Leitfaden zu:
Mathematische Methoden der Biowissenschaften II:
Wahrscheinlichkeitstheorie und Statistik

Ellen Baake, Michael Baake

LaTeX Satz von Nils Hoffmann SoSe 2006
überarbeitet von Christoph Richard SoSe 2008
Korrekturen SoSe 2013
Korrekturen SoSe 2020
Korrekturen SoSe 2024

25. April 2024

Inhaltsverzeichnis

Kapitel 1

Elementare Konzepte und Beispiele

1.1 Der Münzwurf

Das Ergebnis eines Münzwurfs lautet entweder "Kopf" oder "Zahl". Nach n-maligem Werfen bezeichnen wir mit n_K die Zahl der Würfe, bei denen Kopf aufgetreten ist, und mit n_Z die Zahl der Würfe, bei denen Zahl aufgetreten ist, $n_K + n_Z = n$. Intuitiv "erwarten" wir $n_K \approx n_Z$ – obwohl jeder einzelne Wurf zufällig ist! Darüber hinaus erwarten wir, dass die relative Häufigkeit n_K/n des Auftretens von Kopf sich immer besser einer festen Zahl p_K annähert, je größer n ist. Für eine faire Münze erwarten wir $p_K = 1/2$. Wir wollen ein mathematisches Modell aufstellen, das die obige Situation beschreibt und innerhalb dessen man die obigen Vermutungen präzisieren und beweisen kann. Gegen Ende dieses Kurses werden wir sogar statistische Verfahren zur Hand haben, mit denen die Gültigkeit der obigen Vermutungen quantitativ getestet werden kann.

1.1.1 Einfacher Münzwurf

$1 \mathrel{\widehat{=}} K; \quad 0 \mathrel{\widehat{=}} Z; \quad \mathbb{P}(1) = p =: p_K; \quad \mathbb{P}(0) = q = 1 - p =: p_Z$

$\Omega = \{0, 1\}$ bildet den **Grundraum** (oder *Wahrscheinlichkeitsraum*). Den eingeklammerten Begriff wollen wir nicht verwenden, da er später anders eingesetzt wird (vgl. Def. 1.6).

$\mathcal{P}(\Omega) = \{\varnothing, \{0\}, \{1\}, \{0, 1\}\}$ bildet den **Ereignisraum** und entspricht hier der Potenzmenge von Ω, also der Menge aller Teilmengen von Ω.

$\{1\}, \{0\}$ sind die **Elementarereignisse**.

Ereignis	Deutung	Wahrscheinlichkeit
$\varnothing$	weder K noch Z	0
$\{1\}$	K	p
$\{0\}$	Z	$q = 1 - p$
$\{1, 0\}$	K oder Z	1

Hierbei ist $\varnothing$ das "unmögliche" Ereignis, $\{1,0\}$ ist das "sichere" Ereignis.

Definition 1.1 (Faire Münze)**.** Eine Münze heißt **fair**, falls $p_K = p_Z$ gilt. (Dabei gilt dann $p_K = p_Z = \frac{1}{2}$.)

1.1.2 Mehrfacher Münzwurf

Wie groß ist die Wahrscheinlichkeit, bei n Würfen genau ℓ-mal eine "1" zu werfen?

Der Grundraum besteht aus den n-Tupeln $(s_1, s_2, \ldots, s_n)$ mit $s_i \in \{0,1\}$. Es gibt also 2^n Elementarereignisse. Von diesen tritt in $\binom{n}{\ell} = \frac{n!}{\ell!(n-\ell)!}$ Fällen genau ℓ-mal eine 1 auf. Jedes Elementarereignis sei gleich wahrscheinlich (faire Münze!), habe also die Wahrscheinlichkeit 2^{-n}. Dann ist

$$P_\ell := \mathbb{P}(\text{“}\ell\text{-mal 1”}) = \binom{n}{\ell}\frac{1}{2^n} \geq 0. \tag{1.1}$$

Die Summe aller Wahrscheinlichkeiten P_ℓ ist nach Konstruktion 1. Das bedeutet[1]

$$\sum_{\ell=0}^{n} P_\ell = \frac{1}{2^n}\sum_{\ell=0}^{n}\binom{n}{\ell} \stackrel{!}{=} 1. \tag{1.2}$$

Die oben auftretende Identität für Binomialkoeffizienten (und einige weitere) können Sie sich am Pascal'schen Dreieck gut veranschaulichen. Ein algebraischer Beweis gelingt rasch mit dem binomischen Lehrsatz, siehe Gleichung (1.5).

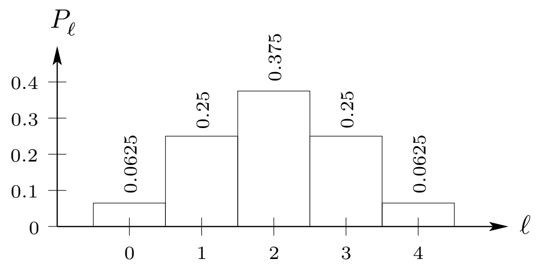

Abbildung 1.1: Säulendiagramm der Binomialverteilung für $n = 4$

Jetzt "zinken" wir die Münze: $\mathbb{P}(K) = p$, $\mathbb{P}(Z) = q = 1 - p$. Was ist jetzt die Wahrscheinlichkeit, genau ℓ-mal die 1 zu würfeln?

Wir verwenden denselben Grundraum wie im fairen Fall. Die Wahrscheinlichkeit eines Elementarereignisses mit ℓ-mal "1" sei nun $p^\ell\ q^{n-\ell}$. (Überlegen Sie sich,

[1] Das "!" über den Gleichheitszeichen verweist darauf, dass man über die Gleichheit nicht hinweglesen soll, sondern über sie nachdenken, bzw. sie nachrechnen soll.

warum dies die richtige Wahl der Wahrscheinlichkeiten ist!) Damit ergibt sich wie oben

$$P_\ell := \mathbb{P}(\text{“}\ell\text{-mal 1”}) = \binom{n}{\ell} p^\ell \, q^{n-\ell}. \tag{1.3}$$

Dann gilt für die Summe der Wahrscheinlichkeiten immer noch

$$\sum_{\ell=0}^{n} \binom{n}{\ell} p^\ell \, (1-p)^{n-\ell} = \big(p + (1-p)\big)^n = 1^n = 1. \tag{1.4}$$

Hierbei wurde der binomische Lehrsatz verwendet,

$$(x+y)^n = \sum_{\ell=0}^{n} \binom{n}{\ell} x^\ell \, y^{n-\ell}. \tag{1.5}$$

Definition 1.2 (Binomialverteilung)**.** Sei $0 \le p \le 1$ fest. Die durch die Wahrscheinlichkeiten

$$P_\ell = \binom{n}{\ell} p^\ell \, (1-p)^{n-\ell} \qquad (0 \le \ell \le n)$$

gegebene Verteilung heißt **Binomialverteilung** zu den Parametern n und p, notiert als $\mathcal{B}(n,p)$.

