) \cup \dots = (A_1 \cup A_2 \cup A_3 \cup \dots) \cap T$$

und der Abgeschlossenheit von  $\mathcal{A}$  unter Vereinigung ist auch  $\mathcal{A}_T$  abgeschlossen unter Vereinigung. **qed.**

**Beispiel 1.5 (Kleinstmögliche sigma-Algebra)**  $\{\emptyset, \Omega\}$  ist eine  $\sigma$ -Algebra über  $\Omega$ .

**Beispiel 1.6 (Größtmögliche sigma-Algebra)**  $\mathcal{P}(\Omega)$  ist eine  $\sigma$ -Algebra über  $\Omega$ .

Sinnvolle Beispiele liegen zwischen diesen beiden Extremen. Wir beschränken uns im Folgenden auf das besonders wichtige Beispiel der sogenannten Borelschen Mengen.

**Beispiel 1.7 (Borelsche Mengen (eindimensionaler Fall))** Betrachte die Grundmenge  $\Omega = \mathbb{R}$  der reellen Zahlen. Das System der eindimensionalen Borelschen Mengen, notiert als  $\mathcal{B}_1$ , sei die kleinste  $\sigma$ -Algebra, die alle Intervalle der Form  $[a, b]$  mit  $a \leq b \in \mathbb{R}$  enthält. Induktiv ergibt sich  $\mathcal{B}_1$  somit als kleinstes Mengensystem mit folgenden Eigenschaften:

1. Alle abgeschlossenen Intervalle gehören zu  $\mathcal{B}_1$ .
2. Wenn  $A_1, A_2, A_3, \dots$  zu  $\mathcal{B}_1$  gehören, so auch  $A_1 \cup A_2 \cup A_3 \dots$ .
3. Wenn  $A$  zu  $\mathcal{B}_1$  gehört, so auch  $\bar{A} = \mathbb{R} \setminus A$ .

**Beispiel 1.8 (Borelsche Mengen (mehrdimensionaler Fall))** Betrachte die Grundmenge  $\Omega = \mathbb{R}^k$ . Wir bezeichnen das kartesische Produkt von  $k$  abgeschlossenen Intervallen als einen ( $k$ -dimensionalen) achsenparalleler Quader oder kurz als ( $k$ -dimensionale) Box. Das System der  $k$ -dimensionalen Borelschen Mengen, notiert als  $\mathcal{B}_k$ , sei die kleinste  $\sigma$ -Algebra, die alle  $k$ -dimensionalen Boxen enthält. Induktiv ergibt sich  $\mathcal{B}_k$  somit als kleinstes Mengensystem mit folgenden Eigenschaften:

1. Alle  $k$ -dimensionalen Boxen gehören zu  $\mathcal{B}_k$ .
2. Wenn  $A_1, A_2, A_3, \dots$  zu  $\mathcal{B}_k$  gehören, so auch  $A_1 \cup A_2 \cup A_3 \dots$ .
3. Wenn  $A$  zu  $\mathcal{B}_k$  gehört, so auch  $\bar{A} = \mathbb{R}^k \setminus A$ .

Die Definition von Borelschen Mengen ist aus folgenden Gründen klug gewählt:

- Borelsche Mengen decken einen breiten Bereich von Mengen in  $\mathbb{R}^k$  ab. Zum Beispiel ist jede topologisch offene Menge<sup>1</sup> Borelsch (wie man sich anhand der Dichtheit der rationalen Zahlen leicht überlegen kann.<sup>2</sup> Da topologisch abgeschlossene Mengen gerade die Komplemente der topologisch offenen Mengen sind, ist auch jede topologisch abgeschlossene Menge Borelsch.
- Das Volumen Borelscher Mengen ist „vernünftig“ bestimmbar. Man kann nämlich zeigen, dass es genau ein „Maß“ auf  $\mathcal{B}_k$  mit nichtnegativen Werten gibt, welches
  - der leeren Menge die Maßzahl 0 zuordnet,
  - den  $k$ -dimensionalen Boxen ihr Volumen (= Produkt der Kantenlängen), zuordnet,
  - und sich additiv auf disjunkten Vereinigungen Borelscher Mengen verhält.

---

<sup>1</sup>Erinnerung: Eine Teilmenge  $M$  des  $\mathbb{R}^k$  heißt *topologisch offen*, wenn zu jedem Punkt  $x \in M$  eine positive Zahl  $\rho$  existiert, so dass auch die Kugel um  $x$  mit Radius  $\rho$  noch zu  $M$  gehört.

<sup>2</sup>Da wir innerhalb einer offenen Menge jeden Punkt mit einer (hinreichend kleinen) Box umgeben können, ist klar, dass die offene Menge sich als (überabzählbar unendliche) Vereinigung von Boxen schreiben lässt. Um zu einer abzählbar unendlichen Vereinigung zu gelangen, nutzen wir aus, dass  $\mathbb{Q}^k$  dicht in  $\mathbb{R}^k$  liegt.

Dieses Maß wird *Lebesgue-Maß* genannt. In unserer kurzen Übersicht werden wir uns damit aber nicht weiter beschäftigen.

Beachte, dass die Spuren der  $\sigma$ -Algebra der  $k$ -dimensionalen Borelschen Mengen auf Teilmengen von  $\mathbb{R}^k$  weitere Beispiele für  $\sigma$ -Algebren liefern (wie zum Beispiel die Spur von  $\mathcal{B}_2$  auf der Zielscheibe von Doc Holliday).

Wir kommen nun zu der zentralen Definition des laufenden Abschnittes:

**Definition 1.9 (Wahrscheinlichkeitsmaß, Wahrscheinlichkeitsraum)** *Es sei  $\mathcal{A}$  eine  $\sigma$ -Algebra über Grundmenge  $\Omega$ . Eine Abbildung  $\Pr : \mathcal{A} \rightarrow [0, 1]$  heißt Wahrscheinlichkeitsmaß, wenn sie folgende Bedingungen erfüllt:*

1.  $\Pr(\Omega) = 1$ .
2. Für paarweise disjunkte Mengen  $A_1, A_2, A_3, \dots \in \mathcal{A}$  gilt:

$$\Pr(A_1 \cup A_2 \cup A_3 \cup \dots) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) + \dots .$$

Ein Wahrscheinlichkeitsraum ist gegeben durch eine Grundmenge  $\Omega$  versehen mit einer  $\sigma$ -Algebra  $\mathcal{A}$  und einem Wahrscheinlichkeitsmaß  $\Pr : \mathcal{A} \rightarrow [0, 1]$ .

Von einem Wahrscheinlichkeitsmaß wird hier (in dieser von Kolmogorov stammenden Definition) wirklich nur das Nötigste verlangt. Die erste Bedingung normiert die Gesamtwahrscheinlichkeit der „Universalereignisses“ zu 1. Die zweite Bedingung entspricht dem Additionssatz wie wir ihn für diskrete Wahrscheinlichkeitsräume kennengelernt haben.

Die meisten für diskrete Wahrscheinlichkeitsräume bekannten Tatsachen<sup>3</sup> übertragen sich (mit leichtem Beweis) auf kontinuierliche Wahrscheinlichkeitsräume. Wir machen im Folgenden von solchen Übertragungen kommentarlos Gebrauch.

## 1.2 Stetige Zufallsvariable

Bevor wir uns den Zufallsvariablen widmen, stellen wir eine kleine Vorbetrachtung an, die es uns anschließend ermöglicht, das wichtige Konzept der „Dichtefunktion“ einzuführen:

**Satz 1.10** *Eine stetige<sup>4</sup> Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$ , welche die Bedingung*

$$\lim_{x \rightarrow \infty} F(x) = \int_{-\infty}^{+\infty} f(x) dx = 1 \tag{1}$$

*erfüllt, induziert ein Wahrscheinlichkeitsmaß auf  $\mathcal{B}_1$  (den 1-dimensionalen Borelschen Mengen) über die Gleichung*

$$\Pr(A) := \int_A f(x) dx .$$

---

<sup>3</sup>wie zum Beispiel Siebformel, Multiplikationssatz, Satz von der totalen Wahrscheinlichkeit, Satz von Bayes, sowie weitere Regeln für unabhängige oder abhängige Ereignisse und zum Rechnen mit Wahrscheinlichkeiten oder bedingten Wahrscheinlichkeiten

<sup>4</sup>Hier wie im Folgenden würde „stückweise stetig“ anstelle von „stetig“ genügen.

