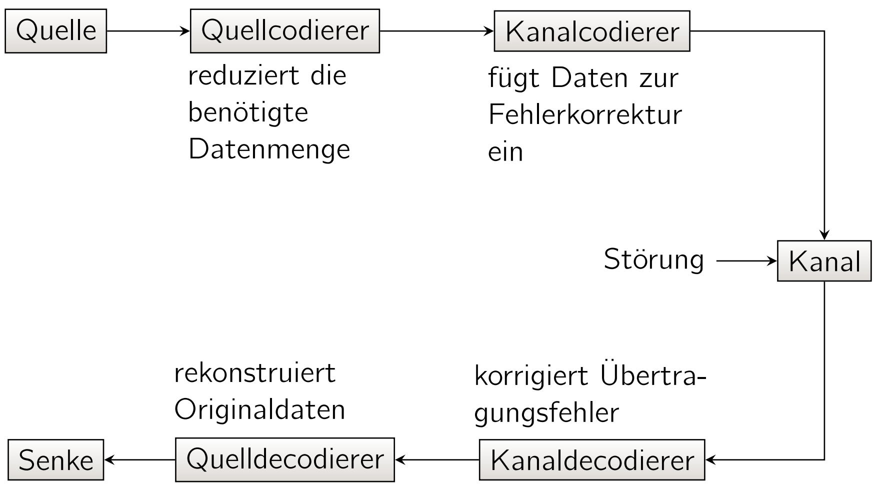
1.1 Modell der Nachrichtenübertragung

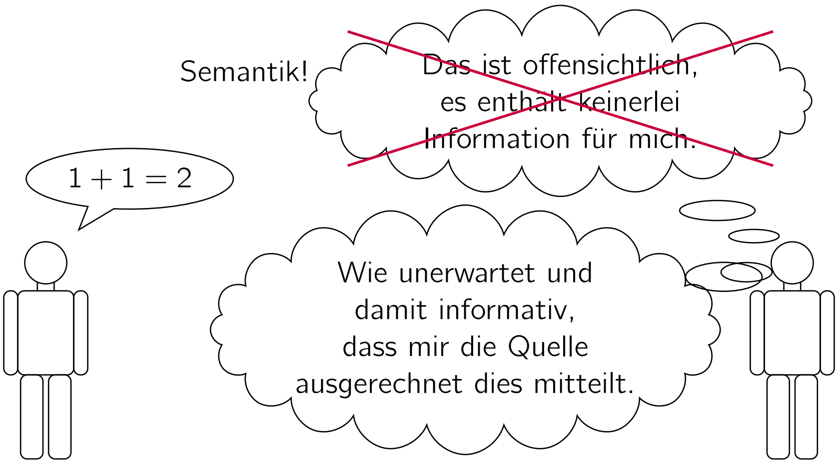
Aufgabe: Effiziente Übertragung von Daten über einen gestörten Kanal.



Die Aufteilung in Quell- und Kanalcodierung dient der Vereinfachung durch Betrachtung von Teilproblemen.

1.2 Informationsbegriff

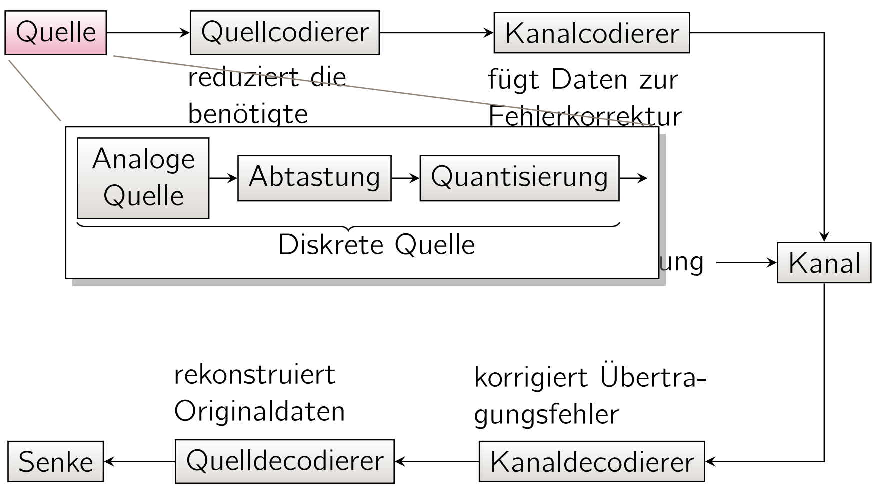
Der Shannonschen[1] Informationstheorie liegt ein *statistischer* Informationsbegriff zu Grunde.



