2 Information und Zufall

2.1 Wahrscheinlichkeit und Informationsmenge

Im Folgenden suchen wir eine Funktion I definiert auf dem Intervall $[0, 1]$, die jeder Wahrscheinlichkeit $p \in [0, 1]$ eine Informationsmenge $I(p)$ zuordnet; von dieser Funktion I werden gewisse Eigenschaften gefordert:

(I1) Die Funktion $I : [0, 1] \to [0, \infty]$ soll auf dem offenen Intervall $(0, 1)$ stetig sein.
Diese Forderung bedarf wohl keiner Erklärung. Niemand wird ernsthaft Unstetigkeiten fordern oder zulassen wollen.

(I2) $I\left(\frac{1}{2}\right) = 1$.
Man kann nur messen, wenn man eine Einheit festgelegt hat (wie zum Beispiel das Urmeter als Einheit der Längenmessung in Paris). Diese Forderung legt nun als Einheit die Informationsmenge Eins für die Wahrscheinlichkeit $\frac{1}{2}$ fest.

(I3) $I(pq) = I(p) + I(q)$ für alle $p, q \in (0, 1)$.
Tritt ein Ereignis A mit der Wahrscheinlichkeit p auf und tritt ein Ereignis B mit der Wahrscheinlichkeit q auf, so gelten diese Ereignisse als stochastisch unabhängig, wenn das gemeinsame Auftreten von A und B mit Wahrscheinlichkeit pq erfolgt. In diesem Fall beeinflusst das Auftreten von A nicht die Wahrscheinlichkeit für das Auftreten von B und umgekehrt. Es ist daher sinnvoll, die Informationsmenge, die das gemeinsame Auftreten von A und B beinhaltet, als Summe der einzelnen Informationsmengen (von A und von B) festzulegen.

(I4) $I(0) = \lim\limits_{\substack{p \to 0 \\ p \in (0,1)}} I(p)$, $I(1) = \lim\limits_{\substack{p \to 1 \\ p \in (0,1)}} I(p)$
Diese vierte Forderung ist wiederum ein Stetigkeitsargument.

Nun soll in einem ersten Resultat gezeigt werden, dass die Funktion I auf dem Intervall $(0, 1)$ durch die ersten drei Eigenschaften eindeutig festgelegt ist.

S. Schäffler, *Mathematik der Information*, Springer-Lehrbuch Masterclass,
DOI 10.1007/978-3-662-46382-6_2

Theorem 2.1 (Eindeutigkeit der Funktion I) *Es gibt genau eine Funktion*

$$h : (0,1) \to (0,\infty)$$

mit:

(i) h *ist stetig.*
(ii) $h\left(\frac{1}{2}\right) = 1.$
(iii) $h(pq) = h(p) + h(q)$ *für alle* $p, q \in (0,1)$.

Diese Funktion ist die Umkehrfunktion zu

$$f : (0,\infty) \to (0,1), \quad x \mapsto 2^{-x}$$

und damit der negative Logarithmus dualis auf dem Intervall $(0,1)$ *(bezeichnet mit:* $-\operatorname{ld}_{(0,1)}$*). Es gilt:*

$$\lim_{\substack{x \to 0 \\ x > 0}} x \cdot (-\operatorname{ld}_{(0,1)}(x)) = 0. \qquad \triangleleft$$

Beweis Seien $n, m \in \mathbb{N}$, so gilt für eine Funktion

$$h : (0,1) \to (0,\infty)$$

mit den Eigenschaften (i)–(iii):

$$h\left(2^{-n}\right) = h\left(\left(\frac{1}{2}\right)^n\right) = n \cdot h\left(\frac{1}{2}\right) = n.$$

Ferner erhalten wir aus

$$n = h\left(\left(\frac{1}{2}\right)^n\right) = h\left(\left(\left(\frac{1}{2}\right)^{\frac{n}{m}}\right)^m\right) = mh\left(\left(\frac{1}{2}\right)^{\frac{n}{m}}\right)$$

die Gleichung

$$h\left(2^{-\frac{n}{m}}\right) = h\left(\left(\frac{1}{2}\right)^{\frac{n}{m}}\right) = \frac{n}{m}.$$