Den einfachen Beweis lassen wir aus.

**Definition 1.11 (Zufallsvariable, Dichte, Verteilung)** Eine (numerische, stetige) Zufallsvariable auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \Pr)$  ist eine Abbildung  $X : \Omega \rightarrow \mathbb{R}$ , die bezüglich einer stetigen Funktion  $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$  die Gleichung<sup>5</sup>

$$\Pr[X \in A] = \int_A f(x) dx \quad (2)$$

für jede eindimensionale Borelsche Menge  $A$  erfüllt. Funktion  $f$  heißt in diesem Zusammenhang die Dichtefunktion (oder kurz Dichte) zur Zufallsvariablen  $X$  und wird auch als  $f_X$  notiert. Die Verteilungsfunktion (oder kurz Verteilung) zu  $X$  ist dann definiert als die Stammfunktion von  $f_X$ :

$$F_X(x) := \Pr[-\infty < X \leq x] = \int_{-\infty}^x f_X(t) dt$$

Wenn eine Zufallsvariable durch ihre Dichte gegeben ist, brauchen wir offensichtlich über den Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \Pr)$  nichts zu wissen, da die Berechnung von  $\Pr[X \in A]$  und davon abgeleiteten Größen über die Gleichung (2) erfolgen kann. Dies unterstreicht erneut die wichtige Rolle der Borelschen Mengen. Wir werden später alle (für uns) wichtigen stetigen Zufallsvariablen über ihre Dichte einführen.

Da  $F_X$  Stammfunktion von  $f_X$  ist, gilt natürlich

$$F'_X \equiv f_X \text{ ,}$$

d.h., die Dichte ergibt sich als erste Ableitung der Verteilung. Weiterhin ergibt sich die Wahrscheinlichkeit, dass  $X$  sich in einem Intervall von  $a$  bis  $b$  aufhält, mit Hilfe der Verteilung gemäß

$$\Pr[a \leq X \leq b] = F_X(b) - F_X(a) \text{ .} \quad (3)$$

Bei der Diskussion der Momente einer Zufallsvariablen konzentrieren wir uns (wie schon im Falle von diskreten Zufallsvariablen) auf die Momente der Ordnung 1 und 2:

**Definition 1.12 (Erwartungswert, Varianz)** Der Erwartungswert und die Varianz einer stetigen, numerischen Zufallsvariable  $X$  mit Dichte  $f_X$  sind (sofern die folgenden Integrale existieren) gegeben durch

$$\begin{aligned} \mathbb{E}[X] &:= \int_{-\infty}^{+\infty} t f_X(t) dt \text{ ,} \\ \text{Var}[X] &:= \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{+\infty} (t - \mathbb{E}[X])^2 f_X(t) dt \text{ .} \end{aligned}$$

---

<sup>5</sup>Wir werden nur Fälle betrachten, wo das in (2) vorkommende Integral mit dem aus einer Mathematik-Grundvorlesung her bekannten (sogenannten Riemannschen) Integral übereinstimmt.

Freilich kann die Varianz auch (wie gewohnt) gemäß

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{+\infty} t^2 f_X(t) dt \\ \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2\end{aligned}$$

berechnet werden.

Die Rechenregeln für Erwartungswert und Varianz, die wir für diskrete Zufallsvariable kennengelernt haben, gelten sinngemäß auch für stetige Zufallsvariable. Davon machen wir im Folgenden (zumeist kommentarlos) Gebrauch.

Bevor wir zu konkreten Beispielen für stetige Zufallsvariable kommen, möchten wir eine vorbereitende Überlegung anstellen, die uns hilft, überflüssige Rechnungen zu vermeiden. Es sei  $X$  eine Zufallsvariable mit Dichte  $f$  und  $Y = cX + d$  für Konstanten  $c, d \in \mathbb{R}$  mit  $c > 0$ , d.h.,  $Y$  ist eine „affine Transformation“ von  $X$ . Dann können wir die Berechnung von  $\Pr[a \leq Y \leq b]$  auf die Berechnung von

$$\Pr[a \leq X \leq b] = \int_a^b f(t) dt$$

zurückführen wie folgt:

$$\Pr[a \leq Y \leq b] = \Pr[a \leq cX + d \leq b] = \Pr\left(\frac{a-d}{c} \leq X \leq \frac{b-d}{c}\right).$$

Dies gilt natürlich insbesondere für  $a = -\infty$  und  $b = x$  und somit für die Verteilung:

$$F_Y(x) = \Pr[Y \leq x] = \Pr\left(X \leq \frac{x-d}{c}\right) = F_X\left(\frac{x-d}{c}\right).$$

Da die Dichte sich als Ableitung der Verteilung ergibt, gilt weiterhin

$$f_Y(x) = F'_Y(x) = \frac{1}{c} F'_X\left(\frac{x-d}{c}\right) = \frac{1}{c} f_X\left(\frac{x-d}{c}\right).$$

Desweiteren gelten natürlich die üblichen Regeln zum Rechnen mit Erwartungswert und Varianz. Wir fassen diese Diskussion zusammen in dem

**Lemma 1.13** *Für Zufallsvariable  $X, Y$  mit  $Y = cX + d$ ,  $c, d \in \mathbb{R}$  mit  $c \geq 0$ , gilt:*

$$\begin{aligned}\Pr[a \leq Y \leq b] &= \Pr\left(\frac{a-d}{c} \leq X \leq \frac{b-d}{c}\right) \\ F_Y(x) &= F_X\left(\frac{x-d}{c}\right) \\ f_Y(x) &= \frac{1}{c} f_X\left(\frac{x-d}{c}\right) \\ \mathbb{E}[Y] &= c\mathbb{E}[X] + d \\ \text{Var}[Y] &= c^2 \text{Var}[X]\end{aligned}$$



Für negative Konstanten gilt eine analoge Regel, deren Formulierung und Herleitung wir dem Leser und der Leserin überlassen.

Als Aufwärmübung zu den vielen technischen Definitionen des laufenden Abschnittes diskutieren wir im Folgenden die einfachste Verteilung einer stetigen Zufallsvariable:

**Beispiel 1.14 (uniforme Verteilung auf einem Einheitsintervall)** Eine uniform auf  $[0, 1]$  verteilte Zufallsvariable  $U \equiv U_{0,1}$  hat den Wertebereich  $W_U = [0, 1]$  und die Dichte

$$f_U(x) = \begin{cases} 1 & \text{falls } x \in [0, 1] \\ 0 & \text{sonst} \end{cases} .$$

Für alle  $0 \leq a \leq b \leq 1$  gilt dann<sup>6</sup>

$$\Pr[a \leq U \leq b] = \int_a^b f(x) dx = \int_a^b 1 dx = [x]_a^b = b - a .$$

Für die Verteilung (der Fall  $a = 0, b = 1$ ) erhalten wir

$$F_U(x) = \begin{cases} 0 & \text{falls } x < 0 \\ x & \text{falls } x \in [0, 1] \\ 1 & \text{falls } x > 1 \end{cases} .$$

Insbesondere ergibt sich wegen  $\lim_{x \rightarrow \infty} F_U(x) = F(1) = 1$ , dass  $f_U$  Bedingung (1) erfüllt (was eine „ordentliche“ Dichte immer tun muss!). Der Erwartungswert errechnet sich (mit wenig überraschendem Ergebnis) wie folgt:

$$\mathbb{E}[U] = \int_0^1 t dt = \frac{1}{2}[t^2]_0^1 = \frac{1}{2}(1^2 - 0^2) = \frac{1}{2} .$$