Information ist beseitigte Unbestimmtheit.

[1] Claude E. Shannon gilt mit seinen Arbeiten vom Ende der 40er als (einer) der Begründer der Informations- und Kodierungstheorie.

2 Diskrete Quellen



2.1 Diskrete Quellen mit unabhängigen Ereignissen

Definition 2.1

Eine *diskrete Quelle* $\mathcal{X}$ liefert eine Sequenz x_k von zufälligen Symbolen aus einer Menge X, dem Alphabet. Wir nehmen eine stationäre Quelle an, sodass die Auftrittswahrscheinlichkeiten $p_X(x) \in [0,1], \sum_{x \in X} p_X(x) = 1$, von k unabhängig sind. Ferner gehen wir von unabhängigen Ereignissen aus, also $p_X(x_k) = p_X(x_k | x_1, \ldots, x_{k-1}, x_{k+1}, \ldots)$.

Beispiel 2.1 (Werfen einer Münze)

$X = \{\text{Kopf}, \text{Zahl}\}$, $p_X(\text{Kopf}) = p_X(\text{Zahl}) = 0{,}5$.

Beispiel 2.2 (Würfeln)

$X = \{1, \ldots, 6\}$, $p_X(1) = \cdots = p_X(6) = \frac{1}{6}$.

2.2 Informationsgehalt eines Symbols

Jedem Symbol x soll ein Informationsgehalt H_x zugeordnet werden, sodass

- $p_X(x) = 1 \Leftrightarrow H_x = 0$, ein sicheres Ereignis also keine Information trägt, da keine Unbestimmtheit beseitigt wird,
- $p_X(x) < p_X(y) \Leftrightarrow H_x > H_y$, unwahrscheinliche Symbole also mehr Information tragen und
- sich der Informationsgehalt H_{x_1,x_2} zweier *unabhängiger* Symbole x_1, x_2 (d.h. $p_{X,X}(x_1, x_2) = p_X(x_1) \cdot p_X(x_2)$) durch Addition der Einzelinformationen ergibt.

Diese Bedingungen werden von

$$H_x = -\log_r p_X(x) \tag{2.1}$$

erfüllt, wobei üblicherweise $r = 2$ verwendet wird. Wir schreiben $\text{ld} = \log_2$.

2.3 Entropie

Definition 2.2 (Entropie)

Unter der Entropie $H(\mathcal{X})$ einer Quelle $\mathcal{X}$ versteht man den durchschnittlichen Informationsgehalt der Symbole,

$$H(\mathcal{X}) = \sum_{x \in X} p_X(x) \cdot H_x = -\sum_{x \in X} p_X(x) \operatorname{ld}(p(x)), \qquad (2.2)$$

wobei für $0 \operatorname{ld} 0$ die stetige Fortsetzung $0 \operatorname{ld} 0 = \lim_{x \to 0} x \operatorname{ld} x = 0$ verwendet wird.

Beispiel 2.3

Für das Münzewerfen aus Beispiel 2.1 ergibt sich

$$H(\mathcal{X}) = -\frac{1}{2} \operatorname{ld}\left(\frac{1}{2}\right) - \frac{1}{2} \operatorname{ld}\left(\frac{1}{2}\right) = -\frac{1}{2} \cdot (-1) - \frac{1}{2} \cdot (-1) = 1.$$

Beispiel 2.4

Für den Würfel aus Beispiel 2.2 ergibt sich

$$H(\mathcal{X}) = -6 \cdot \frac{1}{6} \operatorname{ld}\left(\frac{1}{6}\right) = -\operatorname{ld}\left(\frac{1}{6}\right) \approx 2{,}585.$$

