4 Informationskanäle und ihre Kapazität

4.1 Diskrete gedächtnislose Kanäle

Definition 4.1 Ein *diskreter gedächtnisloser Kanal* (engl. discrete memoryless channel, DMC) ist gegeben durch eine stochastische $r \times s$-Matrix $\Pi = (P_{ij})$, $i = 1, \ldots, r$, $j = 1, \ldots, s$.

Dabei heißt eine Matrix stochastisch, wenn ihre Zeilen stochastische Vektoren sind: $\sum_{j=1}^{s} P_{ij} = 1$ für alle $i = 1, \ldots, r$. Die Matrix Π heißt auch *Kanalmatrix* oder *Rauschmatrix*.

Interpretation: Zu einem Kanal gehört ein Quellalphabet $A = \{a_1, \ldots, a_r\}$ und ein "Zielalphabet" $B = \{b_1, \ldots, b_s\}$. Es ist

$$P_{ij} = Pr(b_j \mid a_i) = Pr(b = b_j \mid a = a_i)$$

die Wahrscheinlichkeit, b_j zu empfangen, wenn a_i gesendet wurde.

Beispiel 4.2 a) Der *binäre symmetrische Kanal* BSC:
$\Pi = \begin{pmatrix} P & 1-P \\ 1-P & P \end{pmatrix}$ für ein festes $P \in [0,1]$.
Oft schreibt man auch $1 - P =: \overline{P}$.
Typischerweise ist $A = B = \{0, 1\}$.

b) Der *binäre Kanal mit Auslöschungen* BEC (binary erasure channel):
$A = \{0,1\}$, $B = \{0, 1, ?\}$, $\Pi = \begin{pmatrix} P & 0 & \overline{P} \\ 0 & P & \overline{P} \end{pmatrix}$.

D.h. jedes Zeichen wird entweder richtig übertragen oder mit einer gewissen (kleinen) Wahrscheinlichkeit $\overline{P}$ ausgelöscht, formal: als ? übertragen. Allgemeiner kann hier $\Pi = \begin{pmatrix} P_1 & P_2 & P_3 \\ P_2 & P_1 & P_3 \end{pmatrix}$ sein (falsche Übertragung mit Wahrscheinlichkeit P_2 möglich).

Wir betrachten nun Systeme $(\mathcal{A}, \mathcal{B}, \Pi)$, wobei $\mathcal{A} = (A = \{a_1, \ldots, a_r\}, \mathbf{p} = (p_1, \ldots, p_r))$ und $\mathcal{B} = (B = \{b_1, \ldots, b_s\}, \mathbf{q})$ formal beides Quellen sind. In der Interpretation befindet sich $\mathcal{A}$ am Eingang und $\mathcal{B}$am Ausgang des Kanals. Ein externer Beobachter sieht $\mathcal{A}$ und $\mathcal{B}$ und interpretiert den Zusammenhang bzw. Unterschied als Wirkung des Kanals. Neben Π, $\mathbf{p}$ und $\mathbf{q}$ unter der Nebenbedingung $\mathbf{q} = \mathbf{p}\Pi$ setzen wir noch folgende Wahrscheinlichkeiten als bekannt an:

$Q_{ij} = Pr(a = a_i \mid b = b_j)$
$R_{ij} = Pr(a_i, b_j) = Pr(a = a_i \wedge b = b_j)$

Warum können wir in unserem Modell die Existenz der Q_{ij} und R_{ij} fordern? Die p_i der Quelle sind eigentlich als Verteilung einer Zufallsvariablen X, nämlich eines gesendeten Symbols, zu interpretieren, entsprechend sind die q_j die Verteilung von Y, der am Ausgang des Kanals beobachteten Zufallsvariablen. Nun besitzen zwei Zufallsvariable X und Y, die auf demselben Wahrscheinlichkeitsraum Ω definiert sind, immer eine gemeinsame Verteilung:

$$Pr(X = a_i \wedge Y = b_j) = Pr\{\omega \mid X(\omega) = a_i \wedge Y(\omega) = b_j\}$$

ist definiert. Die obigen bedingten Wahrscheinlichkeiten sind hieraus abgeleitete Größen:

$$P_{ij} = Pr(b_j|a_i) = \frac{R_{ij}}{p_i} \quad \text{, falls } p_i \neq 0$$
$$Q_{ij} = Pr(a_i|b_j) = \frac{R_{ij}}{q_j} \quad \text{, falls } q_j \neq 0 \ .$$