Eine analoge Rechnung ergibt

$$\mathbb{E}[U^2] = \int_0^1 t^2 dt = \frac{1}{3}[t^3]_0^1 = \frac{1}{3}(1^3 - 0^3) = \frac{1}{3} .$$

Die Varianz ergibt sich dann aus

$$\text{Var}[U] = \mathbb{E}[U^2] - \mathbb{E}[U]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} .$$

**Beispiel 1.15 (uniforme Verteilung auf einem beliebigen Intervall)** Zu einer uniform auf  $[0, 1]$  verteilten Zufallsvariable  $U$  betrachten wir die (uniform auf  $[a, b]$  verteilte) Zufallsvariable  $U' = a + (b - a)U$ . Weiter gelte  $a \leq a' \leq b' \leq b$ , so dass  $[a', b']$  ein Teilintervall von

---

<sup>6</sup>Hier wie im Folgenden machen wir Gebrauch von der bekannten Tatsache, dass  $\int_{-\infty}^{+\infty} f(x) dx = \int_A f(x) dx$ , wenn  $f$  außerhalb der Menge  $A$  den Wert 0 annimmt.

$[a, b]$  ist. Mit Hilfe von Lemma 1.13 ergibt sich (mit  $c := b - a$  und  $d := a$ ) sofort:

$$\begin{aligned}
 f_{U'}(x) &= \begin{cases} \frac{1}{b-a} & \text{falls } x \in [a, b] \\ 0 & \text{sonst} \end{cases} \\
 F_{U'}(x) &= \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } x \in [a, b] \\ 1 & \text{falls } x > b \end{cases} \\
 \Pr[a' \leq U' \leq b'] &= \Pr\left(\frac{a' - a}{b - a} \leq U \leq \frac{b' - a}{b - a}\right) \\
 &= \frac{b' - a}{b - a} - \frac{a' - a}{b - a} = \frac{b' - a'}{b - a} \\
 \mathbb{E}[U'] &= a + (b - a)\mathbb{E}[U] \\
 &= a + \frac{b - a}{2} = \frac{a + b}{2} \\
 \text{Var}[U'] &= (b - a)^2 \text{Var}[U] = \frac{(b - a)^2}{12}
 \end{aligned}$$

Zum Abschluss des laufenden Abschnittes diskutieren wir das sogenannte „Bertrand’sche Paradox“. Es ist aber kein eigentliches logisches Paradox, sondern zeigt lediglich, dass man beim Spezifizieren einer „uniformen Verteilung“ penibel sein muss: wenn sich der Begriff der uniformen Verteilung auf verschiedene Zufallsvariable bezieht, kann man beim gleichen Problem zu verschiedenen Lösungen gelangen.

**Beispiel 1.16 (Bertrand’sches Paradox)** Gegeben ein Kreis mit Radius  $r$  (und folglich Fläche  $\pi r^2$ ) und ein diesem Kreis einbeschriebenes gleichseitiges Dreieck. Diese Situation ist (mit ein paar elementar geometrisch ableitbaren Zusatzdaten) in Abbildung 1 illustriert. Wir stellen nun folgende Frage: mit welcher Wahrscheinlichkeit ist eine „zufällige“ Kreissehne länger als die Dreiecksseite (Ereignis  $E$ )?

Um zu einer ersten Antwort zu gelangen, nehmen wir an, dass der Abstand  $R$  der Sehne zum Kreismittelpunkt zufällig gemäß der uniformen Verteilung aus dem Intervall  $[0, r]$  ausgewählt wird. Wie aus Abbildung 1 implizit hervorgeht, ist  $E$  äquivalent zu  $R < r/2$  und hat daher Wahrscheinlichkeit  $1/2$ .

Wir können die Kreissehne aber auch erzeugen, indem wir eine zufällige Inkreisfläche  $F$  gemäß der uniformen Verteilung aus dem Intervall  $[0, \pi r^2]$  auswählen und die Sehne tangential an den Inkreis setzen. Wie aus Abbildung 1 implizit hervorgeht, ist  $E$  äquivalent zu  $F < \pi r^2/4$  und hat daher Wahrscheinlichkeit  $1/4$ .

Schließlich könnten wir auch am Kreismittelpunkt einen zufälligen Winkel von  $\varphi$  Grad ansetzen, um die Kreissehne als Verbindungslinie der Schnittpunkte von Kreislinie mit den Schenkeln des Winkels zu erhalten. Hierbei wird  $\varphi$  uniform aus dem Intervall von 0 bis 360 Grad gewählt. Wie aus Abbildung 1 implizit hervorgeht, ist  $E$  äquivalent zu  $\varphi \in (120, 240)$  und hat daher Wahrscheinlichkeit  $1/3$ .

Je nach Modellierung haben wir drei verschiedene Antworten auf die eingangs gestellte Frage erhalten.

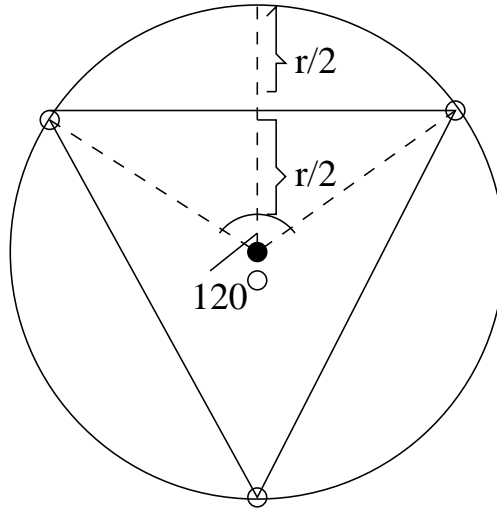


Abbildung 1: Illustration des Bertrand'schen Paradox'.

### 1.3 Wichtige stetige Verteilungen

Die erste wichtige Beispielklasse stellen die uniform auf einem Intervall  $[a, b]$  verteilten Zufallsvariablen dar, die wir bereits in Beispiel 1.15 kennengelernt haben. In diesem Abschnitt besprechen wir darüber hinaus die Normalverteilung und die exponentielle Verteilung.

Beim Hantieren mit konkreten Verteilungen sind wir öfter mal genötigt, Integrale zu berechnen. Zur Vorbereitung erinnern wir an die partielle Integration und an die Substitutionsregel. Partielle Integration erfolgt nach dem Schema

$$\int u'v = uv - \int uv' .$$

Substitution  $x = g(t)$  für eine umkehrbare Funktion  $g$  ermöglicht im Integranden einen Wechsel von Variable  $x$  zur Variable  $t$  gemäß

$$\int f(x) dx = \int f(g(t)) \frac{dx}{dt} dt = \int f(g(t)) g'(t) dt$$

bzw.

$$\int_a^b f(x) dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(t)) g'(t) dt = - \int_{g^{-1}(b)}^{g^{-1}(a)} f(g(t)) g'(t) dt .$$

Die zweite Gleichung (die nützlich ist im Falle  $g^{-1}(b) < g^{-1}(a)$ , d.h., im Falle dass  $g$  streng monoton fällt) ergibt sich aus der allgemeinen Regel

$$\int_{\alpha}^{\beta} h(t) dt = - \int_{\beta}^{\alpha} h(t) dt .$$

### 1.3.1 Normalverteilung

Eine Zufallsvariable  $X$  mit Wertebereich  $\mathbb{R}$ , Dichte

$$f_X(x) := \varphi(x; \mu, \sigma) := \frac{1}{\sqrt{2\pi} \sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

und Verteilung

$$F_X(x) := \Phi(x; \mu, \sigma) := \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^x e^{-(t-\mu)^2/(2\sigma^2)} dt$$

heißt *normalverteilt mit Parametern  $\mu, \sigma$* .