Sei nun $y \in (0,1)$, so gibt es ein eindeutiges $x \in (0,\infty)$ mit $y = 2^{-x}$. Da $\mathbb{Q}$ dicht in $\mathbb{R}$ liegt, gibt es zwei Folgen $\{m_i\}_{i\in\mathbb{N}}$ und $\{n_i\}_{i\in\mathbb{N}}$ natürlicher Zahlen mit

$$\lim_{i\to\infty} \frac{n_i}{m_i} = x.$$

Aus der Stetigkeit von f und h folgt:

$$\begin{aligned} h\left(2^{-x}\right) &= h\left(2^{-\lim\limits_{i\to\infty}\frac{n_i}{m_i}}\right) = h\left(\lim_{i\to\infty} 2^{-\frac{n_i}{m_i}}\right) = \\ &= \lim_{i\to\infty} h\left(2^{-\frac{n_i}{m_i}}\right) = \lim_{i\to\infty} h\left(\left(\frac{1}{2}\right)^{\frac{n_i}{m_i}}\right) = \\ &= \lim_{i\to\infty} \frac{n_i}{m_i} = x. \end{aligned}$$

Es gilt also: $h = -\operatorname{ld}_{(0,1)}$.
Für $x > 0$ ist

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \sum_{k=3}^{\infty} \frac{x^k}{k!} > \frac{x^2}{2}.$$

Somit erhalten wir:

$$\begin{aligned} 0 \leq \lim_{\substack{x\to 0\\ x>0}} x \cdot (-\operatorname{ld}_{(0,1)}(x)) &= \lim_{y\to\infty} \frac{1}{y\ln(2)} \cdot \left(-\ln_{(0,1)}\left(\frac{1}{y}\right)\right) = \lim_{y\to\infty} \frac{\ln(y)}{y\ln(2)} = \\ = \lim_{y\to\infty} \frac{\ln(y)}{e^{\ln(y)}\ln(2)} &= \lim_{z\to\infty} \frac{z}{e^z\ln(2)} \leq \lim_{z\to\infty} \frac{z}{\frac{z^2}{2}\ln(2)} = \\ = 0. \end{aligned}$$

q.e.d.

Aus diesem Resultat folgt, dass unsere gesuchte Funktion I auf dem Intervall $(0, 1)$ durch die Funktion $-\operatorname{ld}_{(0,1)}$ festgelegt ist. Da

$$\begin{aligned} &\lim_{\substack{p\to 1\\ p\in(0,1)}} (-\operatorname{ld}_{(0,1)}(p)) = -\operatorname{ld}(1) = 0 \quad \text{und} \\ &\lim_{\substack{p\to 0\\ p\in(0,1)}} (-\operatorname{ld}_{(0,1)}(p)) = \lim_{\substack{p\to 0\\ p\in(0,1)}} (-\operatorname{ld}(p)) = \infty, \end{aligned}$$

folgt (siehe Abb. 2.1):

$$I(0) = \infty \quad \text{und} \quad I(1) = 0.$$

Mit der in der Maßtheorie üblichen Festlegung

$$\infty + a = a + \infty = \infty \quad \text{für alle} \quad a \in \mathbb{R} \cup \{\infty\}$$

gilt sogar

$$I(pq) = I(p) + I(q) \quad \text{für alle} \quad p, q \in [0, 1].$$

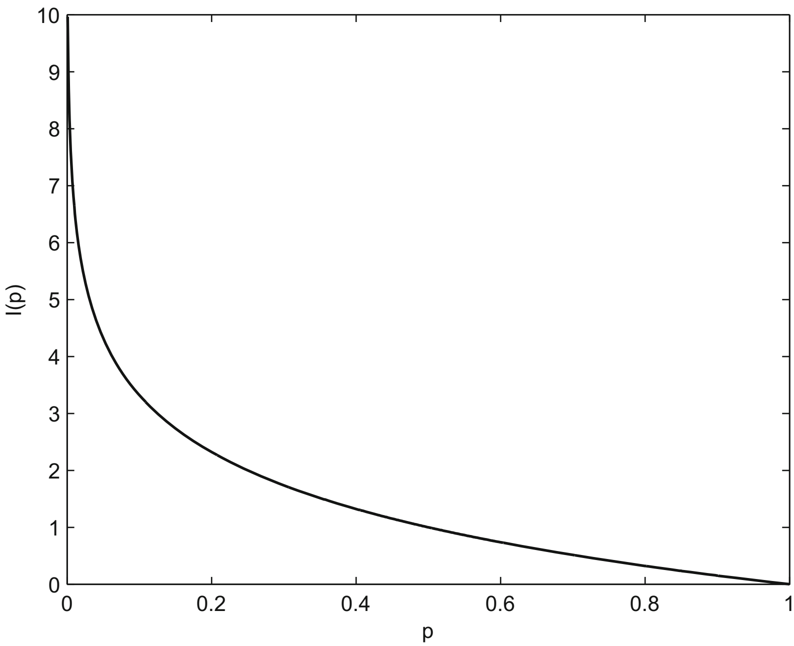

Abb. 2.1 Die Funktion I

Die Informationsmenge besitzt auch eine Einheit; sie wird in **bit** gemessen. Diese Wahl ist naheliegend, wenn man sich folgende Spezialfälle betrachtet, wobei der Index „b" bedeutet, dass das Binärsystem zugrunde gelegt ist:

