Zusammenfassung der 10. Vorlesung (07.01.08)

1.4.3 *Der klassische Informationskanal*:

Um an einem Beispiel zu zeigen , wie die Shannon Entropie verwendet werden kann, betrachten wir einen klassischen gestörten Informationskanal. Sei $\mathbf{A} := \{a_i\}_{i=0,1,2,\dots,(M-1)}$ ein Alphabet. Zu einem Wort, das auch unendlich lang sein kann, gehört ein Wahrscheinlichkeitsraum $(\mathbf{A}, \mathcal{A}, \mu)$, wobei $\mathcal{A}$ die Potenzmenge von $\mathbf{A}$ und $\mu(a_i)$ die Wahrscheinlichkeit für das zufällige Erkennen des Zeichens a_i ist. Die Entropie des Wortes ist $H = -\sum_{i=0}^{M-1} \mu(a_i) \log_2 \mu(a_i)$.

Der Einfachheit halber wollen wir einen Kanal betrachten, der keine Kodierung erzeugt, d.h. das gesendete und das empfangene Wort besteht aus Zeichen des Alphabets $\mathbf{A}$, und im Idealfall stimmt das empfangene Wort mit dem gesendeten überein. Bei der Beschreibung eines solchen Kanals sind die Elementarereignisse (a_i, a_k), wobei erste Faktor ein gesendetes und der zweite Faktor das daraufhin empfangenes Zeichen angibt. Zur Beschreibung des Kanals betrachtet man daher einen Wahrscheinlichkeitsraum $(\mathbf{A} \times \mathbf{A}, \mathcal{A} \otimes \mathcal{A}, \sigma)$ mit den marginalen Wahrscheinlichkeitsräumen $(\mathbf{A}, \mathcal{A}, \mu_s)$ für das gesendete und $(\mathbf{A}, \mathcal{A}, \mu_e)$ für das empfangene Wort:

$$\boxed{\text{Sender}} \longrightarrow \boxed{\text{Kanal}} \longrightarrow \boxed{\text{Empfänger}}$$

$$(\mathbf{A}, \mathcal{A}, \mu_s) \qquad (\mathbf{A} \times \mathbf{A}, \mathcal{A} \otimes \mathcal{A}, \sigma) \qquad (\mathbf{A}, \mathcal{A}, \mu_e)$$

Im Fall eines idealen Kanals ist $\sigma(\{a_i\}, \{a_k\}) = \delta_{i,k}\sigma(\{a_i\}, \{a_i\})$. Mit $\sigma_{ik} := \sigma(\{a_i\}, \{a_k\})$ sei im allgemeinen Fall

$$p_i := \mu_s(a_i) = \sigma(\{a_i\}, \mathbf{A}) = \sum_k \sigma_{ik}, \quad q_k. = \mu_e(a_k) = \sigma(\mathbf{A}, \{a_k\}) = \sum_i \sigma_{ik}$$

und

$$p_{ki} := \sigma((\mathbf{A}, \{a_k\})|(\{a_i\}), \mathbf{A})) = \frac{\sigma_{ik}}{p_i}, \quad q_{ik} := \sigma((\{a_i\}), \mathbf{A})|(\mathbf{A}, \{a_k\})|) = \frac{\sigma_{ik}}{q_k}.$$

Beim realen Kanal muss die Entropie des gesendeten Wortes, $H_s = -\sum_i p_i \log_2 p_i$, nicht mit der Entropie des empfangenen Wortes, $H_e = -\sum_i q_i \log_2 q_i$, übereinstimmen. Zum Beispiel ist $H_s = 0$, wenn das gesendete Wort nur das Zeichen a_i enthält. Die Entropie des empfangenen Wortes ist dann jedoch $H_e = H(q|a_i)$, wobei

$$H(q|a_i) := -\sum_k p_{ki} \log_2 p_{ki}$$

die **bedingte Entropie** genannt wird. Die mit $H(q|a_i)$ gemessene Information, die der Empfänger erhält, ist fur die zu übertragende Information nicht relevant. Allgemein wird der Erwartungswert von $H(q|a_i)$ **Irrelevanz** genannt:

$$H(q|p) = -\sum_{i,k} p_i p_{ki} \log_2 p_{ki} = -\sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 p_i) = H_v - H_s,$$

wobei $H_v := -\sum_{i,k} \sigma_{ki}(\log_2 \sigma_{ki}$ **Verbundentropie** genannt wird. Im Idealfall gilt $\sigma_{ik} = \delta_{ik}\sigma_{ii}$ und damit $p_i = \sigma_{ii}$. Dann ist $H_v = H_s$ und die Irrelevanz verschwindet.

Entsprechend kann im realen Fall bei Empfang des Zeichens a_k nicht darauf geschlossen werden, dass auch a_k gesendet wurde. Mit der Wahrscheinlichkeit $q_{i,k}$ könnte auch a_i gesendet worden sein. Die bedingte Entropie

$$H(p|a_k) = -\sum_i q_{ik} \log_2 q_{ik}$$

misst die Information der Zeichenreihe, die "hinter" dem Empfang des Zeichens stehen könnte, also gleich gelautet hätten. Der Erwartungswert dieser bedingten Entropie über der Verteilung der empfangenen Zeichen wird **Äquivokation** genannt:

$$H(p|q) = -\sum_{i,k} q_k q_{ik} \log_2 q_{ik} = -\sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 q_k) = H_v - H_e.$$

Im Idealfall, $\sigma_{ik} = \delta_{ik}\sigma_{ii}$, verschwindet auch die Äquivokation.