**Notation**  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

Parameter  $\mu$  wird sich später als Erwartungswert und  $\sigma^2$  wird sich als Varianz von  $X \sim \mathcal{N}(\mu, \sigma^2)$  erweisen.

Im Falle von  $\mu = 0$  und  $\sigma = 1$  heißt  $X$  *standard-normalverteilt*.  $X$  hat dann die Dichte

$$f_X(x) := \varphi(x) := \varphi(x; 0, 1) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

und die Verteilung

$$F_X(x) := \Phi(x) := \Phi(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt . \quad (4)$$

Die Dichte der Normalverteilung hat die (vermutlich den meisten bekannte) Form einer „Glockenkurve“. Zur Illustration s. Abbildung 2.4 im Lehrbuch.

Wir zitieren das folgende nützliche Resultat ohne Beweis:

**Lemma 1.17**  $I := \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ .

Es folgt sofort

$$\int_{-\infty}^{+\infty} \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \cdot I = 1 ,$$

d.h.,  $\varphi(x)$  ist eine „ordentliche“ Dichte.

Mit Hilfe von Lemma 1.17 können wir leicht Erwartungswert und Varianz von standard-normalverteilten Zufallsvariablen berechnen:

**Lemma 1.18** Für  $X \sim \mathcal{N}(0, 1)$  gilt  $\mathbb{E}[X] = 0$  und  $\text{Var}[X] = 1$ .

**Beweis** Offensichtlich gilt  $\varphi(-t) = \varphi(t)$ . Aus dieser Symmetrie läßt sich leicht  $\mathbb{E}[X] = 0$  folgern. Zur Berechnung der Varianz verwenden wir Lemma 1.17 sowie partielle Integration und erhalten

$$\sqrt{2\pi} = \int_{-\infty}^{+\infty} \underbrace{1}_{u'} \cdot \underbrace{e^{-x^2/2}}_v dx = [x e^{-x^2/2}]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx = 0 - 0 + \sqrt{2\pi} \mathbb{E}[X^2] .$$

Somit gilt  $\mathbb{E}[X^2] = 1$  und  $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - 0^2 = 1$ .

**qed.**

Wir zeigen nun, dass eine beliebig normalverteilte Variable sich als affine Transformation einer standard-normalverteilten Variablen darstellen lässt.

**Lemma 1.19** Falls  $X \sim \mathcal{N}(0, 1)$ , dann folgt  $Y \sim \mathcal{N}(\mu, \sigma^2)$  für  $Y := \sigma \cdot X + \mu$ .

**Beweis** Der Beweis ist eine einfache Anwendung von Lemma 1.13 (mit  $c := \sigma$  und  $d := \mu$ ):

$$f_Y(x) = \frac{1}{\sigma} f_X\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi} \sigma} e^{-(x-\mu)^2/(2\sigma^2)} = \varphi(x; \mu, \sigma)$$

qed.

**Folgerung 1.20** Falls  $Y \sim \mathcal{N}(\mu, \sigma^2)$ , dann folgt  $Z \sim \mathcal{N}(c\mu + d, c^2\sigma^2)$  für  $Z := cY + d$ .

**Beweis** Die Aussage folgt sofort aus Lemma 1.19 wegen

$$Z = cY + d = c(\sigma X + \mu) + d = c\sigma X + c\mu + d .$$

qed.

Es sei  $X \sim \mathcal{N}(0, 1)$  und somit  $\mathbb{E}[X] = 0$  und  $\text{Var}(X) = 1$ . In Verbindung mit Lemma 1.13 lassen sich nun Erwartungswert und Varianz von  $Y = \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$  leicht bestimmen:

$$\begin{aligned} \mathbb{E}[Y] &= \sigma \mathbb{E}[X] + \mu = \mu \\ \text{Var}[Y] &= \sigma^2 \text{Var}[X] = \sigma^2 \end{aligned}$$

Wie oben angekündigt repräsentieren die Parameter  $\mu$  und  $\sigma^2$  also gerade den Erwartungswert und die Varianz der Normalverteilung.

Schließlich sind wir noch an der Wahrscheinlichkeit interessiert, dass eine normalverteilte Zufallsvariable sich in einem bestimmten Intervall (in Anwendungen meist ein um den Erwartungswert zentriertes Intervall) aufhält. Aus der allgemeinen Gleichung (3) folgt insbesondere für  $X \sim \mathcal{N}(0, 1)$ :

$$\Pr[a \leq X \leq b] = \Phi(b) - \Phi(a) .$$

$\Phi$  ist analytisch nicht auswertbar, da wir keine analytische Lösung für das Integral in der Definitionsgleichung (4) von  $\Phi$  kennen. Es existieren aber Funktionstabellen für  $\Phi$ , die mit numerischen Verfahren erstellt sind (s. Lehrbuch im Anhang A). Für eine beliebige normalverteilte Variable  $Y = \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$  können wir Lemma 1.13 anwenden und erhalten die Formel

$$\Pr[a \leq Y \leq b] = \Pr\left[\frac{a - \mu}{\sigma} \leq X \leq \frac{b - \mu}{\sigma}\right] = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) . \quad (5)$$

### 1.3.2 Exponentialverteilung

Eine Zufallsvariable  $X$  mit Wertebereich  $\mathbb{R}_0^+$ , Dichte

$$f_X(x) := \lambda e^{-\lambda x}$$

für alle  $x \geq 0$  und Verteilung

$$F_X(x) := \lambda \int_0^x e^{-\lambda t} dt = -[e^{-\lambda t}]_0^x = 1 - e^{-\lambda x} \quad (6)$$

heißt *exponentialverteilt mit Parameter*  $\lambda > 0$ . Es folgt sofort

$$\int_0^\infty f_X(x) dx = \lim_{x \rightarrow \infty} F_X(x) = 1 - 0 = 1 \quad ,$$

d.h.,  $f_X$  ist eine „ordentliche“ Dichte. Weiterhin ergeben sich aus (6) die nützlichen Gleichungen

$$1 - F_X(x) = e^{-\lambda x} \quad (7)$$

sowie

$$\Pr[a \leq X \leq b] = F_X(b) - F_X(a) = (1 - e^{-\lambda b}) - (1 - e^{-\lambda a}) = e^{-\lambda a} - e^{-\lambda b} \quad .$$

Die Verteilung  $F_X(x)$  ist durch die Gleichung (7) eindeutig bestimmt. Wenn also eine Zufallsvariable  $X$  der Gleichung (7) genügt, dann ist sie exponentialverteilt mit Parameter  $\lambda$ .

Wie folgendes Resultat lehrt, stimmt Parameter  $\lambda$  überein mit dem Kehrwert des Erwartungswertes (und dem Kehrwert der Standardabweichung):

**Lemma 1.21** *Für eine exponentialverteilte Zufallsvariable  $X$  gilt*

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{und} \quad \text{Var}[X] = \frac{1}{\lambda^2} \quad .$$

**Beweis** Der Beweis ergibt sich aus einer einfachen Rechnung unter Verwendung von partieller Integration:

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty \underbrace{t}_u \underbrace{\lambda e^{-\lambda t}}_{v'} dt = \left[ -te^{-\lambda t} \right]_0^\infty + \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda} \\ \mathbb{E}[X^2] &= \int_0^\infty \underbrace{t^2}_u \underbrace{\lambda e^{-\lambda t}}_{v'} dt = \left[ -t^2 e^{-\lambda t} \right]_0^\infty + 2 \underbrace{\int_0^\infty t e^{-\lambda t} dt}_{\mathbb{E}[X]/\lambda} = \frac{2}{\lambda^2} \\ \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \end{aligned}$$

**qed.**

Skalierung einer exponentialverteilten Zufallsvariable mit einem konstanten Faktor liefert wieder eine exponentialverteilte Zufallsvariable:

**Satz 1.22** Falls  $X$  exponentialverteilt ist mit Parameter  $\lambda$ , dann ist  $Y := aX$  (mit einer Konstanten  $a > 0$ ) exponentialverteilt mit Parameter  $\lambda/a$ .