Wieviele 1'er dürfen wir erwarten?

$$\begin{aligned}
\underline{m} &:= \sum_{\ell=0}^{n} \ell \cdot P_\ell = \sum_{\ell=0}^{n} \ell \binom{n}{\ell} p^\ell \, (1-p)^{n-\ell} \\
&\stackrel{!}{=} \sum_{\ell=1}^{n} n \binom{n-1}{\ell-1} p^\ell \, (1-p)^{n-\ell} \\
&\stackrel{!}{=} n \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} \, (1-p)^{n-k-1} \\
&\stackrel{!}{=} np.
\end{aligned} \tag{1.6}$$

Dabei heißt $\underline{m}$ der **Mittelwert** oder (besser!) **Erwartungswert** der Verteilung.
Wie stark weicht das Ergebnis vom Erwartungswert ab?

$$\begin{aligned}
V &:= \sum_{\ell=0}^{n} (\ell - \underline{m})^2 \cdot P_\ell = \sum_{\ell=0}^{n} \Big(\ell^2 - 2\underline{m}\ell + \underline{m}^2\Big) \cdot P_\ell \\
&= \Big(\sum_{\ell=0}^{n} \ell^2 \cdot P_\ell\Big) + \underbrace{\underline{m}^2 \Big(\sum_{\ell} P_\ell\Big) - 2\underline{m}\Big(\sum_{\ell} \ell \cdot P_\ell\Big)}_{= -\underline{m}^2} \\
&= \Big(\sum_{\ell=0}^{n} \ell^2 \cdot P_\ell\Big) - \underline{m}^2.
\end{aligned} \tag{1.7}$$

Diese Größe heißt die **Varianz**, die Größe $\sigma := \sqrt{V}$ die **Standardabweichung** der Verteilung. (Überlegen Sie sich, dass die Größe $\sum_{\ell=0}^{n} (\ell - \underline{m})\, P_\ell$ nicht geeignet ist, um die Abweichung vom Erwartungswert zu beschreiben. Hingegen ist die Größe $\sum_{\ell=0}^{n} |\ell - \underline{m}|\, P_\ell$ im Prinzip geeignet, wird aber nicht häufig verwendet, weil man mit ihr nicht so gut umgehen kann wie mit der Standardabweichung.)

Um die Varianz für den ℓ-fachen Münzwurf zu erhalten, berechnen wir jetzt das sogenannte **2. Moment**:

$$\begin{aligned}
\sum_{\ell=0}^{n} \ell^2 \cdot P_\ell &= \sum_{\ell=1}^{n} \ell^2 \binom{n}{\ell} p^\ell\,(1-p)^{n-\ell} \\
&\overset{!}{=} n \sum_{\ell=1}^{n} \ell \binom{n-1}{\ell-1} p^\ell\,(1-p)^{n-\ell} \\
&\overset{!}{=} np \sum_{\ell=0}^{n-1} (\ell+1) \binom{n-1}{\ell} p^\ell\,(1-p)^{n-1-\ell} \\
&= np\Big(1 + \underbrace{\sum_{\ell=0}^{n-1} \ell \binom{n-1}{\ell} p^\ell (1-p)^{n-1-\ell}}_{\overset{!}{=}(n-1)p}\Big) \\
&= np + n(n-1)p^2. \qquad (1.8)
\end{aligned}$$

Damit erhalten wir

$$V \;=\; np + n(n-1)p^2 - n^2p^2 \;=\; np(1-p) \;=\; npq. \qquad (1.9)$$

Wir fassen unsere Überlegungen im folgenden Satz zusammen.

Satz 1.3. Die Binomialverteilung $\mathcal{B}(n,p)$ mit den Wahrscheinlichkeiten

$$P_\ell \;=\; \binom{n}{\ell} p^\ell\,(1-p)^{n-\ell} \qquad (0 \le \ell \le n)$$

besitzt den Erwartungswert $\underline{m} = np$ und die Varianz $V = npq$. □

Beispiel. Die faire Münze mit $p = \frac{1}{2}$ hat Erwartungswert $\underline{m} = \frac{1}{2}n$ und Varianz $V = \frac{1}{4}n$ bzw. Standardabweichung $\sigma = \frac{1}{2}\sqrt{n}$. Vergleichen Sie dieses Ergebnis mit der Motivation für den Münzwurf zu Beginn des Abschnitts! Die "Fluktuation" $\frac{1}{2}\sqrt{n}$ um den Erwartungswert ist größer, als Sie es anschaulich vielleicht erwarten. Würfeln Sie doch einmal "von Hand" oder mit dem Computer!

1.2 Binomialkoeffizienten

Im letzten Abschnitt haben wir Binomialkoeffizienten verwendet. Hier zur Erinnerung deren Definition und grundlegende Eigenschaften.