$$\begin{aligned} I\left(\frac{1}{2}\right) &= I(0.1_b) = 1 \text{ bit}, \\ I\left(\left(\frac{1}{2}\right)^k\right) &= I(0.\underbrace{0\ldots01_b}_{k \text{ Stellen}}) = k \text{ bit}, \\ 3 \text{ bit } &= I\left(\frac{1}{8}\right) = I(0.001_b) \leq \\ &\leq I(0.\overline{0001}_b) = I\left(\frac{1}{15}\right) = 3.9069\ldots \text{bit} < \\ &< I(0.0001_b) = 4 \text{ bit}. \end{aligned}$$

Die Informationsmenge einer Zahl $p \in (0, 1]$ kann also mit

$$\lfloor p \rfloor := \min\left\{k \in \mathbb{N}_0;\ \left(\frac{1}{2}\right)^k \geq p\right\}$$

durch

$$\lfloor p \rfloor \leq I(p) < \lfloor p \rfloor + 1$$

abgeschätzt werden.

Da sich die Funktionen $-\mathrm{ld}_{(0,1)}$ und $-\mathrm{ld}$ auf dem Intervall $(0, 1)$ nicht unterscheiden, verwenden wir im Folgenden nur noch die Funktion $-\mathrm{ld}$ bzw. ld. Hätten wir in Forderung [I2] für $\zeta > 1$ statt $I\left(\frac{1}{2}\right) = 1$ die Forderung $I\left(\frac{1}{\zeta}\right) = 1$ aufgestellt, so hätten wir als Ergebnis statt dem Logarithmus dualis den Logarithmus zur Basis ζ erhalten.

Um uns vom Begriff **Informationsmenge** gegeben durch die Funktion I eine Vorstellung machen zu können, betrachten wir folgendes

Beispiel 2.2 Am 31. Mai 2010 erhalten zwei Personen, A und B, von einer dritten Person – nennen wir sie C – die Nachricht, dass heute Bundespräsident Horst Köhler zurückgetreten ist. Person A wusste das bereits, während Person B nichts wusste und den Rücktritt eines Bundespräsidenten für unmöglich hielt. Ein und diesselbe Nachricht beinhaltet somit für die beiden Personen A und B völlig unterschiedliche Mengen an Information. Für Person A war die Wahrscheinlichkeit p_A, dass Horst Köhler zurücktritt, in dem Moment, als sie die Nachricht von Person C erhält, gleich Eins, denn sie kannte den Inhalt der Nachricht bereits. Somit war die Nachricht mit keinerlei Information verbunden:

$$I(p_A) = I(1) = -\mathrm{ld}(1) = 0.$$

Für Person B war die Überraschung unendlich groß, da sie diesen Rücktritt für unmöglich hielt ($p_B = 0$):

$$I(p_B) = I(0) = \lim_{\substack{x \to 0 \\ x > 0}} -\mathrm{ld}(x) = \infty.$$

Person C hatte eine weitere Nachricht parat, nämlich dass ebenfalls an diesem Tag die israelische Armee einen Schiffskonvoi des Free Gaza Movement geentert hatte. Beide Personen A und B haben mit Wahrscheinlichkeit $q_A = q_B = 0.75$ mit dieser Handlung gerechnet, da der Staat Israel dieses Vorgehen bereits mehrfach angekündigt hatte, wussten aber noch nichts davon. Intuitiv wird man die Gesamtmenge an Information, die die Person A durch diese beiden Nachrichten erhalten hat, auf

$$I(1) + I(0.75) = 0 - \mathrm{ld}(0.75) \approx 0.415\ \mathrm{bit}$$

festlegen. Dies liegt daran, dass sich beide Ereignisse (Rücktritt des Bundespräsidenten und Militäraktion Israels) gegenseitig nicht beeinflussen. Die Wahrscheinlichkeit für das Eintreten beider Ereignisse ist somit gleich $p_A q_A$ für Person A bzw. $p_B q_B$ für Person B und es gilt wegen (iii) für Person A:

$$I(p_A q_A) = I(p_A) + I(q_A) = 0 - \mathrm{ld}(0.75) = -\mathrm{ld}(0.75) \approx 0.415\ \mathrm{bit}.$$