**Beweis** Es genügt zu zeigen, dass  $F_Y(x) = 1 - e^{-(\lambda/a)x}$ :

$$F_Y(x) = \Pr[Y \leq x] = \Pr[aX \leq x] = \Pr\left[X \leq \frac{x}{a}\right] = F_X\left(\frac{x}{a}\right) = 1 - e^{-(\lambda/a)x} .$$

**qed.**

Exponentialverteilte Zufallsvariable lassen sich mit der Eigenschaft der „Gedächtnislosigkeit“ (Stetigkeit und positiver Wertebereich vorausgesetzt) charakterisieren:

**Satz 1.23 (Gedächtnislosigkeit der Exponentialverteilung)** Eine stetige Zufallsvariable  $X$  mit Wertebereich  $\mathbb{R}^+$  ist genau dann exponentialverteilt, wenn für alle  $x, y > 0$  gilt:

$$\Pr[X > x + y | X > y] = \Pr[X > x] . \quad (8)$$

**Beweis** Wir zeigen zunächst, dass eine exponentialverteilte Zufallsvariable  $X$  die Bedingung (8) erfüllt:

$$\begin{aligned} \Pr[X > x + y | X > y] &= \frac{\Pr[X > x + y, X > y]}{\Pr[X > y]} \\ &= \frac{\Pr[X > x + y]}{\Pr[X > y]} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} = e^{-\lambda x} = \Pr[X > x] . \end{aligned}$$

Für die umgekehrte Beweisrichtung nehmen wir an, dass  $X$  die Bedingung (8) erfüllt, und machen einen Parameter  $\lambda$  ausfindig, mit welchem  $X$  exponentialverteilt ist. Hierfür genügt es die Gleichung  $g(x) := \Pr[X > x] = e^{-\lambda x}$  nachzuweisen. Die folgenden drei Behauptungen dienen diesem Zweck. Vorab weisen wir darauf hin, dass  $0 \leq g(x) \leq 1$ , da  $g$  als Wahrscheinlichkeit eines Ereignisses definiert ist.

**Behauptung 1**  $g(x + y) = g(x)g(y)$ .

Dies ergibt sich unter Verwendung des Satzes von der totalen Wahrscheinlichkeit und der Gleichung (8) durch eine einfache Rechnung:

$$\begin{aligned} g(x + y) &= \Pr[X > x + y] \\ &= \underbrace{\Pr[X > x + y | X > y]}_{=\Pr[X > x]} \Pr[X > y] + \underbrace{\Pr[X > x + y | X \leq y]}_{=0} \Pr[X \leq y] \\ &= \Pr[X > x] \Pr[X > y] = g(x)g(y) \end{aligned}$$

**Behauptung 2**  $g(p/q) = g(1)^{p/q}$  für alle  $p, q \in \mathbb{N}$ .

Mit Behauptung 1 ergeben sich die folgenden beiden Gleichungen:

$$g\left(\frac{p}{q}\right) = g\left(\underbrace{\frac{1}{q} + \cdots + \frac{1}{q}}_{p\text{-mal}}\right) = g\left(\frac{1}{q}\right)^p$$

$$g(1) = g\left(\underbrace{\frac{1}{q} + \cdots + \frac{1}{q}}_{q\text{-mal}}\right) = g\left(\frac{1}{q}\right)^q$$

Aus der zweiten Gleichung folgt  $g(1/q) = g(1)^{1/q}$  und zusammen mit der ersten Gleichung ergibt sich dann die Behauptung.

**Behauptung 3**  $0 < g(1) < 1$ .

Es sei an  $0 \leq g(x) \leq 1$  erinnert. Wir haben also lediglich  $g(1) \notin \{0, 1\}$  zu zeigen. Wir führen den Beweis durch Widerspruch. Da  $X$  eine stetige Zufallsvariable mit Wertebereich  $\mathbb{R}^+$  ist, muss  $g(0) = \Pr[X > 0] = 1$  und  $\lim_{x \rightarrow \infty} g(x) = \lim_{x \rightarrow \infty} \Pr[X > x] = 0$  gelten. Aus  $g(1) = 0$  würde nun aber mit Behauptung 2 folgen, dass  $g(1/q) = g(1)^{1/q} = 0$  für alle  $q \in \mathbb{N}$ . Aus Stetigkeitsgründen müsste dann auch  $g(0) = 0$  gelten im Widerspruch zu  $g(0) = 1$ . Aus  $g(1) = 1$  würde mit Behauptung 2 folgen, dass  $g(p) = g(1)^p = 1$  für alle  $p \in \mathbb{N}$ . Aus Stetigkeitsgründen müsste dann auch  $\lim_{p \rightarrow \infty} g(p) = 1$  gelten im Widerspruch zu  $\lim_{p \rightarrow \infty} g(p) = 0$ . Somit gilt  $0 < g(1) < 1$ .

Aus  $0 < g(1) < 1$  folgt  $\lambda := -\ln g(1) \in \mathbb{R}^+$ . Mit dieser Wahl von  $\lambda$  gilt dann  $g(1) = e^{-\lambda}$  und somit  $g(p/q) = e^{-(p/q)\lambda}$  für alle  $p, q \in \mathbb{N}$ . Da  $p/q$  alle positiven rationalen Zahlen durchläuft und  $\mathbb{Q}$  dicht in  $\mathbb{R}$  liegt, folgt dann aus Stetigkeitsgründen  $g(x) = e^{-\lambda x}$  für alle  $x \in \mathbb{R}^+$ . Folglich ist  $X$  exponentialverteilt mit Parameter  $\lambda$ . **qed.**

Die Exponentialverteilung lässt sich als Grenzwert von (geeignet skalierten) geometrischen Verteilungen auffassen. Wir wollen diesen Zusammenhang kurz skizzieren. Für alle  $n \in \mathbb{N}$  sei  $X_n$  eine geometrisch verteilte Zufallsvariable mit Parameter  $p_n := \lambda/n$ . Dann gilt für alle Vielfachen  $x$  von  $1/n$  (so dass  $xn \in \mathbb{N}$ )

$$\begin{aligned} \Pr[X_n \leq xn] &= \sum_{i=1}^{xn} (1 - p_n)^{i-1} p_n \\ &= p_n \sum_{i=0}^{xn-1} (1 - p_n)^i \\ &= p_n \frac{1 - (1 - p_n)^{xn}}{1 - (1 - p_n)} \\ &= 1 - (1 - p_n)^{xn} = 1 - \left(1 - \frac{\lambda}{n}\right)^{xn}. \end{aligned}$$



Für  $Y_n := X_n/n$  ergibt sich dann asymptotisch eine Exponentialverteilung mit Parameter  $\lambda$ :<sup>7</sup>

$$\lim_{n \rightarrow \infty} \Pr[Y_n \leq x] = \lim_{n \rightarrow \infty} \Pr[X_n \leq xn] = \lim_{n \rightarrow \infty} \left( 1 - \left( 1 - \frac{\lambda}{n} \right)^{xn} \right) = 1 - e^{-\lambda x} ,$$

wobei in der letzten Gleichung die Formel  $\lim_{n \rightarrow \infty} (1 - \lambda/n)^n = e^{-\lambda}$  verwendet wurde.