Definition 1.4 (allgemeiner Binomialkoeffizient)**.** Für $\alpha \in \mathbb{R}$ und $k \in \mathbb{N}_0$ ist der **Binomialkoeffizient** $\binom{\alpha}{k}$ definiert durch

$$\binom{\alpha}{k} := \frac{\alpha \cdot (\alpha - 1) \cdot \ldots \cdot (\alpha - k + 1)}{1 \cdot 2 \cdot \ldots \cdot k}. \tag{1.10}$$

Speziell gelten folgende Formeln, die man leicht einsieht:

$$\begin{aligned} \binom{\alpha}{0} &= 1 \quad \text{(leeres Produkt im Zähler!)} \\ \binom{n}{k} &= 0 \quad \text{für } n \in \mathbb{N}_0 \text{ und } k > n, \\ \binom{n}{k} &= \frac{n!}{k!(n-k)!} = \binom{n}{n-k} \quad \text{für } n \in \mathbb{N}_0 \text{ und } k \leq n. \end{aligned} \tag{1.11}$$

Der Binomialkoeffizient taucht im allgemeinen binomischen Lehrsatz (Newton, 1667) auf,

$$(1+x)^{\alpha} = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k.$$

Die obige (i.A. unendliche) Summe ist für $|x| < 1$ absolut konvergent. Gilt $\alpha \in \mathbb{N}$, so ist die Summe endlich,

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k,$$

und die Gleichung kann auch durch ein einfaches Abzähl-Argument bewiesen werden (“k aus n”). Zur näherungsweisen Berechnung der Fakultät $n!$ für große n kann man die **Stirling’sche Formel** verwenden,

$$n! \sim \sqrt{2\pi n} \cdot \left(\frac{n}{\mathrm{e}}\right)^n \qquad (\text{für } n \to \infty).$$

Dabei ist $\mathrm{e} \approx 2,718282$ die **Euler’sche Zahl**. Die Schreibweise $\sim$ ist eine Abkürzung für $n!/\big(\sqrt{2\pi n} \cdot \left(\frac{n}{\mathrm{e}}\right)^n\big) \longrightarrow 1$ im Grenzfall $n \to \infty$.

1.3 Der klassische Würfel

Die Elementarereignisse beim Würfeln sind $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$. Deren Wahrscheinlichkeiten sind beim fairen Würfel gleich, $P_1 = P_2 = \ldots = P_6 = \frac{1}{6}$. Damit ergibt sich als Erwartungswert $\underline{m} = \sum_{\ell=1}^{6} \ell \cdot \frac{1}{6} = \frac{7}{2}$ und als Varianz $V = \frac{1}{6}\Big(\sum_{\ell=1}^{6} \ell^2\Big) - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$. Allgemeiner gilt der folgende Satz.

Definition und Satz 1.5 (Gleichverteilung, fairer Würfel)**.** Der faire Würfel mit n Seiten wird durch die sogenannte **Gleichverteilung** beschrieben, also hier durch $P_1 = P_2 = ... = P_n = \frac{1}{n}$. Deren Erwartungswert $\underline{m}$ und die Varianz V sind gegeben durch

$$\underline{m} = \frac{(n+1)}{2} \quad \text{und} \quad V = \frac{(n+1)(n-1)}{12}.$$

Beweis. Durch Ausrechnen. Für den Erwartungswert gilt

$$\underline{m} = \frac{1}{n}\,(1+2+\ldots+n) = \frac{1}{n}\,\frac{n(n+1)}{2} = \frac{(n+1)}{2},$$

wobei wir eine wohlbekannte Formel (von Gauß) verwendet haben, die man durch vollständige Induktion oder auch durch ein einfaches graphisches Argument beweisen kann. Für das 2. Moment gilt

$$\sum_{\ell=1}^{n} \ell^2 \cdot P_\ell = \frac{1}{n}\sum_{\ell=1}^{n} \ell^2 = \frac{1}{n}\,\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$

Auch die hier verwendete Gleichung kann man graphisch oder durch Induktion beweisen. Für die Varianz folgt damit

$$V = \frac{1}{6}(n+1)(2n+1) - \frac{1}{4}(n+1)^2 = \frac{n^2-1}{12}. \qquad \square$$

1.4 Der Begriff des Wahrscheinlichkeitsraums

Beim klassischen Würfel ist der Grundraum gegeben durch $\Omega = \{1,2,3,4,5,6\}$. Der Ereignisraum ist

$$\mathcal{P}(\Omega) = \big\{\varnothing, \{1\}, \{2\}, \ldots, \{6\}, \{1,2\}, \ldots, \{1,2,3,4,5,6\}\big\}.$$

Er hat 2^6 Elemente, die wir auch **Ereignisse** nennen. Dabei deuten wir das Ereignis $\varnothing$ als den unmöglichen Wurf, das Ereignis $\{\ell\}$ als Elementarwurf, $1 \le \ell \le 6$, das Ereignis $\{1,2\}$ als Wurf von 1 oder 2 usw. Das Ereignis $\{1,2,\ldots,6\}$ deuten wir als den "sicheren" Wurf. Überlegen Sie sich, dass dies alle "Ereignisse" sind, die man beim Würfeln überhaupt beobachten kann!

Jedes Ereignis $A \in \mathcal{P}(\Omega)$ hat eine Wahrscheinlichkeit $\mathbb{P}(A)$. Diese Wahrscheinlichkeiten haben die folgenden (offensichtlichen) Eigenschaften:

$$\begin{aligned}
&0 \le \mathbb{P}(A) \le 1 \quad \text{für alle } A \in \mathcal{P}(\Omega),\\
&\mathbb{P}(\varnothing) = 0, \quad \mathbb{P}(\Omega) = 1,\\
&A \cap B = \varnothing \;\Rightarrow\; \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B),\\
&\bar{A} = \Omega \setminus A \;\Rightarrow\; \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A).\\
&\qquad \text{("Komplementärereignis")}
\end{aligned} \tag{A}$$

Dies motiviert die folgende Definition.

Definition 1.6 (Wahrscheinlichkeitsraum)**.** Ist ein endlicher Grundraum Ω gegeben, zusammen mit einer Abbildung $\mathbb{P} : \mathcal{P}(\Omega) \to [0, 1]$, die (A) erfüllt, so heißt das Tripel $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein **Wahrscheinlichkeitsraum**. Ein endlicher Wahrscheinlichkeitsraum mit gleichen Wahrscheinlichkeiten für alle Elementarereignisse heißt auch **Laplace-Raum**.

Beispiel. Der Münzwurf (Abschnitt 1.1) und der Würfel (Abschnitt 1.3) mit jeweils endlichem Grundraum. Beachten Sie, dass wir dort stillschweigend benutzt haben, dass die Wahrscheinlichkeit eines Ereignisses die Summe der Wahrscheinlichkeiten seiner Elementarereignisse ist, was eine direkte Folge aus (A) ist! Überzeugen Sie sich von der Gültigkeit von (A) in diesen Beispielen!