Es gelten nun die folgenden Sätze:

Satz 1:

$$H_s - H(q|p) = H_e - H(p|q) = \sum_{ik} \sigma_{ik} \log_2 \frac{\sigma_{ik}}{p_i q_k} =: I.$$

I heißt die **Transinformation** (mutual information). Im Idealfall, $\sigma_{ik} = \delta_{ik}\sigma_{ii}$, gilt $I = H_s = H_e$.

Beweis (von Satz 1):

$$\begin{aligned} I &= \sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 p_i - \log_2 q_k) \\ &= -H_v + H_s + H_e = H_s - H(q|p) = H_e - H(p|q). \end{aligned}$$

Satz 2:

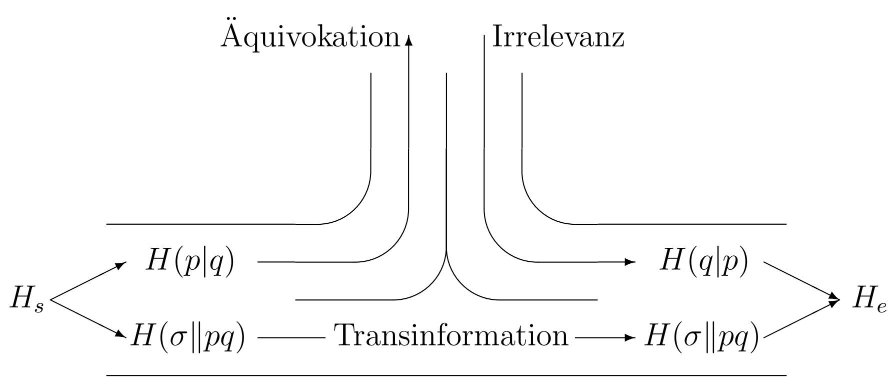
$$I \geq O, \qquad (I = 0 \quad \Leftrightarrow \quad \sigma_{ik} = p_i q_k)$$

Beweis: Die Behauptung folgt unmittelbar daraus, dass

$$I = \sum_{i,k} \sigma_{ik}(\log_2 \sigma_{ik} - \log_2 p_i q_k) = H(\sigma \| pq)$$

ist, wobei $pq = \{p_i q_k\}$ die Verteilung ist, bei der keine Korrelation zwischen gesendetem und empfagenen Zeichen besteht. $H(\sigma \| pq)$ ist die **relative Entropie** der Verteilung σ nach pq, die allgemein die Aussage des Satzes als Eigenschaft hat, wie wir im anschließenden Abschnitt zeigen werden.

Zunächst veranschaulichen wir uns das Ergebnis der Analyse des klassischen Informationskanals im Bergerschen Diagramm:



1.4.4 *Die relative Entropie*:

Für zwei diskrete Verteilungen $p_i \geq 0$, $\sum_{i=0}^{M-1} p_i = 1$, und $q_i \geq 0$, $\sum_{i=0}^{M-1} q_i = 1$, ist die **relative Entropie** von p nach q durch

$$H(p\|q) = \sum_{i=0}^{M-1} p_i(\log_2 p_i - \log_2 q_i) = -H(p) - \sum_{i=0}^{M-1} p_i(\log_2 q_i)$$

definiert. Die relative Entropie ist nicht symmetrisch in p und q, liefert jedoch ein Maß dafür, wie weit sich p von q unterscheidet. Dies zeigt der folgende Satz:

Satz(Kleinsche Ungleichung):

$$H(p\|q) \geq 0, \qquad (H(p\|q) = 0 \Leftrightarrow p_i = q_i)$$

Beweis: Der Beweis beruht auf der strikten Konkavität des Logarrithmus: Seien $x, y > 0$, dann gilt für $\lambda \in [0, 1]$ stets

$$\lambda \log_2 x + (1 - \lambda) \log_2 y \leq \log_2(\lambda x + (1 - \lambda))y),$$

und die Gleichheit gilt genau dann, wenn $x = y$ oder $\lambda \in \{0, 1\}$ ist. Damit folgt zum einen

$$\begin{aligned} H(p\|q) &= -\sum_{\{i|p_i \neq 0\}} p_i \log_2 \frac{q_i}{p_i} \\ &\geq -\log_2(\sum_{\{i|p_i \neq 0\}} p_i \frac{q_i}{p_i}) = -\log_2(\sum_{\{i|p_i \neq 0\}} q_i) \geq 0 \end{aligned}$$

Ist nun $H(p\|q) = 0$, dann gilt $-\log_2(\sum_{\{i|p_i \neq 0\}} q_i) = 0$, also $\sum_{\{i|p_i \neq 0\}} q_i = 1$, und damit $p_i = 0 \Leftrightarrow q_i = 0$. Ferner ist

$$-\sum_{\{i|p_i \neq 0\}} p_i \log_2 \frac{q_i}{p_i} = -\log_2(\sum_{\{i|p_i \neq 0\}} p_i \frac{q_i}{p_i}).$$

Aus der strikten Konkavität von $\log_2$ folgt deshalb zum anderen für $p_i \neq 0$, dass $p_i \in \{0, 1\}$ oder $\frac{q_i}{p_i} = \frac{q_k}{p_k} =: \alpha$ gelten muss. Im ersten Fall gibt es dann genau ein i mit $p_i = 1$, und deshalb muss auch $q_i = 1$ sein. Im zweiten Fall ist $-\log_2 \alpha = 0$, also $p_i = q_i$.