Eine geometrisch verteilte Zufallsvariable mit Parameter  $p$  gibt an, im wievielten Versuch einer Folge unabhängiger Bernoulli-Experimente mit Erfolgswahrscheinlichkeit  $p$  der erste Erfolg aufgetreten ist (Warten auf den Erfolg). Die obige Diskussion zur Darstellung einer exponentiell mit Parameter  $\lambda$  verteilten Zufallsvariable  $Y$  als Grenzwert der Verteilungen  $Y_n = X_n/n$  legt folgende Interpretation nahe:  $Y$  repräsentiert die (kontinuierliche) Wartezeit bis zum ersten Eintreten eines Ereignisses, welches in diesem Zusammenhang auch „Treffer“ genannt wird. Wir erhalten einen Treffer im Mittel nach  $1/\lambda$  Zeiteinheiten. Daher wird Parameter  $\lambda$  auch „Trefferrate“ genannt.<sup>8</sup>

Es ist daher nicht verwunderlich, dass die Exponentialverteilung überall dort Anwendung findet, wo es um (kontinuierliche) Wartezeiten geht.<sup>9</sup> Das Konzept der kontinuierlichen Wartezeit finden wir zum Beispiel in der Warteschlangentheorie und bei der Analyse der Lebenszeit (bis zum Zerfall) von radioaktivem Material.

**Beispiel 1.24** *Das Cäsium-Isotop  $^{134}_{55}\text{Cs}$  hat eine mittlere Lebenszeit von ungefähr 3.03 Jahren bzw.  $1.55 \cdot 10^6$  Minuten. Wir können seine Lebenszeit als Zufallsvariable  $X$  auffassen, die (bei einer Minute als Zeiteinheit) exponentiell mit Parameter*

$$\lambda = \frac{1}{\mathbb{E}[X]} = \frac{1}{1.5 \cdot 10^6} \approx 0.645 \cdot 10^{-6}$$

*verteilt ist. Da ein „Treffer“ beim Zerfall des Atoms vorliegt, heißt Parameter  $\lambda$  in diesem Kontext „Zerfallsrate“.*

## 1.4 Zusammenspiel mehrerer numerischer Zufallsvariablen

Um die wesentlichen Konzepte beim Zusammenspiel mehrerer numerischer Zufallsvariablen möglichst einfach und einprägsam einzuführen, beschränken wir uns zumeist auf den Fall zweier Variablen, sagen wir  $X$  und  $Y$ . Mehr als zwei Zufallsvariable lassen sich aber völlig analog abhandeln.

Eine stetige (oder stückweise stetige) Funktion  $f : \mathbb{R}^2 \rightarrow \mathbb{R}_0^+$ , welche die Bedingung

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) d(x, y) = 1$$

---

<sup>7</sup>Damit die folgende Rechnung schlüssig ist, muss  $x$  wiederum ein Vielfaches von  $1/n$  sein. Da eine beliebige Zahl  $x \in \mathbb{R}^+$  bis auf einen additiven Fehler von  $1/n$  durch ein Vielfaches von  $1/n$  approximiert werden kann, verallgemeinert sich das Ergebnis aber aus Stetigkeitsgründen auf alle  $x \in \mathbb{R}^+$ .

<sup>8</sup>Wir werden später sehen (s. Folgerung 1.34), dass wir umgekehrt auch pro Zeiteinheit im Mittel  $\lambda$  Treffer haben.

<sup>9</sup>Wegen der Gedächtnislosigkeit der Exponentialverteilung handelt kann sie aber nur dann zur Modellierung herangezogen werden, wenn das zugrunde liegende System „nicht altert“.

erfüllt, induziert ein Wahrscheinlichkeitsmaß auf  $\mathcal{B}_2$  (der  $\sigma$ -Algebra der 2-dimensionalen Borelschen Mengen) über die Gleichung

$$\Pr(A) := \int_A f(x, y) d(x, y) .$$

**Definition 1.25 (gemeinsame Dichte zweier Zufallsvariablen)** *Ein Paar  $(X, Y)$  von numerischen Zufallsvariablen hat  $f$  als Dichte, falls*

$$\Pr[(X, Y) \in A] := \int_A f(x, y) d(x, y) .$$

*In diesem Fall notieren wir die Dichte auch als  $f_{X,Y}$ .*

Wir konzentrieren uns wieder auf den Fall von stetigen Dichten  $f_{X,Y}$  (und sprechen dann auch von einem Paar stetiger, numerischer Zufallsvariablen).

**Definition 1.26 (gemeinsame Verteilung zweier Zufallsvariablen)** *Für ein Paar  $(X, Y)$  von stetigen, numerischen Zufallsvariablen mit Dichte  $f_{X,Y}$  ist die Verteilung gegeben durch*

$$F_{X,Y}(x, y) := \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) d(u, v) .$$

Die Dichte  $f_{X,Y}$  ergibt sich aus der Verteilung durch 2-faches partielles Ableiten:

$$f_{X,Y}(x_0, y_0) = \frac{F_{X,Y}}{\partial x \partial y}(x_0, y_0) .$$

Aus der gemeinsamen Dichte bzw. Verteilung zweier Zufallsvariablen kann man die Dichte bzw. Verteilung jeder einzelnen von ihnen als Randdichte bzw. Randverteilung erhalten:

**Bemerkung und Definition 1.27 (Randdichte, Randverteilung)** *Wenn die Zufallsvariablen  $X, Y$  über ihre gemeinsame Dichte gegeben sind, dann bezeichnen wir*

$$f_X(u) := \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv$$

*als die Randdichte von  $X$  und*

$$F_X(x) := \Pr[X \leq x] = \int_{-\infty}^x f_X(u) du = \int_{-\infty}^x \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv du$$

*als die Randverteilung von  $X$ .*

Wir setzen im Folgenden voraus, dass ein Paar  $(X, Y)$  von stetigen, numerischen Zufallsvariablen über die gemeinsame Dichte (oder die gemeinsame Verteilung) gegeben ist.

Der Fall unabhängiger Paare von Zufallsvariablen verdient eine besondere Würdigung:

**Definition 1.28 (Unabhängigkeit stetiger Zufallsvariablen)**  $X, Y$  heißen unabhängig, wenn für alle  $x, y \in \mathbb{R}$  die Bedingung

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

erfüllt ist (mit den Randverteilungen auf der rechten Seite).

Durch partielles Ableiten überträgt sich diese „Multiplikatitivität“ auf die Dichten:

$$f_{X,Y}(x_0, y_0) = \frac{F_{X,Y}}{\partial x \partial y}(x_0, y_0) = \frac{F_X}{\partial x}(x_0) \frac{F_Y}{\partial y}(y_0) = f_X(x_0) f_Y(y_0) .$$

Die Multiplikatitivität der Dichten bewirkt, dass sich Integrale über zweidimensionale Borelsche Mengen in zwei Integrale über eindimensionale Borelsche Mengen separieren lassen. Um diese Aussage präziser zu formulieren, benötigen wir die folgende (in Abbildung 2 illustrierte)

**Definition 1.29 (Projektion auf eindimensionale Mengen)** Zu einer 2-dimensionalen Borelschen Menge  $A \in \mathcal{B}_2$  betrachten wir die folgende eindimensionale Borelsche Menge:

$$A_1 := \{x \mid \exists y : (x, y) \in A\}$$

Für jedes  $x \in A_1$  betrachten wir weiterhin auch die eindimensionale Borelsche Menge

$$A_2(x) := \{y \mid (x, y) \in A\} .$$

Im Falle unabhängiger Zufallsvariablen  $X, Y$  kann ein Integral über  $A \in \mathcal{B}_2$  nach folgendem Schema berechnet werden:

$$\int_A f_{X,Y}(x, y) d(x, y) = \int_{A_1} \left( \int_{A_2(x)} f_Y(y) dy \right) f_X(x) dx .$$

Besonders einfach gestaltet sich die Berechnung, wenn  $A$  die Form  $A_1 \times A_2$  für  $A_1, A_2 \in \mathcal{B}_1$  hat, weil dann  $A_2(x) = A_2$  für alle  $x \in A_1$ . In diesem Fall erhalten wir

$$\int_{A_1 \times A_2} f_{X,Y}(x, y) d(x, y) = \left( \int_{A_1} f_X(x) dx \right) \cdot \left( \int_{A_2} f_Y(y) dy \right) ,$$

was gleichbedeutend zu

$$\Pr[(X \in A_1, Y \in A_2)] = \Pr[X \in A_1] \cdot \Pr[Y \in A_2]$$

ist.