Wir wollen den Begriff des Wahrscheinlichkeitsraums auch für unendliche Grundräume erklären. Im Fall endlicher Grundräume folgt aus (A) die Eigenschaft

$$A_i \cap A_j = \varnothing \quad (i \neq j) \quad \Longrightarrow \quad \mathbb{P}\Big(\bigcup_i A_i\Big) = \sum_i \mathbb{P}(A_i), \tag{1.12}$$

also für endlich viele (eventuell leere) Ereignisse A_i, welche man σ-Additivität nennt. Im Fall unendlicher Grundräume hat diese Eigenschaft eine neue Bedeutung, weil die Vereinigung nun auch aus *unendlich* vielen nichtleeren Mengen bestehen kann. Wir fordern in diesem Fall zusätzlich zu (A) die Eigenschaft der σ-Additivität und bezeichnen das System aller dieser Forderungen wieder mit (A). Falls Ω abzählbar ist, besteht ein Wahrscheinlichkeitsraum nun wieder aus dem Tripel $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ wie in der obigen Definition. Zwei Beispiele mit abzählbar unendlichem Grundraum tauchen im folgenden Abschnitt auf. Endliche oder abzählbar unendliche Grundräume nennt man auch diskrete Grundräume.

Im Fall nicht-diskreter Grundräume Ω, die weiter unten eine große Rolle spielen werden, wählt man für die Definition einer Wahrscheinlichkeitsfunktion gemäß (A) *nicht* mehr die Potenzmenge $\mathcal{P}(\Omega)$, weil so eine Wahrscheinlichkeitsfunktion dann “pathologische” Eigenschaften hätte. Als Ausweg definiert man in diesem Fall Wahrscheinlichkeitsfunktionen gemäß (A) auf speziellen Teilmengen $\Sigma(\Omega)$ der Potenzmenge, sogenannten σ-Algebren, für die solche Pathologien nicht auftreten. Sie enthalten üblicherweise aber alle Ereignisse, an denen man interessiert ist. Für $\Omega = \mathbb{R}^d$ wählt man die sogenannte **Borel'sche σ-Algebra** $\Sigma(\mathbb{R}^d) = \mathcal{B}_d$, deren Elemente man **Borel-Mengen** nennt. Alle offenen und alle abgeschlossenen Teilmengen von $\mathbb{R}^d$ sind Borel-Mengen, und gleichfalls deren abzählbare Vereinigungen sowie Komplemente davon. Es gibt Teilmengen von $\mathbb{R}^d$, die keine Borel-Mengen sind, allerdings sind solche Mengen ziemlich kompliziert und haben “seltsame” Eigenschaften. Die Ereignisse, welche wir im kontinuierlichen Fall betrachten, werden meist Intervalle sein. Für einen gegebenen Grundraum Ω ist auch seine Potenzmenge $\mathcal{P}(\Omega)$ eine σ-Algebra, welche wir stets im Fall eines diskreten Grundraums wählen. (Für Details sei auf die Lehrbücher von Krengel und Georgii verwiesen.)

1.5 Weitere diskrete Verteilungen

1.5.1 Geometrische Verteilung

Wir betrachten den Münzwurf mit gezinkter Münze, $0 < p < 1$, welcher auch **Bernoulli-Experiment** genannt wird. Für die Wahrscheinlichkeit, dass "Kopf" genau beim n-ten Versuch zum ersten Mal auftritt, gilt offenbar

$$P_n := \mathbb{P}(\text{“1” zum ersten Mal nach } n \text{ Würfen}) = (1-p)^{n-1}\, p \qquad (n \in \mathbb{N}).$$

Wir bestimmen einige wichtige Kenngrößen dieser Verteilung. Dabei tritt immer wieder die geometrische Reihe auf, die aus der Analysis bekannt ist.

Normierung: Da $1 - p < 1$ gilt, bekommen wir

$$\sum_{n=1}^{\infty}(1-p)^{n-1}p = p\sum_{m=0}^{\infty}(1-p)^m = p \cdot \frac{1}{1-(1-p)} = 1. \tag{1.13}$$

Erwartungswert: Wir verwenden einen "Ableitungstrick". Es gilt

$$\begin{aligned}
\sum_{n=1}^{\infty} np\,(1-p)^{n-1} &= p\sum_{n\geq 1} n\, x^{n-1}\Big|_{x=1-p} \\
&= p\sum_{n\geq 1}\frac{\mathrm{d}}{\mathrm{d}x}\,x^n\Big|_{x=1-p} \overset{!}{=} p\cdot\frac{\mathrm{d}}{\mathrm{d}x}\sum_{n\geq 0} x^n\Big|_{x=1-p} \\
&= p\cdot\frac{\mathrm{d}}{\mathrm{d}x}\,\frac{1}{1-x}\Big|_{x=1-p} = p\cdot\frac{1}{(1-x)^2}\Big|_{x=1-p} \\
&= \frac{1}{p}.
\end{aligned} \tag{1.14}$$

Hier wurde die absolute Konvergenz der geometrischen Reihe verwendet, man darf dann Summation und Ableitung vertauschen.

2. Moment:

$$\begin{aligned}
\sum_{n\geq 1} n^2p\,(1-p)^{n-1} &= p\sum_{n\geq 1} n\,\frac{\mathrm{d}}{\mathrm{d}x}\,x^n\Big|_{x=1-p} \\
&= p\sum_{n\geq 1}\frac{\mathrm{d}}{\mathrm{d}x}\left(x\cdot\frac{\mathrm{d}}{\mathrm{d}x}\,x^n\right)\Big|_{x=1-p} && (1.15)\\
&\overset{!}{=} p\cdot\frac{\mathrm{d}}{\mathrm{d}x}\left(x\cdot\frac{\mathrm{d}}{\mathrm{d}x}\,\frac{1}{1-x}\right)\Big|_{x=1-p} \\
&= p\cdot\frac{\mathrm{d}}{\mathrm{d}x}\,\frac{x}{(1-x)^2}\Big|_{x=1-p} = p\cdot\frac{1+x}{(1-x)^3}\Big|_{x=1-p} \\
&= \frac{2-p}{p^2}. && (1.16)
\end{aligned}$$

Varianz und Standardabweichung:

$$V = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}, \quad \sigma = \frac{\sqrt{1-p}}{p}. \tag{1.17}$$

Definition und Satz 1.7 (geometrische Verteilung)**.** Die **geometrische Verteilung** mit Parameter[2] $0 < p < 1$ ist durch die Wahrscheinlichkeiten

$$P_n = (1-p)^{n-1}p \qquad (n \in \mathbb{N}) \tag{1.18}$$

gegeben. Sie besitzt den Erwartungswert $\frac{1}{p}$ und die Varianz $\frac{1-p}{p^2}$. □

Beispiel. Wann erscheint beim idealen Würfel zuerst eine “6”? Wir haben

$$p = \frac{1}{6}, \qquad \underline{m} = 6, \qquad V = 30.$$

Intuitiv würde man vielleicht einen kleineren Mittelwert erwarten (wenn man sich von der großen Varianz täuschen läßt)!