Beispiel 2.5

Für einen gezinkten Würfel mit $p_X(1) = \frac{1}{21}$, $p_X(2) = \frac{2}{21}$, $p_X(3) = \frac{3}{21}$, $p_X(4) = \frac{4}{21}$, $p_X(5) = \frac{5}{21}$, $p_X(6) = \frac{6}{21}$ ergibt sich

$$H(\mathcal{X}) = -\frac{1}{21} \operatorname{ld}\left(\frac{1}{21}\right) - \cdots - \frac{6}{21} \operatorname{ld}\left(\frac{6}{21}\right) \approx 2{,}398.$$

Die Symbole, die der gezinkte Würfel erzeugt, tragen weniger Information, da sie sich besser vorhersagen lassen.

Satz 2.1

Für die Entropie $H(\mathcal{X})$ einer Quelle $\mathcal{X}$ mit einem Alphabet X aus N Symbolen gilt

$$0 \leq H(\mathcal{X}) \leq \operatorname{ld} N. \tag{2.3}$$

Dabei ist $H(\mathcal{X}) = 0$ g.d.w. es ein sicheres Ereignis x' mit $p_X(x') = 1$ gibt ($p_X(x) = 0$ für $x \neq x'$), und $H(\mathcal{X}) = \operatorname{ld} N$ g.d.w. alle Symbole gleich wahrscheinlich sind mit $p_X(x) = \frac{1}{N}$ für alle $x \in X$.

Beweis für die untere Schranke.

Offenbar gilt

$$-p_X(x) \operatorname{ld}(p_X(x)) > 0 \text{ für } 0 < p_X(x) < 1 \tag{2.4}$$

und

$$-p_X(x) \operatorname{ld}(p_X(x)) = 0 \text{ für } p_X(x) \in \{0, 1\}. \tag{2.5}$$

Daraus folgt unmittelbar

$$H(\mathcal{X}) = \sum_{x \in X} -p_X(x) \operatorname{ld}(p_X(x)) \geq 0. \tag{2.6}$$

Gleichheit gilt dabei nur, wenn $-p_X(x) \operatorname{ld}(p_X(x)) = 0$ für alle x, also alle $p_X(x) \in \{0, 1\}$. Da $\sum_{x \in X} p_X(x) = 1$, muss in diesem Fall genau ein x' das sichere Ereignis mit $p_X(x') = 1$ sein, für die übrigen $x \neq x'$ muss $p_X(x) = 0$ gelten. □

Beweis für die obere Schranke.

Unter Verwendung eines Lagrange-Multiplikators λ für die Nebenbedingung $\sum_{x \in X} p_X(x) = 1$ und mit der abkürzenden Schreibweise $p_x = p_X(x)$ wird die Hilfsfunktion

$$f = -\sum_{x \in X} p_x \operatorname{ld} p_x + \lambda \left(1 - \sum_{x \in X} p_x\right) \tag{2.7}$$

gebildet. Nullsetzen der partiellen Ableitungen

$$\frac{\partial f}{\partial p_x} = -1 \cdot \operatorname{ld} p_x - p_x \cdot \frac{\operatorname{ld} e}{p_x} - \lambda = -\operatorname{ld} p_x - \operatorname{ld} e - \lambda = 0 \tag{2.8}$$

führt auf

$$\operatorname{ld} p_x = -\operatorname{ld} e - \lambda, \tag{2.9}$$

es sind also p_x gleich, was mit der Nebenbedingung auf das eindeutige Maximum $p_x = \frac{1}{N}$ führt, für das der Wert

$$-\sum_{x \in X} \frac{1}{N} \operatorname{ld}\left(\frac{1}{N}\right) = \operatorname{ld} N \tag{2.10}$$

erreicht wird. □

2.4 Verbundquellen

Definition 2.3

Die diskreten Quellen $\mathcal{X}$ und $\mathcal{Y}$ mit den Alphabeten X und Y und der Verbundwahrscheinlichkeit $p_{X,Y}(x,y)$ bilden eine *Verbundquelle* $(\mathcal{X},\mathcal{Y})$.