In jedem Fall gilt

$$p_i P_{ij} = R_{ij} = Q_{ij} q_j \ .$$

Insbesondere haben wir die Bayes'sche Regel $Q_{ij} = \frac{p_i}{q_j} P_{ij}$ und weiter

$$Q_{ij} = \frac{p_i P_{ij}}{\sum_{k=1}^{r} p_k P_{kj}} \quad .$$

Die "Rückwärtswahrscheinlichkeiten" Q_{ij} ergeben sich also aus der Kanalmatrix Π und dem Input $\mathbf{p}$. Für eine feste Inputverteilung bekommt man unter Umständen nicht ganz Q, sondern nur gewisse Spalten. Man erhält die j-te Spalte, wenn der j-Eintrag von $\mathbf{q} = \mathbf{p}\Pi$ ungleich Null ist.

4.2 System-Entropien

Wir betrachten ein System $(\mathcal{A}, \mathcal{B}, \Pi, (R_{ij}))$ aus Input, Kanal und Output wie im vorigen Abschnitt. Alle beteiligten Verteilungen besitzen eine Entropie. Das sind

Input-Entropie	$H(\mathcal{A})$	$= -\sum_i p_i \log p_i$
Output-Entropie	$H(\mathcal{B})$	$= -\sum_j q_j \log q_j$
bedingte Entropie	$H(\mathcal{A}\|b_j)$	$= -\sum_i Pr(a_i\|b_j) \log Pr(a_i\|b_j)$
bedingte Entropie	$H(\mathcal{B}\|a_i)$	$= -\sum_j Pr(b_j\|a_i) \log Pr(b_j\|a_i)$
Äquivokation	$H(\mathcal{A}\|\mathcal{B})$	$= \sum_j q_j H(\mathcal{A}\|b_j)$
Irrelevanz	$H(\mathcal{B}\|\mathcal{A})$	$= \sum_i p_i H(\mathcal{B}\|a_i)$
gemeinsame Entropie	$H(\mathcal{A}, \mathcal{B})$	$= -\sum_{i,j} R_{ij} \log R_{ij}$

Die betrachteten Entropien heißen gemeinsam die *System-Entropien* des betrachteten Systems. Die Bezeichnungen sind nicht ganz glücklich gewählt, weil nämlich die Werte im Allgemeinen nicht nur von $\mathcal{A}$ und $\mathcal{B}$ (d.h. von den Wahrscheinlichkeitsvektoren $\mathbf{p}$ und $\mathbf{q}$) abhängen, sondern wesentlich auch von der Kanalmatrix Π (wobei dann wegen $\mathbf{q} = \mathbf{p}\Pi$ natürlich $\mathbf{q}$ nicht mehr gebraucht wird). Übrigens hat man dieses Problem nicht, wenn man von Zufallsvariablen ausgeht: dann sind die analogen Definitionen $H(X|Y)$ usw. alle unmittelbar sinnvoll.

Noch klarer werden die Zusammenhänge, wenn wir die Formeln etwas umstellen. Dabei wollen auch die Systementropien noch inhaltlich interpretieren. Einfaches Einsetzen ergibt zunächst für die zweitgenannte relative Entropie:

$$H(\mathcal{B}|a_i) = -\sum_j P_{ij} \log P_{ij}.$$

Dieser Wert ist also gar nicht von $\mathcal{A}$ oder $\mathcal{B}$, sondern nur von der Kanalmatrix, genauer ihrer i-ten Zeile abhängig. Es handelt sich hier um die Unsicherheit (oder, wie immer bei Entropie, um den möglichen Informationsgewinn) von $\mathcal{B}$ im Fall, daß bekanntermaßen das Symbol a_i gesendet wurde.