1.5.2 Poisson-Verteilung

Beim n-maligen gezinkten Münzwurf wird die Wahrscheinlichkeit des Ereignisses “ℓ-mal 1” durch die Binomialverteilung $\mathcal{B}(n,p)$ beschrieben. Im Fall n “riesig” und p “winzig” sind die entsprechenden Wahrscheinlichkeiten selbst mit dem Computer kaum berechenbar. Allerdings ergibt sich im Grenzfall $n \to \infty$ und $p \to 0$ mit $\lambda := np \geq 0$ fest eine einfache Formel. Dann gilt nämlich

$$P_\ell := \mathbb{P}(\text{“}\ell\text{-mal 1”}) = \frac{\lambda^\ell}{\ell!}\, \mathrm{e}^{-\lambda} \qquad (\ell \geq 0). \tag{1.19}$$

Beweis. Mit $\lambda = np \geq 0$ gilt für alle $\ell \in \mathbb{N}_0$ fest

$$\begin{aligned}
\binom{n}{\ell} p^\ell (1-p)^{n-\ell} &= \frac{n \cdot (n-1) \cdot \ldots \cdot (n-\ell+1)}{1 \cdot 2 \cdot \ldots \cdot \ell}\, p^\ell\, (1-p)^n\, \frac{1}{(1-p)^\ell} \\
&= \frac{(np)^\ell}{\ell!} \left(1 - \frac{np}{n}\right)^n \frac{1 \cdot (1-\frac{1}{n}) \cdot \ldots \cdot (1 - \frac{\ell-1}{n})}{(1-p)\cdot(1-p)\cdot \ldots \cdot (1-p)} \\
&= \frac{\lambda^\ell}{\ell!} \left(1 - \frac{\lambda}{n}\right)^n \frac{1 \cdot (1-\frac{1}{n}) \cdot \ldots \cdot (1 - \frac{\ell-1}{n})}{(1-p)\cdot(1-p)\cdot \ldots \cdot (1-p)}.
\end{aligned}$$

Beachten Sie die Konvention $0^0 = 1$. Jetzt betrachten wir den gemeinsamen Grenzwert $p \to 0$ und $n \to \infty$ mit $\lambda = np$ fest. Wegen $\lim_{n\to\infty}\bigl(1 - \frac{\lambda}{n}\bigr)^n = \mathrm{e}^{-\lambda}$ gilt dann

$$\binom{n}{\ell} p^\ell (1-p)^{n-\ell} \longrightarrow \frac{\lambda^\ell}{\ell!}\, \mathrm{e}^{-\lambda}.$$ □

[2]Man überlege sich, was der Grenzfall $p = 1$ bedeutet.

Der obige Beweis zeigt auch, dass man die Binomialverteilung $\mathcal{B}(n,p)$ für große n und kleine p annähern kann durch

$$P_\ell \simeq \frac{(np)^\ell}{\ell!}\,\mathrm{e}^{-np}. \tag{1.20}$$

Als Faustregel für eine gute Approximation gilt $p \leq 0.1$ und $n \geq 100$.
Wir bestimmen einige Kenngrößen dieser Verteilung. Dabei tritt immer wieder die Exponentialreihe aus der Analysis auf.

Normierung:

$$\sum_{n\geq 0} \frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda} \;=\; \mathrm{e}^{-\lambda}\sum_{n\geq 0}\frac{1}{n!}\,\lambda^n \;=\; \mathrm{e}^{-\lambda}\,\mathrm{e}^{\lambda} \;=\; 1. \tag{1.21}$$

Erwartungswert:

$$\underline{m} = \sum_{n\geq 0} n\,\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda} \;=\; \lambda\,\mathrm{e}^{-\lambda}\sum_{m=1}^{\infty}\frac{1}{(m-1)!}\,\lambda^{m-1} \;=\; \lambda. \tag{1.22}$$

2. Moment: Berechnung z.B. mit dem Ableitungstrick.

$$\begin{aligned}
\sum_{n\geq 0} n^2\frac{\lambda^n}{n!}\,\mathrm{e}^{-\lambda} \;&=\; \lambda\,\mathrm{e}^{-\lambda}\sum_{n=1}^{\infty} n\cdot\frac{\lambda^{n-1}}{(n-1)!}\\
&=\; \lambda\,\mathrm{e}^{-\lambda}\sum_{n\geq 1}\frac{\mathrm{d}}{\mathrm{d}\lambda}\,\frac{\lambda^n}{(n-1)!} \;=\; \lambda\,\mathrm{e}^{-\lambda}\frac{\mathrm{d}}{\mathrm{d}\lambda}\Big(\lambda\sum_{n\geq 1}\frac{\lambda^{n-1}}{(n-1)!}\Big)\\
&=\; \lambda\,\mathrm{e}^{-\lambda}\frac{\mathrm{d}}{\mathrm{d}\lambda}\big(\lambda\,\mathrm{e}^{\lambda}\big) \;=\; \lambda\,\mathrm{e}^{-\lambda}\big(\mathrm{e}^{\lambda}+\lambda\,\mathrm{e}^{\lambda}\big)\\
&=\; \lambda+\lambda^2.
\end{aligned} \tag{1.23}$$

Varianz:

$$V = \lambda+\lambda^2-\lambda^2 = \lambda. \tag{1.24}$$

Definition und Satz 1.8 (Poisson-Verteilung)**.** Die **Poisson-Verteilung** mit Parameter $\lambda > 0$ ist gegeben durch die Wahrscheinlichkeiten

$$P_n \;=\; \mathrm{e}^{-\lambda}\,\frac{\lambda^n}{n!} \qquad (\text{für } n \in \mathbb{N}_0). \tag{1.25}$$