Beispiel 2.6

Zwei Spieler A und B spielen eine Partie Schach. Das Recht auf den ersten Zug wird durch Münzwurf entschieden. Die Quelle $\mathcal{X}$ mit $X = \{\mathrm{A},\mathrm{B}\}$ gebe das Ergebnis des Münzwurfs an, die Quelle $\mathcal{Y}$ mit $Y = \{\mathrm{A},\mathrm{B},\mathrm{R}\}$ (A gewinnt, B gewinnt, Remis) den Ausgang der Schachpartie. Die Verbundquelle $(\mathcal{X},\mathcal{Y})$ hat die Verbundwahrscheinlichkeit

$$
\begin{aligned}
&p_{X,Y}(\mathrm{A},\mathrm{A}) = 0{,}25 \qquad && p_{X,Y}(\mathrm{A},\mathrm{B}) = 0{,}15 \qquad && p_{X,Y}(\mathrm{A},\mathrm{R}) = 0{,}1\\
&p_{X,Y}(\mathrm{B},\mathrm{A}) = 0{,}15 && p_{X,Y}(\mathrm{B},\mathrm{B}) = 0{,}25 && p_{X,Y}(\mathrm{B},\mathrm{R}) = 0{,}1.
\end{aligned}
$$

2.5 Verbundentropie

Definition 2.4

Wird in Definition 2.2 $p_X(x)$ durch die Verbundwahrscheinlichkeit $p_{X,Y}(x,y)$ ersetzt, erhält man die *Verbundentropie*

$$H(\mathcal{X},\mathcal{Y}) = -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\,\mathrm{ld}\left(p_{X,Y}(x,y)\right). \tag{2.11}$$

Beispiel 2.7

Für die Verbundquelle aus Beispiel 2.6 ergibt sich

$$H(\mathcal{X},\mathcal{Y}) = -2\cdot 0{,}25\cdot \mathrm{ld}\,0{,}25 - 2\cdot 0{,}15\cdot \mathrm{ld}\,0{,}15 - 2\cdot 0{,}1\cdot \mathrm{ld}\,0{,}1 \approx 2{,}486.$$

Für die Wahrscheinlichkeiten der Einzelquellen gilt $p_X(x) = \sum_{y\in Y} p_{X,Y}(x,y)$ bzw. $p_Y(y) = \sum_{x\in X} p_{X,Y}(x,y)$, also

$$p_X(\mathrm{A}) = p_X(\mathrm{B}) = 0{,}5 \quad \text{und} \quad p_Y(\mathrm{A}) = p_Y(\mathrm{B}) = 0{,}4,\ p_Y(\mathrm{R}) = 0{,}2,$$

was auf die Einzelentropien $H(\mathcal{X}) = 1$ und $H(\mathcal{Y}) \approx 1{,}522$ führt.

Satz 2.2

Mit obigen Definitionen gilt

$$H(\mathcal{X}, \mathcal{Y}) \leq H(\mathcal{X}) + H(\mathcal{Y}) \tag{2.12}$$

mit Gleichheit g.d.w. $\mathcal{X}$ und $\mathcal{Y}$ unabhängig sind.

Für den Beweis benötigen wir

Lemma 2.1

Seien $p_1, \ldots, p_N, q_1, \ldots, q_N \geq 0$ mit $\sum_{i=1}^N p_i = 1$ und $\sum_{i=1}^N q_i = 1$, so gilt

$$-\sum_{i=1}^{N} p_i \operatorname{ld}(p_i) \leq -\sum_{i=1}^{N} p_i \operatorname{ld}(q_i) \tag{2.13}$$

mit Gleichheit g.d.w. $p_i = q_i$ für $i = 1, \ldots, N$.

Beweis von Lemma 2.1.