Die Irrelevanz ist dann das gewichtete Mittel all dieser Werte, gemittelt über Symbole der Quelle:

$$H(\mathcal{B}|\mathcal{A}) = -\sum_{ij} p_i P_{ij} \log P_{ij}.$$

Dieses ist also die mittlere Unsicherheit von $\mathcal{B}$ bei bekannter gesendeter Symbolfolge. Es handelt sich um einen "scheinbareren" möglichen Informationsgewinn aus $\mathcal{B}$, der nur durch das Rauschen (die Störungen) von $\mathcal{B}$ bedingt ist. Für die eigentliche Information von $\mathcal{A}$ ist er irrelevant, daher der Name.

Für die erste relative Entropie und die Äquivokation ergibt sich entsprechend

$$H(\mathcal{A}|b_j) = -\sum_i Q_{ij} \log Q_{ij}$$
$$H(\mathcal{A}|\mathcal{B}) = -\sum_{ij} Q_{ij} q_j \log Q_{ij}$$

Die Äquivokation stellt die Unsicherheit über die von $\mathcal{A}$ gesendete Nachricht bei bekannter empfangener Nachricht dar, sozusagen die noch fehlende Information, die die exakte Kenntnis der Sendung zusätzlich liefern würde.

Die bisherigen Formeln zeigen, daß die Irrelevanz nur vom Input und vom Kanal (d.h. der Kanalmatrix) abhängt; für die Äquivokation ist das noch nicht klar, weil in der Formel die Matrix der Q_{ij} vorkommt. Für die gemeinsame Entropie ist die Situation wieder wie beim Input, denn die zugehörige Matrix

$R_{ij} = p_i P_{ij}$ ist wieder sofort aus Input und Kanal abzulesen.

Letzlich sind alle Systementropien, auch die Äquivokation, nur von der Verteilung des Inputs und vom Kanal abhängig, wie sich aus folgendem Satz direkt ergibt.

Satz 4.3 *Für die gemeinsame Entropie eines Systems* $(\mathcal{A}, \mathcal{B}, \Pi)$ *gelten die Zerlegungen*

$$\begin{aligned} H(\mathcal{A}, \mathcal{B}) &= H(\mathcal{A}) + H(\mathcal{B}|\mathcal{A}) \\ &= H(\mathcal{B}) + H(\mathcal{A}|\mathcal{B}) \end{aligned}$$

Die erste Gleichung bestätigt noch einmal die Interpretation der Irrelevanz: die Gesamtinformation des Systems besteht aus der Information von $\mathcal{A}$ zusammen mit der irrelevanten Information, die das Rauschen des Kanals scheinbar hinzufügt. Entsprechend bestätigt die zweite Gleichung, daß die Äquivokation der noch fehlende Teil der Gesamtinformation bei schon bekanntem Output des Kanals ist.

Zum Beweis der ersten Formel setzt man die Gleichung $R_{ij} = p_i P_{ij}$ in die Definition ein, zieht die Faktoren $\log R_{ij}$ entsprechend auseiander, bekommt so zwei Summen, von denen sich die erste wegen $\sum_j R_{ij} = p_i$ zu $H(\mathcal{A})$ vereinfacht.
Die zweite Formel sieht man völlig analog.

Beispiel: die Systementropien des BSC

Wir betrachten den binären symmetrischen Kanal mit Kanalmatrix mit Zeilen und Spalten P, $\overline{P} = 1 - P$ sowie Inputverteilung $(p,\ \overline{p} = 1 - p)$.

Wir erinnern ferner an die Entropiefunktion $H(x) = -x \log x - (1 - x) \log(1 - x)$ mit ihrem bekannten Verlauf: $H(0) = H(1) = 0$, strikt konvex (oft auch konkav genannt) wegen zweiter Ableitung; all dieses gilt für die beiden Summanden einzeln; deren Maximum liegt bei $1/r$ bzw. $1 - 1/r$, wobei r die Basis der verwendeten Logarithmus, hier $r = 2$ ist. H selbst ist symmetrisch um den Wert $x = \frac{1}{2}$.