Wie sieht nun die Gesamtmenge an Information für Person B aus? Wegen $0 \cdot 0.75 = 0$ und wegen der Festlegung

$$\infty + a = a + \infty = \infty \quad \text{für alle} \quad a \in \mathbb{R} \cup \{\infty\}$$

gilt:

$$\infty = I(0) = I(p_B q_B) = I(0 \cdot 0.75) = I(0) + I(0.75) = \infty - \mathrm{ld}(0.75) = \infty. \quad \triangleleft$$

2.2 Die mittlere Informationsmenge eines Zeichens

Wie in Tab. 1.1 zusammengefasst, gibt es in einem deutschen Text für jedes Zeichen eine bestimmte Auftrittswahrscheinlichkeit. Wählt man nun in einem deutschen Text irgendeine Stelle aus und betrachtet man das Zeichen, welches dort steht, so erhält man durch das Erkennen dieses Zeichens eine bestimmte Menge an Information, die in folgender Tab. 2.1 dokumentiert ist.

Tab. 2.1 Informationsmenge pro Zeichen

Nr.	Zeichen	W	$-\mathrm{ld}(W)$ [bit]	Nr.	Zeichen	W	$-\mathrm{ld}(W)$ [bit]
1	„Leerzeichen“	0.15149	2.72271	16	O	0.01772	5.81848
2	E	0.14700	2.76611	17	B	0.01597	5.96849
3	N	0.08835	3.50063	18	Z	0.01423	6.13492
4	R	0.06858	3.86607	19	W	0.01420	6.13797
5	I	0.06377	3.97098	20	F	0.01360	6.20025
6	S	0.05388	4.21411	21	K	0.00956	6.70877
7	T	0.04731	4.40171	22	V	0.00735	7.08804
8	D	0.04385	4.51128	23	Ü	0.00580	7.42973
9	H	0.04355	4.52118	24	P	0.00499	7.64674
10	A	0.04331	4.52916	25	Ä	0.00491	7.67006
11	U	0.03188	4.97120	26	Ö	0.00255	8.61529
12	L	0.02931	5.09246	27	J	0.00165	9.24331
13	C	0.02673	5.22540	28	Y	0.00017	12.52218
14	G	0.02667	5.22864	29	Q	0.00015	12.70275
15	M	0.02134	5.55030	30	X	0.00013	12.90920

Die mittlere Informationsmenge $\bar{I}$ pro Zeichen ergibt sich nun zu

$$\begin{aligned}
\bar{I} = \quad & 0.15149 \cdot 2.72271 + 0.14700 \cdot 2.76611 + 0.08835 \cdot 3.50063 + \\
+ \; & 0.06858 \cdot 3.86607 + 0.06377 \cdot 3.97098 + 0.05388 \cdot 4.21411 + \\
+ \; & 0.04731 \cdot 4.40171 + 0.04385 \cdot 4.51128 + 0.04355 \cdot 4.52118 + \\
+ \; & 0.04331 \cdot 4.52916 + 0.03188 \cdot 4.97120 + 0.02931 \cdot 5.09246 + \\
+ \; & 0.02673 \cdot 5.22540 + 0.02667 \cdot 5.22864 + 0.02134 \cdot 5.55030 + \\
+ \; & 0.01772 \cdot 5.81848 + 0.01597 \cdot 5.96849 + 0.01423 \cdot 6.13492 + \\
+ \; & 0.01420 \cdot 6.13797 + 0.01360 \cdot 6.20025 + 0.00956 \cdot 6.70877 + \\
+ \; & 0.00735 \cdot 7.08804 + 0.00580 \cdot 7.42973 + 0.00499 \cdot 7.64674 + \\
+ \; & 0.00491 \cdot 7.67006 + 0.00255 \cdot 8.61529 + 0.00165 \cdot 9.24331 + \\
+ \; & 0.00017 \cdot 12.52218 + 0.00015 \cdot 12.70275 + 0.00013 \cdot 12.90920 = \\
= \; & 4.11461 \text{ bit.}
\end{aligned}$$

Vergleicht man dieses Ergebnis mit der aus Tab. 1.1 berechneten mittleren Wortlänge $L = 4.14834$ Bits, so zeigt sich, dass die Codierung in Tab. 1.1 praktisch nicht mehr verbessert werden kann (vgl. Abschn. 3.3).

http://www.springer.com/978-3-662-46381-9

Mathematik der Information
Theorie und Anwendungen der Shannon-Wiener Information
Schäffler, S.
2015, XV, 160 S. 26 Abb., Softcover
ISBN: 978-3-662-46381-9