Sie besitzt Erwartungswert λ und Varianz λ. □

Beispiel. Die Anzahl von Regentropfen pro Pflasterstein bei Regenbeginn wird näherungsweise durch eine Poisson-Verteilung beschrieben: Wir betrachten ein Pflaster mit N Steinen, auf das n Regentropfen fallen. Dann ist $\lambda := n/N$ die Regendichte, also die (mittlere) Regenmenge pro Stein. Die Wahrscheinlichkeit, dass ein vorgegebener Stein von einem Regentropfen getroffen wird, sei $p = 1/N$. Die Wahrscheinlichkeit, dass genau ℓ Tropfen auf einen vorgegebenen Stein auftreffen, sei binomial mit Parametern n und p verteilt. (Überlegen Sie sich, weshalb diese Annahmen vernünftig sind!) Im Grenzfall eines unendlich großen Pflasters ($N \to \infty$ bzw. $p \to 0$) erhalten wir damit bei konstanter Regendichte λ (also insbesondere für $n \to \infty$) die Poisson-Verteilung mit Parameter λ.

Versuchen Sie, das näherungsweise Auftreten einer Poisson-Verteilung in den folgenden Beispielen auf ähnliche Weise zu begründen.

- Anzahl Rosinen pro Volumeneinheit englischen Kuchens
- Anzahl Pferdehufschlagtote in preußischen Kavallerieregimenten
- Zahl der Bakterienkolonien pro Fläche in einer Petri-Schale
- Zahl der Tore pro Spiel in der 1. Fußball-Bundesliga
- radioaktive Zerfallsereignisse pro Zeiteinheit
- Schadensmeldungen bei einer Kfz-Haftpflichtversicherung pro Zeitintervall

1.5.3 Hypergeometrische Verteilung

Die hypergeometrische Verteilung tritt bei der Entnahme einer sog. **Stichprobe ohne Zurücklegen** auf. Gegeben ist eine Urne mit W weißen Kugeln und S schwarzen Kugeln, aus der ohne Zurücklegen gezogen wird. Beim ersten Zug ist die Wahrscheinlichkeit, eine weiße Kugel zu ziehen, gegeben durch $p = W/(W + S)$. Es werden insgesamt n Kugeln gezogen. Was ist dann die Wahrscheinlichkeit, k-mal eine weiße Kugel zu ziehen? Wir betrachten nur den interessanten Fall $0 \leq k \leq \min(n, W)$. Offenbar gibt es

$\binom{W+S}{n}$ gleichwahrscheinliche Möglichkeiten, n Kugeln zu ziehen,

$\binom{W}{k}$ Möglichkeiten, k von den weißen Kugeln zu ziehen,

$\binom{S}{\ell}$ Möglichkeiten, ℓ von den schwarzen Kugeln zu ziehen, mit $\ell = n - k$.

Damit erhalten wir für die gesuchte Wahrscheinlichkeit

$$P_k := \mathbb{P}(\text{“}k\text{-mal weiß”}) = \frac{\binom{W}{k}\binom{S}{n-k}}{\binom{W+S}{n}}. \tag{1.26}$$

Diese Formel kann man auf eine für Rechnungen vorteilhaftere Form bringen:

$$\begin{aligned} P_k &= \frac{W!}{k!(W-k)!} \cdot \frac{S!}{(n-k)!(S-(n-k))!} \cdot \frac{n!(W+S-n)!}{(W+S)!} \\ &= \binom{n}{k}\binom{W+S-n}{W-k}\binom{W+S}{W}^{-1}. \end{aligned} \tag{1.27}$$

Normierung: Wir verwenden die Identität

$$\binom{\alpha}{0}\binom{\beta}{k} + \binom{\alpha}{1}\binom{\beta}{k-1} + \ldots + \binom{\alpha}{k}\binom{\beta}{0} = \binom{\alpha+\beta}{k}, \tag{1.28}$$

welche für $\alpha, \beta \in \mathbb{R}$ und $k \in \mathbb{N}_0$ gültig ist. (Man erhält sie übrigens recht einfach aus $(1+x)^\alpha(1+x)^\beta = (1+x)^{\alpha+\beta}$ unter Verwendung des binomischen Lehrsatzes durch Koeffizientenvergleich.) Dann folgt die Normierung aus

$$\sum_{k=0}^{W} \binom{n}{k}\binom{W+S-n}{W-k} = \binom{W+S}{W}. \tag{1.29}$$

Definition und Satz 1.9 (hypergeometrische Verteilung)**.** Die **hypergeometrische Verteilung** mit den Parametern $W, S, n \in \mathbb{N}$ mit der Wahrscheinlichkeitsverteilung

$$P_k = \frac{\binom{W}{k}\binom{S}{n-k}}{\binom{W+S}{n}} \qquad (0 \le k \le n) \tag{1.30}$$

hat Erwartungswert und Varianz

$$\underline{m} = np\,, \quad V = np(1-p)\,\frac{W+S-n}{W+S-1}, \tag{1.31}$$

mit $p = W/(W+S)$. □

Den Beweis kann man wieder mit Identitäten für Binomialkoeffizienten führen, siehe auch das Lehrbuch von Krengel. Beachten Sie, dass der Erwartungswert und die Varianz denen der Binomialverteilung ähneln. Tatsächlich wird die Binomialverteilung beim **Ziehen mit Zurücklegen** erhalten, wie Sie sich leicht klarmachen können. Warum stimmen die Erwartungswerte sogar *exakt* überein?

Beispiel. Wir fragen nach der Wahrscheinlichkeit für vier Richtige beim Lotto 6 aus 49. Dazu malen wir von den 49 Kugeln die sechs getippten weiß an und die restlichen schwarz. Die Wahrscheinlichkeit, bei sechsmaligem Ziehen ohne Zurücklegen vier weiße Kugeln zu ziehen, ist gegeben durch

$$\mathbb{P}(\text{“4 Richtige”}) = \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} = \frac{15 \cdot 903}{13\,983\,816} \approx 0.000\,097. \tag{1.32}$$

1.6 Diskrete und kontinuierliche Zufallsvariablen

Wir betrachten einen allgemeinen Wahrscheinlichkeitsraum $(\Omega, \Sigma(\Omega), \mathbb{P})$, vergleiche 1.4. Da Ω auch recht “abstrakt” sein kann (Farbe, Kopf), benötigt man noch ein Konzept zur Quantifizierung.