Wir nutzen $\ln(x) = \operatorname{ld}(x)\ln(2)$ und beweisen

$$-\sum_{i=1}^{N} p_i \ln(p_i) \leq -\sum_{i=1}^{N} p_i \ln(q_i). \tag{2.14}$$

Ausgehend von $\ln x \leq x - 1$ mit Gleichheit nur für $x = 1$ ergibt sich durch Ersetzen

$$\ln\left(\frac{q_i}{p_i}\right) \leq \frac{q_i}{p_i} - 1 \qquad \text{mit Gleichheit nur für } p_i = q_i. \tag{2.15}$$

Multiplikation mit p_i und Summieren über $i = 1, \ldots, N$ führt auf

$$\sum_{i=1}^{N} p_i \ln\left(\frac{q_i}{p_i}\right) \leq \sum_{i=1}^{N}(q_i - p_i) = \sum_{i=1}^{N} q_i - \sum_{i=1}^{N} p_i = 1 - 1 = 0, \tag{2.16}$$

also

$$\sum_{i=1}^{N} p_i \left(\ln(q_i) - \ln(p_i)\right) = \sum_{i=1}^{N} p_i \ln(q_i) - \sum_{i=1}^{N} p_i \ln(p_i) \leq 0 \tag{2.17}$$

mit Gleichheit nur für $p_i = q_i$. □

Beweis von Satz 2.2.

$$H(\mathcal{X}) = -\sum_{x\in X} p_X(x)\,\mathrm{ld}(p_X(x)) = -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\,\mathrm{ld}(p_X(x)) \tag{2.18}$$

$$H(\mathcal{Y}) = -\sum_{y\in Y} p_Y(y)\,\mathrm{ld}(p_Y(y)) = -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\,\mathrm{ld}(p_Y(y)) \tag{2.19}$$

$$\begin{aligned} H(\mathcal{X}) + H(\mathcal{Y}) &= -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\big(\mathrm{ld}(p_X(x)) + \mathrm{ld}(p_Y(y))\big) && (2.20)\\ &= -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\,\mathrm{ld}\left(p_X(x)\cdot p_Y(y)\right) && (2.21)\end{aligned}$$

$$H(\mathcal{X},\mathcal{Y}) = -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\,\mathrm{ld}\left(p_{X,Y}(x,y)\right) \le H(\mathcal{X}) + H(\mathcal{Y}) \tag{2.22}$$

nach Lemma 2.1, mit Gleichheit nur für $p_{X,Y}(x,y) = p_X(x)\cdot p_Y(y)$, also Unabhängigkeit. □

Korollar 2.1

Für die spezielle Verbundentropie $H(\mathcal{X},\mathcal{X}) = H(\mathcal{X}^2)$ gilt unter der Voraussetzung voneinander unabhängiger Einzelereignisse

$$H(\mathcal{X}^2) = 2H(\mathcal{X}) \tag{2.23}$$

und über Induktion folgt

$$H(\mathcal{X}^k) = k\cdot H(\mathcal{X}). \tag{2.24}$$

Werden mehrere Symbole auf einmal betrachtet, bleibt die Entropie pro Symbol konstant.

2.6 Bedingte Entropie

Definition 2.5

Der Ausdruck

$$H(\mathcal{Y}|\mathcal{X}) = -\sum_{x\in X}\sum_{y\in Y} p_X(x)p_{Y|X}(y|x)\,\mathrm{ld}\left(p_{Y|X}(y|x)\right) \tag{2.25}$$

wird als *bedingte Entropie* bezeichnet.

Die bedingte Entropie $H(\mathcal{Y}|\mathcal{X})$ gibt an, wie viel zusätzliche Information die Kenntnis von $\mathcal{Y}$ bei bereits bekanntem $\mathcal{X}$ bringt.

Satz 2.3

Mit obigen Definitionen gilt

$$H(\mathcal{X},\mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) = H(\mathcal{Y}) + H(\mathcal{X}|\mathcal{Y}). \tag{2.26}$$

Beweis.