Als Anwendung dieser Überlegungen betrachten wir folgenden

**Satz 1.30** Für unabhängige mit Parametern  $\lambda_1, \dots, \lambda_n$  exponentialverteilte Zufallsvariable  $X_1, \dots, X_n$  ist dann auch die Zufallsvariable

$$X := \min\{X_1, \dots, X_n\}$$

exponentialverteilt mit Parameter  $\lambda_1 + \dots + \lambda_n$ .

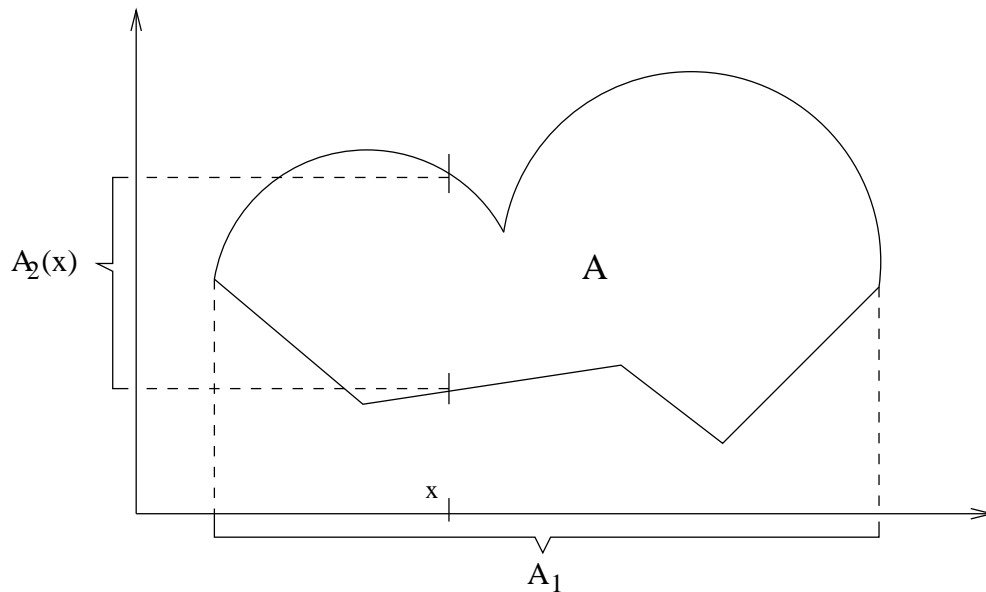


Abbildung 2: Die von einer zweidimensionalen Borelschen Menge  $A$  induzierten Mengen  $A_1$  und  $A_2(x)$ .

**Beweis** Wir führen den Beweis für  $n = 2$ . (Der allgemeine Fall ergibt sich analog.) Es genügt zu zeigen, dass  $1 - F_X(t) = e^{-(\lambda_1 + \lambda_2)t}$  erfüllt ist:

$$\begin{aligned}
 1 - F_X(t) &= \Pr[X > t] \\
 &= \Pr[\min\{X_1, X_2\} > t] \\
 &= \Pr[X_1 > t, X_2 > t] \\
 &= \Pr[X_1 > t] \cdot \Pr[X_2 > t] \\
 &= e^{-\lambda_1 t} \cdot e^{-\lambda_2 t} = e^{-(\lambda_1 + \lambda_2)t}
 \end{aligned}$$

qed.

Wir betrachten zwei Anwendungsbeispiele für diesen Satz:

**Beispiel 1.31** Bei  $n$  Atomen mit jeweils Zerfallsrate  $\lambda$  (und daher mittlerer Lebensdauer  $1/\lambda$ ) dauert es im Mittel  $1/(n\lambda)$  Zeiteinheiten bis das erste Atom zerfällt. (Hierbei wird die Annahme gemacht, dass die Atome unabhängig voneinander zerfallen.)

**Beispiel 1.32** Bei einer Telefonzelle betrage die mittlere Wartezeit  $m$  Minuten. Dann reduziert sich die Wartezeit bei insgesamt  $n$  Telefonzellen auf  $m/n$  Minuten. (Hier wird die Annahme gemacht, dass die diversen Telefonate sich, was ihre Dauer angeht, nicht gegenseitig beeinflussen.)

Bleiben wir noch etwas bei der Exponentialverteilung. Es sei daran erinnert, dass  $\lambda$  als Trefferrate und  $1/\lambda$  als die mittlere Wartezeit bis zum ersten Treffer interpretiert werden

kann. Wir setzen voraus, dass  $T_1, T_2, T_3, \dots$  unabhängige exponentialverteilte Zufallsvariable mit jeweils Trefferrate  $\lambda$  sind. Somit zählt die Zufallsvariable

$$X(t) := \max\{n \in \mathbb{N}_0 \mid T_1 + \dots + T_n \leq t\} \quad (9)$$

die Anzahl der Treffer bis zum Zeitpunkt  $t$ .

**Satz 1.33 (ohne Beweis)** *Es seien  $T_1, T_2, T_3, \dots$  unabhängige Zufallsvariable und  $X(t)$  sei durch (9) gegeben. Dann ist  $X(t)$  genau dann Poisson-verteilt mit Parameter  $t\lambda$ , wenn  $T_1, T_2, T_3, \dots$  exponentialverteilt mit jeweils Parameter  $\lambda$  sind.*

Wie wir aus dem Kapitel über diskrete Wahrscheinlichkeitstheorie wissen, gibt eine binomialverteilte Zufallsvariable zu Parametern  $t, p$  die Anzahl der Erfolge bei  $t$  unabhängigen Bernoulli-Experimenten mit jeweils Erfolgswahrscheinlichkeit  $p$  an. Das bedeutet, dass für geometrisch verteilte unabhängige  $T_i$  die Zufallsvariable  $X(t)$  binomialverteilt ist. Da der Grenzübergang  $n \rightarrow \infty$  die geometrische Verteilung in die Exponentialverteilung und die Binomialverteilung in die Poisson-Verteilung überführt, ist Satz 1.33 intuitiv einleuchtend.

Da der Parameter einer Poisson-verteilten Zufallsvariable gerade ihren Erwartungswert angibt, führt Satz 1.33 direkt zu der

**Folgerung 1.34**  $\mathbb{E}[X(t)] = t\lambda$ . *Wir haben also bei einer Trefferrate  $\lambda$  im Mittel  $t\lambda$  Treffer nach  $t$  Zeiteinheiten (und insbesondere im Mittel  $\lambda$  Treffer nach einer Zeiteinheit).*

Wir beschließen diesen Abschnitt mit einer weiteren Anwendung unserer Rechenregeln für Integrale über Mengen aus  $\mathcal{B}_2$ .