Definition 1.10 (Zufallsvariable, ZV)**.** Unter einer **ZV** versteht man eine (messbare) Abbildung $X : \Omega \longrightarrow \mathbb{R}$. Die Werte, die X annehmen kann, heißen ihre **Realisierungen**. Je nach Bildbereich nennt man X eine **diskrete** oder **kontinuierliche** ZV.

Bemerkung. Aus Gründen mathematischer Konsistenz fordern wir, dass für alle Borel-Mengen $B \subset \mathbb{R}$ deren Urbilder $X^{-1}(B)$ unter der Zufallsvariablen X Ereignisse in $\Sigma(\Omega)$ sind. Diese Eigenschaft nennt man Messbarkeit. Sie ist für diskrete Grundräume Ω (dann ist X diskret) für beliebige Funktionen X nach Konstruktion erfüllt, weil wir in diesem Fall stets die Potenzmenge als Ereignisraum wählen, $\Sigma(\Omega) = \mathcal{P}(\Omega)$. Im Fall überabzählbarer Grundräume muss die obige Forderung für “komplizierte” Funktionen $X : \Omega \to \mathbb{R}$ nicht erfüllt sein. Wir werden es hier immer mit messbaren kontinuierlichen Zufallsvariablen zu tun haben, allerdings werden wir uns den Nachweis der Messbarkeit jeweils ersparen. (Details dazu findet man in den Lehrbüchern von Krengel oder Georgii.)

Diskrete Zufallsvariablen haben wir in den obigen Beispielen schon immer implizit verwendet.

Beispiel. Der einfache Münzwurf mit $\Omega = \{\text{Kopf, Zahl}\}$, $X : \Omega \to \{0, 1\} \subset \mathbb{R}$. Die Zufallsvariable X ist also eine *Funktion* mit $X(\text{Kopf}) = 1$ und $X(\text{Zahl}) = 0$. Weil der Münzwurf aber ein Zufallsexperiment ist, mit Elementarereignis ω, ist auch die Realisierung $x = X(\omega)$ zufällig.

In diesem Abschnitt untersuchen wir kontinuierliche Zufallsvariablen. Dabei wird nun die sogenannte Verteilungsfunktion einer Zufallsvariablen eine wichtige Rolle spielen. Wir schreiben im folgenden oft X für eine ZV, und x für eine ihrer Realisierungen. Für einen diskreten (!) Grundraum definieren wir in Anlehnung an das obige Beispiel die Wahrscheinlichkeit P_i für die Realisierung x_i der Zufallsvariable X durch

$$P_i := \mathbb{P}(X = x_i) := \sum_{\omega_j \in \Omega : X(\omega_j) = x_i} \mathbb{P}(\omega_j). \tag{1.33}$$

Analog definieren wir Wahrscheinlichkeiten wie $\mathbb{P}(X \leq x)$ oder $\mathbb{P}(X > x)$, also haben wir beispielsweise

$$\mathbb{P}(X \leq x) := \sum_{\omega_j \in \Omega : X(\omega_j) \leq x} \mathbb{P}(\omega_j) = \sum_{y_i \in X(\Omega) : y_i \leq x} \mathbb{P}(X = y_i). \tag{1.34}$$

Definition 1.11 (Verteilungsfunktion)**.** Sei X eine ZV. Dann bezeichnet man mit $F(x) := \mathbb{P}(X \leq x)$ die zugehörige **Verteilungsfunktion**.

Beispiel. Der einfache, faire Münzwurf mit X als Ergebnis des Wurfes, also mit $X(\Omega) = \{0, 1\} \subset \mathbb{R}$. Die Verteilungsfunktion ist nicht stetig (sie hat Sprungstellen), jedoch rechtsseitig stetig.

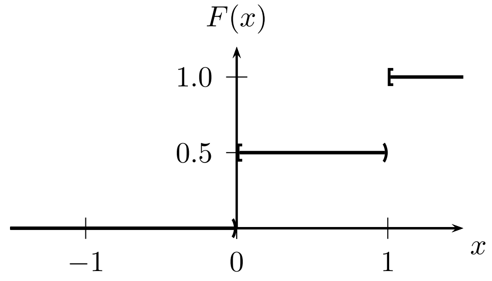

Abbildung 1.2: Verteilungsfunktion des fairen Münzwurfes

Beachten Sie, dass in der obigen Definition $\mathbb{P}(X \leq x)$ gemäß Gleichung (1.34) zunächst nur für den Fall eines diskreten Grundraums erklärt ist. Machen Sie sich klar, dass in diesem Fall die Verteilungsfunktion zur Beschreibung einer ZV ebenso gut geeignet ist wie die Kenntnis der Wahrscheinlichkeiten aller Elementarereignisse. Die Verteilungsfunktion hat die folgenden offensichtlichen Eigenschaften:

1. $0 \leq F(x) \leq 1, \quad \text{mit } \lim F(x) = \begin{cases} 1, & \text{für } x \to +\infty, \\ 0 & \text{für } x \to -\infty. \end{cases}$
2. $x \leq y \Rightarrow F(x) \leq F(y)$ (Monotonie) .

Im kontinuierlichen Fall ist eine Definition von $\mathbb{P}(X \leq x)$ gemäß Gleichung (1.34) nicht sinnvoll, weil wir es dann nicht mit abzählbaren Summen zu tun hätten. Wir werden im folgenden zunächst “naiv” Beispiele für Verteilungsfunktionen kontinuierlicher Zufallsvariablen betrachten. Damit werden wir dann $\mathbb{P}(X \leq x)$ für bestimmte kontinuierliche Zufallsvariablen definieren.

Wir haben im obigen Beispiel gesehen, dass für diskrete ZV deren Verteilungsfunktion $F(x)$ mit Sprungstellen versehen ist. Für kontinuierliche ZV hingegen sieht eine Verteilungsfunktion vielleicht so wie in Abbildung 1.3 oder 1.4 aus.