$$\begin{aligned}
H(\mathcal{X},\mathcal{Y}) &= -\sum_{x\in X}\sum_{y\in Y} p_{X,Y}(x,y)\,\mathrm{ld}\left(p_{X,Y}(x,y)\right) && (2.27)\\
&= -\sum_{x\in X}\sum_{y\in Y} p_X(x)p_{Y|X}(y|x)\,\mathrm{ld}\left(p_X(x)p_{Y|X}(y|x)\right) \qquad (Bayes) && (2.28)\\
&= -\sum_{x\in X}\sum_{y\in Y} p_X(x)p_{Y|X}(y|x)\,\mathrm{ld}\left(p_X(x)\right) && (2.29)\\
&\quad -\sum_{x\in X}\sum_{y\in Y} p_X(x)p_{Y|X}(y|x)\,\mathrm{ld}\left(p_{Y|X}(y|x)\right) && (2.30)\\
&= -\sum_{x\in X} p_X(x)\,\mathrm{ld}\left(p_X(x)\right)\underbrace{\sum_{y\in Y} p_{Y|X}(y|x)}_{=1} + H(\mathcal{Y}|\mathcal{X}) && (2.31)\\
&= H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) && (2.32)
\end{aligned}$$

□

Korollar 2.2

Mit Satz 2.2 ($H(\mathcal{X},\mathcal{Y}) \le H(\mathcal{X}) + H(\mathcal{Y})$) folgt unmittelbar $H(\mathcal{Y}|\mathcal{X}) \le H(\mathcal{Y})$.

Beispiel 2.8

Für die Verbundquelle aus Beispiel 2.6 gilt

$$p_{X|Y}(\mathrm{A}|\mathrm{A}) = 0{,}625 \qquad p_{X|Y}(\mathrm{A}|\mathrm{B}) = 0{,}375 \qquad p_{X|Y}(\mathrm{A}|\mathrm{R}) = 0{,}5$$
$$p_{X|Y}(\mathrm{B}|\mathrm{A}) = 0{,}375 \qquad p_{X|Y}(\mathrm{B}|\mathrm{B}) = 0{,}625 \qquad p_{X|Y}(\mathrm{B}|\mathrm{R}) = 0{,}5$$

und

$$p_{Y|X}(\mathrm{A}|\mathrm{A}) = 0{,}5 \qquad p_{Y|X}(\mathrm{A}|\mathrm{B}) = 0{,}3$$
$$p_{Y|X}(\mathrm{B}|\mathrm{A}) = 0{,}3 \qquad p_{Y|X}(\mathrm{B}|\mathrm{B}) = 0{,}5$$
$$p_{Y|X}(\mathrm{R}|\mathrm{A}) = 0{,}2 \qquad p_{Y|X}(\mathrm{R}|\mathrm{B}) = 0{,}2.$$

Damit ergeben sich

$$H(\mathcal{Y}|\mathcal{X}) = -\sum_{x\in X}\sum_{y\in Y} p_X(x)p_{Y|X}(y|x)\,\mathrm{ld}\left(p_{Y|X}(y|x)\right) \approx 1{,}486$$

und

$$H(\mathcal{X}|\mathcal{Y}) = -\sum_{y\in Y}\sum_{x\in X} p_Y(y)p_{X|Y}(x|y)\,\mathrm{ld}\left(p_{X|Y}(x|y)\right) \approx 0{,}964.$$

Damit gilt $H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) \approx 1 + 1{,}486 = 2{,}486 \approx H(\mathcal{X},\mathcal{Y})$ bzw.
$H(\mathcal{Y}) + H(\mathcal{X}|\mathcal{Y}) \approx 1{,}522 + 0{,}964 = 2{,}486 \approx H(\mathcal{X},\mathcal{Y})$.

2.7 Transinformation

- $H(\mathcal{X})$: Wieviel Information trägt $\mathcal{X}$?
- $H(\mathcal{X}|\mathcal{Y})$: Wieviel zusätzliche Information trägt $\mathcal{X}$, wenn $\mathcal{Y}$ bekannt ist?

⇒ $H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y})$: Wieviel Information trägt $\mathcal{Y}$ über $\mathcal{X}$?

Definition 2.6

$$I(\mathcal{X},\mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) \tag{2.33}$$

wird als *Transinformation* von $\mathcal{X}$ und $\mathcal{Y}$ bezeichnet.

Die Transinformation ist eine allgemeinere Methode, um statistische Abhängigkeiten zu Untersuchen, als die in der Signalverarbeitung gern verwendete Korrelation.