**Satz 1.35** *Für unabhängige, stetige Zufallsvariable  $X, Y$  hat die Zufallsvariable  $Z := X + Y$  die Dichte*

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx .$$

**Beweis** Wir beweisen im Folgenden die Gleichung

$$F_Z(z) = \int_{-\infty}^z \left( \int_{-\infty}^{+\infty} f_X(x) f_Y(t - x) dx \right) dt \quad (10)$$

Durch Bilden der ersten Ableitung folgt hieraus direkt die Aussage des Satzes. Zum Nachweis von (10), betrachten wir die (von  $z$  abhängige) 2-dimensionale Borelsche Menge

$$A := \{(x, y) \in \mathbb{R}^2 \mid x + y \leq z\}$$

und die resultierenden (ebenfalls von  $z$  abhängigen) eindimensionalen Borelschen Mengen

$$A_1 = (-\infty, +\infty) \text{ und } A_2(x) := \{y \in \mathbb{R} \mid x + y \in A\} = (-\infty, z - x] .$$

Nun lässt sich  $F_Z(z)$  berechnen wie folgt:

$$\begin{aligned}
F_Z(z) &= \Pr[Z \leq z] \\
&= \Pr[X + Y \leq z] \\
&= \Pr[(X, Y) \in A] \\
&= \int_A f_{X,Y}(x, y) d(x, y) \\
&= \int_{A_1} \left( \int_{A_2(x)} f_Y(y) dy \right) f_X(x) dx \\
&= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{z-x} f_Y(y) dy \right) f_X(x) dx \\
&\stackrel{(*)}{=} \int_{-\infty}^{+\infty} \left( \int_{-\infty}^z f_Y(y' - x) dy' \right) f_X(x) dx \\
&= \int_{-\infty}^z \left( \int_{-\infty}^{+\infty} f_X(x) f_Y(y - x) dx \right) dy
\end{aligned}$$

Die mit (\*) markierte Gleichung ist dabei eine Anwendung der Substitutionsregel, wobei  $y$  durch  $y' = y + x$  (so dass  $y = y' - x$ ) substituiert wurde. (In der darauf folgenden Gleichung wurde  $y'$  der Einfachheit halber wieder zu  $y$  umbenannt.) **qed.**

**Folgerung 1.36 (ohne Beweis)** Für unabhängige Zufallsvariable  $X_1, \dots, X_n$  mit  $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$  und Konstanten  $a_1, \dots, a_n \in \mathbb{R}$  ist dann auch die Zufallsvariable  $Z := a_1 X_1 + \dots + a_n X_n$  normalverteilt. Genauer:  $Z \sim \mathcal{N}(\mu, \sigma^2)$  mit

$$\mu := a_1 \mu_1 + \dots + a_n \mu_n \text{ und } \sigma^2 := a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2 .$$

Die Gleichungen für  $\mu$  und  $\sigma$  ergeben sich direkt aus den bekannten Rechenregeln für Erwartungswert und Varianz. Die Tatsache, dass  $Z$  normalverteilt ist, lässt sich mit Hilfe von Satz 1.35 (durch Betrachtung der Dichte von  $Z$ ) nachweisen. Wir verzichten aber hier auf weitere Details.

## 1.5 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz besagt im Wesentlichen, dass eine (geeignet normalisierte) Summe von unabhängigen gleichverteilten Zufallsvariablen (mit endlichem Erwartungswert und endlicher Varianz) asymptotisch standard-normalverteilt ist. Dieser Satz unterstreicht nachdrücklich die besondere Rolle der Normalverteilung in der Statistik. Wir präsentieren dieses Resultat hier ohne Beweis:

**Satz 1.37 (Zentraler Grenzwertsatz)** Es sei  $X_1, X_2, X_3, \dots$  eine Folge von unabhängigen gleichverteilten Zufallsvariablen mit Erwartungswert  $\mu$  und Varianz  $\sigma^2 > 0$ . Für alle  $n \in \mathbb{N}$  sei die Zufallsvariable  $Z_n$  gegeben durch

$$\begin{aligned}
Y_n &:= X_1 + \dots + X_n \\
Z_n &:= \frac{Y_n - n\mu}{\sigma\sqrt{n}}
\end{aligned}$$

und  $F_n^*$  bezeichne die Verteilung von  $Z_n$ . Dann gilt

$$\lim_{n \rightarrow \infty} F_n^*(x) = \Phi(x) \quad , \quad (11)$$

d.h.,  $F_n^*$  stimmt beim Grenzübergang  $n \rightarrow \infty$  mit der Verteilung einer standard-normalverteilten Zufallsvariable überein. In Zeichen:

$$Z_n \sim \mathcal{N}(0, 1) \text{ für } n \rightarrow \infty \quad .$$

Anhand der Rechenregeln für Erwartungswert und Varianz kann man leicht nachrechnen, dass  $\mathbb{E}[Z_n] = 0$  und  $\text{Var}[Z_n] = 1$  für alle  $n \in \mathbb{N}$ . Die Gleichung (11) ist jedoch nicht so einfach herzuleiten.<sup>10</sup>

Die Anwendung des zentralen Grenzwertsatzes auf Bernoulli-verteilte Zufallsvariable liefert die folgende Aussage<sup>11</sup>:

**Folgerung 1.38 (Grenzwertsatz von DeMoivre/Laplace)** *Es sei  $X_1, X_2, X_3, \dots$  eine Folge von unabhängigen Bernoulli-verteilten Zufallsvariablen jeweils mit Erfolgswahrscheinlichkeit  $p$  (und somit Erwartungswert  $p$  und Varianz  $p(1-p)$ ). Für alle  $n \in \mathbb{N}$  sei die Zufallsvariable  $H_n^*$  gegeben durch*

$$\begin{aligned} H_n &:= X_1 + \dots + X_n \\ H_n^* &:= \frac{H_n - np}{\sqrt{np(1-p)}} \end{aligned}$$

Dann gilt

$$H_n^* \sim \mathcal{N}(0, 1) \text{ für } n \rightarrow \infty \quad .$$

Wir skizzieren abschließend eine Anwendung des Satzes von DeMoivre/Laplace. Es bezeichne  $F_n$  die Verteilung von (der binomial verteilten Zufallsvariable)  $H_n$  und  $F_n^*$  die Verteilung von  $H_n^*$ . Die direkte Auswertung von  $F_n(t)$  gemäß der exakten Formel

$$F_n(t) = \Pr[H_n \leq t] = \sum_{i=0}^t \binom{n}{i} p^i (1-p)^{n-i}$$

ist leider sehr mühselig, da die hierin auftretende Summe nicht durch eine elementare (und leicht auszuwertende) Funktion ausgedrückt werden kann. Mit Hilfe des Satzes von DeMoivre/Laplace (und der Funktionstabelle für die Verteilung  $\Phi$  der Standard-Normalverteilung im im Anhang A des Lehrbuchs) können wir aber  $F_n(t)$  bequem approximieren:

$$F_n(t) = \Pr[H_n \leq t] = \Pr \left[ H_n^* \leq \frac{t - np}{\sqrt{np(1-p)}} \right] = F_n^* \left( \frac{t - np}{\sqrt{np(1-p)}} \right) \approx \Phi \left( \frac{t - np}{\sqrt{np(1-p)}} \right) \quad . \quad (12)$$

Eine Faustregel besagt, dass diese Approximation bereits recht gut ist ab  $n \geq 5 / \min\{p, 1-p\}$  (also ab  $n \geq 10$  falls  $p = 1/2$ ).

<sup>10</sup>S. aber das Lehrbuch zur Vorlesung, das den Beweis des Grenzwertsatzes enthält.

<sup>11</sup>die aber historisch gesehen zu einem früheren Zeitpunkt bewiesen wurde

**Beispiel 1.39** Wenn wir eine faire Münze  $10^6 = 1000000$ -mal werfen, wird im Mittel 500000-mal „Kopf“ realisiert. Was ist nun die Wahrscheinlichkeit, dass „Kopf“ öfter als 500500-mal realisiert wird?

Offensichtlich ist die gesuchte Wahrscheinlichkeit gegeben durch  $1 - F_{10^6}(500500)$  (bei Erfolgswahrscheinlichkeit  $p = 1/2$ ). Direkte Berechnung führt zu dem Ausdruck

$$1 - F_{10^6}(500500) = \sum_{i=500501}^{10^6} \binom{10^6}{i} 2^{-10^6} ,$$

dessen Auswertung ein „Alptraum“ wäre. Anwendung von (12) und Nachschlagen in der Funktionstabelle für  $\Phi$  (aus Anhang A des Lehrbuchs) liefert

$$1 - F_{10^6}(500500) \approx 1 - \Phi\left(\frac{500500 - 500000}{\sqrt{10^6/4}}\right) = 1 - \Phi(1) \approx 0.1573 .$$