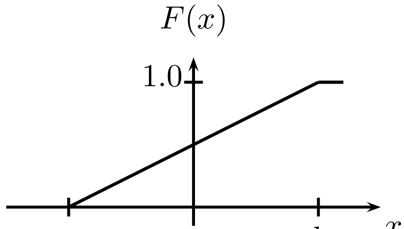

Abbildung 1.3: Gleichverteilung auf $[a, b]$

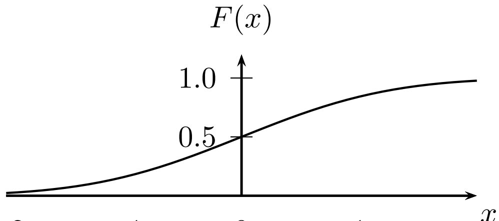

Abbildung 1.4: Eine allgemeine Verteilung auf ganz $\mathbb{R}$

Diese Beobachtung legt es nahe, die Verteilungsfunktion auf folgende Art zu beschreiben:

$$F(x) = \int_{-\infty}^{x} f(\tilde{x})\,\mathrm{d}\tilde{x}. \tag{1.35}$$

Die Verteilungsfunktion $F(x)$ ist hier als Stammfunktion dargestellt, es gilt also auch $f(x) = F'(x)$. Solch eine Darstellung gibt es, falls $F(x)$ im relevanten Intervall differenzierbar ist. Die Funktion $f(x)$ nennt man die **Dichtefunktion**. Hiermit möchte man jetzt zum Beispiel die Wahrscheinlichkeit für das Ereignis angeben, dass die zugehörige Zufallsvariable Werte im Intervall $[\alpha, \beta]$ annimmt. Wir haben

$$\mathbb{P}\big(X \in [\alpha, \beta]\big) \overset{!}{=} F(\beta) - F(\alpha) = \int_a^b f(y)\,\mathrm{d}y, \tag{1.36}$$

wobei wir hier eine Rechenregel für Wahrscheinlichkeiten aus dem diskreten Fall übertragen haben. Die Wahrscheinlichkeit, dass X Werte im Intervall $[\alpha, \beta]$ annimmt, ist also die Fläche zwischen der x-Achse und dem Graphen der Dichtefunktion in diesem Intervall! Für kleine Intervalle folgt für stetige Dichtefunktionen

$$\mathbb{P}\big(X \in [x, x + \mathrm{d}x]\big) = \int_x^{x+\mathrm{d}x} f(y)\,\mathrm{d}y \sim f(x)\,\mathrm{d}x \qquad (\mathrm{d}x \to 0), \tag{1.37}$$

nach dem aus der Analysis bekannten Mittelwertsatz der Integralrechnung. Insbesondere gilt dann $\mathbb{P}(X = x) = 0$. Also kann man im Unterschied zum diskreten Fall eine kontinuierliche Zufallsvariable X nicht durch die Angabe der Wahrscheinlichkeiten $\mathbb{P}(X = x)$ charakterisieren, wohl aber durch die Angabe der Funktion $f(x)$, falls eine solche existiert. Die obige asymptotische Gleichheit rechtfertigt die Bezeichung von $f(x)$ als Wahrscheinlichkeitsdichte. Für geeignete stetige (oder stückweise stetige) Funktionen $f(x)$ definieren wir nun

$$\mathbb{P}(X \le x) := \int_{-\infty}^{x} f(y)\,\mathrm{d}y \tag{1.38}$$

als die kontinuierliche Variante von $\mathbb{P}(X \le x)$ gemäß Gleichung (1.34) im diskreten Fall. Beachten Sie die formale Analogie beider Ausdrücke: Beim Übergang vom diskreten zum kontinuierlichen Fall wird die Summe über i durch ein Integral ersetzt, und $\mathbb{P}(X = y_i)$ wird durch $f(y)\,\mathrm{d}y$ ersetzt. Diese Analogie werden wir später wiederholt verwenden.

Beispiel. Die Gleichverteilung auf $[a, b]$ ist definiert über die folgende Dichtefunktion $f(x)$ oder über die folgende Verteilungsfunktion $F(x)$:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } a \le x < b, \\ 0, & \text{sonst,} \end{cases} \qquad F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \le x < b, \\ 1, & b \le x. \end{cases} \tag{1.39}$$

Die Verteilungsfunktion der Gleichverteilung ist in Abbildung 1.3 skizziert, deren Dichte in Abbildung 1.5.

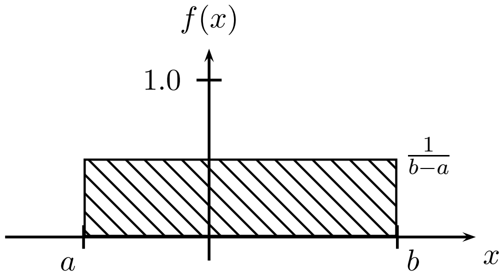

Abbildung 1.5: Dichtefunktion der kontinuierlichen Gleichverteilung auf $[a, b]$

1.7 Erwartungswert und Varianz

Definition 1.12 (Erwartungswert, Varianz)**.** Ist X eine diskrete ZV, so heißt

$$\underline{m} := \mathbb{E}(X) := \sum_i x_i \, \mathbb{P}(X = x_i) \tag{1.40}$$

ihr **Erwartungswert**. Ist X kontinuierlich und besitzt eine Dichte f, setzen wir für eine allgemeine Funktion g

$$\mathbb{E}(g(X)) := \int_{-\infty}^{\infty} g(x) \, f(x) \, \mathrm{d}x. \tag{1.41}$$

Insbesondere heißt dann

$$\underline{m} := \mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) \, \mathrm{d}x \tag{1.42}$$

der **Erwartungswert** von X. Weiter heißt $V := \mathbb{V}(X) := \mathbb{E}\big((X - \underline{m})^2\big)$ die **Varianz** von X, und $\sigma := \sqrt{V}$ die **Standardabweichung** von X. Wir fordern, dass die entsprechenden Summen oder Integrale absolut konvergieren. Dann sind die obigen Größen wohldefiniert und (endliche) reelle Zahlen.

Bemerkung. Absolute Konvergenz bedeutet $\sum_i |x_i| \, \mathbb{P}(X = x_i) < \infty$ für den Erwartungswert einer diskreten ZV oder $\int_{-\infty}^{\infty} |g(x)| \, f(x) \, \mathrm{d}x < \infty$ für Erwartungswerte $\mathbb{E}(g(X))$ für eine kontinuierliche ZV mit Dichte $f(x)$. Der Erwartungswert $\mathbb{E}(g(X))$ für diskrete ZV ist analog erklärt. Überlegen Sie sich diese Definition! Beispiele für Verteilungen ohne Erwartungswert gibt es auch, eines wird in den Übungen behandelt.