Vorlesung 12a

Quellencodieren und Entropie

Teil 3

Sparsames Codieren zufälliger Buchstaben:

Shannon-Codes und die (binäre) Entropie

Sei X eine S-wertige Zufallsvariable

(ein “zufälliger Buchstabe”)

mit Verteilungsgewichten $\rho(a) = \mathbf{P}(X = a)$.

Gefragt ist nach einem binären Präfixcode für S,

der in dem Sinn möglichst günstig ist,

dass die erwartete Codelänge von X,

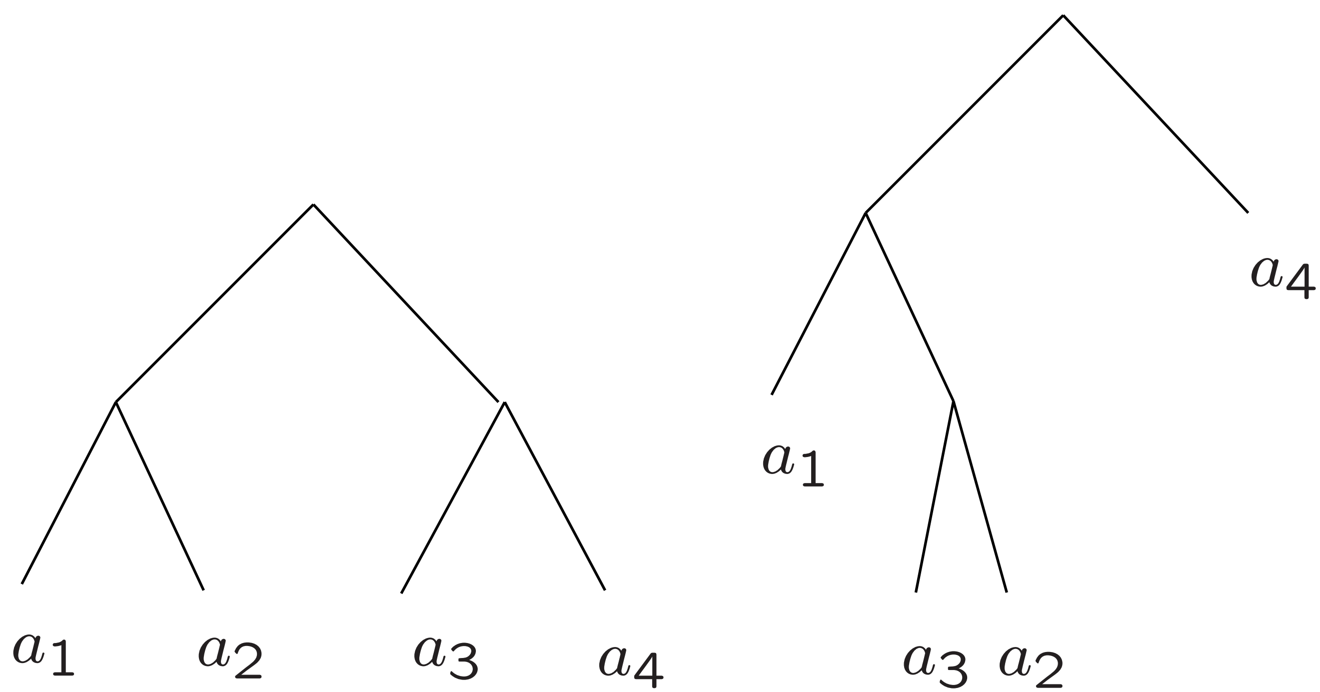
$$\mathbf{E}[\ell(X)] = \sum_{a \in S} \ell(a)\rho(a,)$$

möglichst klein ist.

Beispiel: Beispiel: $S = \{a_1, a_2, a_3, a_4\}$.

Sind die vier Ausgänge gleich wahrscheinlich,

dann ist der Code links günstiger als der rechts.



Für jedes $t \in \mathbb{N}$ ist glaubhaft

(und in der Tat auch eine Folgerung des in Abschnitt 4. formulierten und bewiesenen Quellencodierungssatzes):

Gilt $\#S = 2^t$ mit $\rho(a) = 2^{-t},\ a \in S$,

dann ist es am günstigsten, alle 2^t Buchstaben

mit den 01-Folgen der Länge t zu codieren.

In diesem Fall gilt

$$\ell(a) = t = -\log_2 \rho(a) \quad \text{für alle } a \in S.$$

Gegeben seien also

eine (endliche oder abzählbar unendliche) Menge S

und Verteilungsgewichte $\rho(a)$ auf S.

Wir denken uns

eine S-wertige Zufallsvariablle X mit Verteilung ρ.

Ein *Shannon-Code* für (S, ρ) ist ein Präfixcode, bei dem jeder Buchstabe $a \in S$ mit einer 01-Folge codiert wird, deren Länge $\ell(a)$ durch Aufrunden von $-\log_2 \rho(a)$ auf die nächste ganze Zahl entsteht, also

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

Solche Codes gibt es immer, denn für die so festgelegten $\ell(a)$ folgt

$$\sum_{a \in S} 2^{-\ell(a)} \leq \sum_{a \in S} \rho(a) = 1 \; ,$$

die Fano-Kraft Bedingung ist also erfüllt.

In Teil 4 werden wir zeigen:

Shannon-Codes

verfehlen die minimal mögliche erwartete Codelänge

um höchstens ein Bit.

Bis auf diesen möglichen “Rundungsfehler” ist ihre erwartete Coedelänge gleich der sogenannten binären Entropie

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

$$(\text{mit } 0 \log 0 := 0),$$

und besser geht’s auch nicht!