Vorlesung 12a

Quellencodieren und Entropie

Teil 4:

Der Quellencodierungssatz von Shannon

Satz

(Quellencodierungssatz von Shannon)

a) Für binäre Shannon-Codes gilt

$$\mathbf{E}[\ell(X)] < \mathbf{H}_2[X] + 1.$$

b) Für jeden binären Präfixcode gilt

$$\mathbf{E}[\ell(X)] \geq \mathbf{H}_2[X].$$

Beweis:

Für Shannon-Codes gilt

$$\ell(a) < -\log_2 \rho(a) + 1$$

und folglich

$$\mathbf{E}[\ell(X)] < \sum_a \rho(a)(-\log_2 \rho(a) + 1) = \mathbf{H}_2[X] + 1.$$

Dies beweist erst einmal Teil a) des Satzes.

Jetzt zu Teil b):

Für beliebige Codes gilt

$$\mathbf{H}_2[X] - \mathbf{E}[\ell(X)] = \sum_{a:\rho(a)>0} \rho(a) \log_2 \frac{2^{-\ell(a)}}{\rho(a)}$$

Nun ist die Logarithmusfunktion konkav

und liegt unterhalb ihrer Tangente im Punkt 1.

Folglich gilt $\log_2 x \leq c \cdot (x-1)$ mit geeignetem $c > 0$, und

$$\begin{aligned}\mathbf{H}_2[X] - \mathbf{E}[\ell(X)] &\leq c \sum_{a:\rho(a)>0} \rho(a) \Big(\frac{2^{-\ell(a)}}{\rho(a)} - 1 \Big) \\ &\leq c \cdot \Big(\sum_a 2^{-\ell(a)} - 1 \Big).\end{aligned}$$

Nach Fano-Kraft ist die rechte Seite ≤ 0, also

$$\mathbf{H}_2[X] - \mathbf{E}[\ell(X)] \leq 0. \quad \square$$

Die im Quellencodierungssatz auftretende Größe

$$\mathbf{H}_2[X] := -\sum_{a \in S} \rho(a) \log_2 \rho(a)$$

heißt die **(binäre) Entropie** von X

(die Entropie zur Basis 2).

Es ist sinnvoll, auch andere Basen zu erlauben:

$$\mathbf{H}_b[X] := -\sum_{a \in S} \rho(a) \log_b \rho(a).$$

Die Entropie ist *der* fundamentale Begriff
der Informationstheorie.

Nach dem Quellenkodierungssatz gibt die binäre Entropie
fast genau die mittlere Anzahl von Ja-Nein Fragen an,
die notwendig und
- bei guter Wahl des Codes - auch hinreichend ist,
um den unbekannten Wert von X von jemandem zu erfragen,
der X beobachten kann.

Dies ist gemeint, wenn man die Entropie beschreibt als den
Grad von Unbestimmtheit oder Ungewissheit
über den Wert, den X annimmt.