Satz 2.4

Es gilt

$$I(\mathcal{X},\mathcal{Y}) \geq 0. \tag{2.34}$$

folgt unmittelbar aus Korollar 2.2 ($H(\mathcal{X}|\mathcal{Y}) \leq H(\mathcal{X})$).

Satz 2.5

Es gilt

$$I(\mathcal{X},\mathcal{Y}) = I(\mathcal{Y},\mathcal{X}). \tag{2.35}$$

Beweis.

Unter Ausnutzung von Satz 2.3 folgt

$$\begin{aligned} I(\mathcal{X},\mathcal{Y}) &= H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) && (2.36)\\ &= H(\mathcal{X},\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) && (2.37)\\ &= H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) = I(\mathcal{Y},\mathcal{X}). && (2.38)\end{aligned}$$

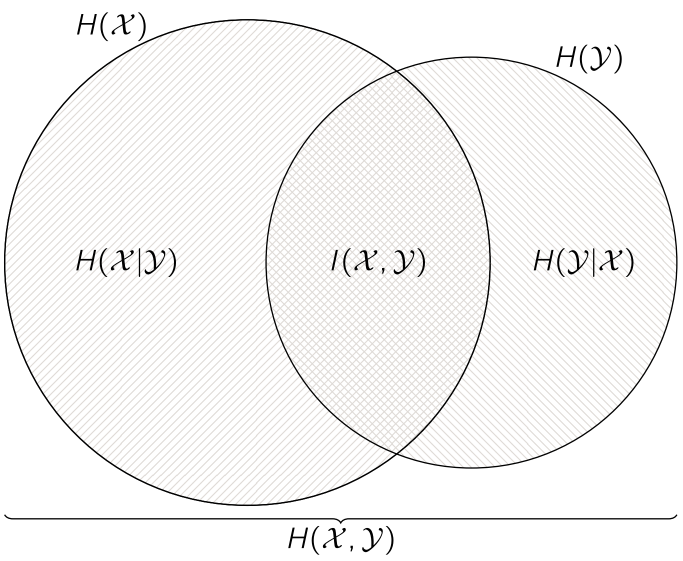
□

Beispiel 2.9

Nochmals seien die Schachspieler aus Beispiel 2.6 betrachtet:

$$I(\mathcal{X},\mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) \approx 1 - 0{,}964 = 0{,}036$$
$$I(\mathcal{Y},\mathcal{X}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) \approx 1{,}522 - 1{,}486 = 0{,}036$$

2.8 Darstellung als Venn-Diagramm



2.9 Zusammenfassung

- Die Entropie $H(\mathcal{X}) = -\sum p_X(x) \operatorname{ld} p_X(x)$ misst den durchschnittlichen Informationsgehalt eines Symbols der Quelle $\mathcal{X}$.
- Gleichverteilte Quelle mit N Symbolen erreicht die maximale Entropie $H(\mathcal{X}) = \operatorname{ld} N$.
- Die Verbundentropie $H(\mathcal{X}, \mathcal{Y})$ misst den Informationsgehalt der Kombination von Symbolen der Quellen $\mathcal{X}$ und $\mathcal{Y}$, es gilt $H(\mathcal{X}, \mathcal{Y}) \leq H(\mathcal{X}) + H(\mathcal{Y})$.
- Für den Spezialfall einer erweiterten Quelle gilt bei Unabhängigkeit der Symbole $H(\mathcal{X}^n) = n \cdot H(\mathcal{X})$.
- Die bedingte Entropie $H(\mathcal{Y}|\mathcal{X})$ gibt an, wieviel zusätzliche Information $\mathcal{Y}$ liefert, wenn $\mathcal{X}$ bekannt ist. Es gilt $H(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X})$ und $H(\mathcal{Y}|\mathcal{X}) \leq H(\mathcal{Y})$.
- Die Transinformation $I(\mathcal{X}, \mathcal{Y}) = H(\mathcal{X}) - H(\mathcal{X}|\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X})$ gibt an, wieviel Information $\mathcal{X}$ und $\mathcal{Y}$ übereinander tragen.