Für die Systementropien des BSC rechnet man nun aus:

$$\begin{aligned} H(\mathcal{A}) = &\quad H(p) \\ H(\mathcal{B}|\mathcal{A}) = &\quad H(P) \\ H(\mathcal{A}, \mathcal{B}) = &\quad H(p) + H(P) \\ H(\mathcal{A}|\mathcal{B}) = &\quad H(p) + H(P) - H(q) \\ &\quad q = pP + \overline{p}\overline{P}. \end{aligned}$$

4.3 Transinformation und Kanalkapazität

Definition 4.4 Gegeben sei ein Kanal Π mit Input $\mathcal{A}$ und Output $\mathcal{B}$. Die *Transinformation* oder Synentropie (engl. auch mutual information) dieses Systems ist definiert als

$$I(\mathcal{A}, \mathcal{B}, \Pi) := H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B}),$$

also als die Entropie von $\mathcal{A}$ abzüglich der Äquivokation des Systems.

Wir erinnern daran, daß nach den Überlegungen des vorigen Abschnittes die Äquivokation tatsächlich von der Quelle $\mathcal{A}$ und dem Kanal Π abhängt, aber von nichts weiterem ($\mathcal{B}$ ergibt sich sowieso hieraus). Die Bezeichnung $I(\mathcal{A}, \mathcal{B}, \Pi)$ ist also sinnvoll, und wir können kürzer auch $I(\mathcal{A}, \Pi)$ schreiben.

Die Bezeichnung $I(\mathcal{A}, \mathcal{B})$ ist in Fortführung entsprechender Bezeichnungen für die Systementropien ebenfalls üblich, zeigt aber nicht die Abhängigkeit von Kanal. Eigentlich ist sie nur gerechtfertigt für Zufallsvariablen X und Y statt ihrer Verteilungen $\mathcal{A}$ und $\mathcal{B}$ (siehe oben).
Die früher besprochenen Interpretationen der Systementropien liefern nun: Die Transinformation ist derjenige Teil der Unsicherheit über den Input, der veschwindet, wenn man den Output des Kanals kennt, oder – äquivalent – derjenige Teil der Information des Inputs, der durch den Output geliefert wird. Kurz: *die Transinformation ist die durch den Kanal übertragene Information der gegebenen Quelle.*

Wir können, auch wenn die Rollen von $\mathcal{A}$ und $\mathcal{B}$ auf Π bezogen nicht gleichwertig sind, auch

$$I(\mathcal{B}, \mathcal{A}, \Pi) := H(\mathcal{B}) - H(\mathcal{B}|\mathcal{A})$$

bilden. Mittels Satz 4.3 folgt dann

$$I(\mathcal{A}, \mathcal{B}, \Pi) = I(\mathcal{B}, \mathcal{A}, \Pi) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A}, \mathcal{B}).$$

Für den binären symmetrischen Kanal BSC mit Übergangswahrscheinlichkeiten P, $\overline{P} = 1 - P$ sowie Inputverteilung $(p,\ \overline{p} = 1 - p)$ folgt aus früheren Formeln

$$\begin{aligned} I(\mathcal{A}, \mathcal{B}) = &\quad H(q) - H(P) \\ = &\quad H(pP + \overline{p}\overline{P}) - H(P), \end{aligned}$$

was wegen des bekannten Verlaufs der Funktion $H \geq 0$ ist (q liegt auf der x-Achse zwischen P und $\overline{P}$.

Bei festem Kanal (also festem P) wird der Maximalwert $1 - H(P)$ der Transinformation $I(\mathcal{A}, \mathcal{B})$ für $p = \frac{1}{2}$ ereicht.

Nicht nur für den BSC gilt:

Satz 4.5 *Die Transinformation $I(\mathcal{A}, \mathcal{B}, \Pi)$ ist immer ≥ 0.*

Sie ist 0 dann und nur dann, wenn Input und Output statistisch unabhängig sind.

Wir merken wieder an, daß der Zusatz vor allem Sinn macht, wenn wir statt $\mathcal{A}$ und $\mathcal{B}$ Zufallsgrößen betrachten.

Definition 4.6 Die *Kanalkapazität* oder kurz *Kapazität* eines durch die stochastische Matrix Π gegebenen Kanals ist definiert als das Maximum der Transinformationen

$$C(\Pi) := \max_{\mathcal{A}} I(\mathcal{A}, \Pi),$$

wobei $\mathcal{A}$ über alle möglichen (Input-)Verteilungen variiert.

Für den BSC folgt aus früheren Rechnungen:

Satz 4.7 *Die Kapazität des BSC mit Übergangs- wahrscheinlichkeiten $P, 1-P$ ist gleich $C = 1 - H(P)$.*