Vorlesung 12a

Quellencodieren und Entropie

Teil 2

Binäre Präfixcodes

als Blattbeschriftungen von Binärbäumen

S sei eine endliche (oder abzählbar unendliche) Menge

(ein "Alphabet").

Die Elemente von S nennen wir *Buchstaben*.

Wir wollen die Buchstaben $a, a', \ldots$ durch

01-Folgen $k(a), k(a'), \ldots$ codieren.

Dabei soll so etwas wie

$$k(a) = 001,\ k(a') = 00101$$

ausgeschlossen sein.

Definition:

Eine Abbildung

$$k : S \to \bigcup_{t \geq 1} \{0,1\}^t$$

heißt **binärer Präfixcode**,

wenn kein $k(a)$ Anfangsstück irgendeines $k(a')$, $a \neq a'$, ist.

Ist $k(a) = k_1(a) \ldots k_t(a)$, dann nennt man

$$\ell(a) := t$$

die *Länge* (oder auch die Anzahl der Bits)

des *Codeworts* $k(a)$.

Binäre Präfixcodes kann man sich auch als

(planare (d.h. “in die Ebene gezeichnete”))

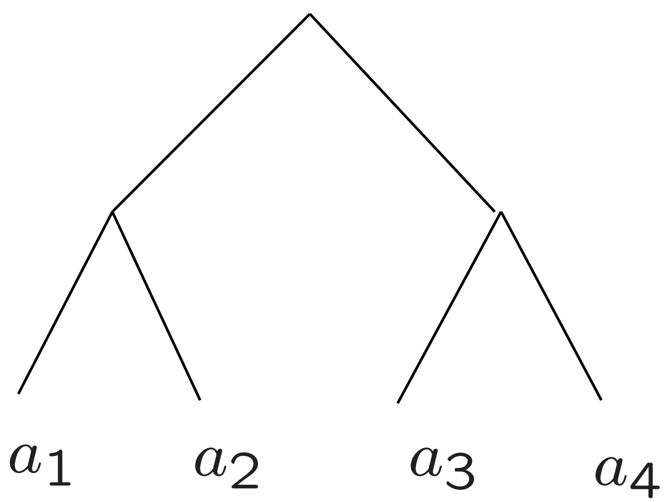
Binärbäume vorstellen,

deren Blätter bijektiv

mit den Buchstaben des Alphabets beschriftet sind.

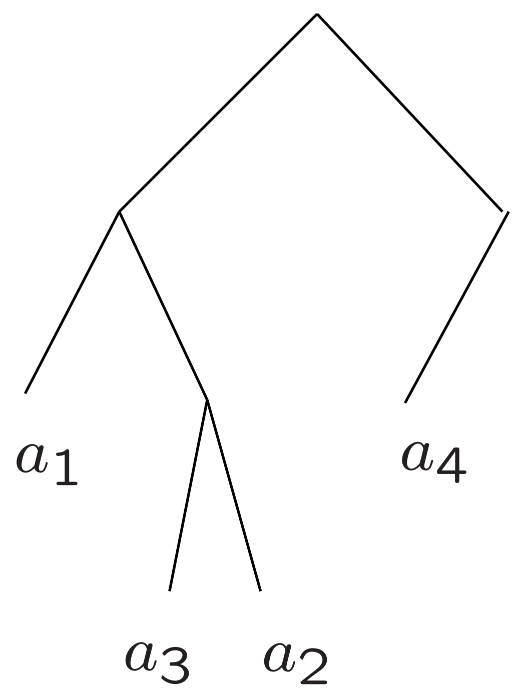
Beispiel: $S = \{a_1, a_2, a_3, a_4\}$

$$k(a_1) := 00,\ k(a_2) := 01,\ k(a_3) := 10,\ k(a_4) := 11.$$



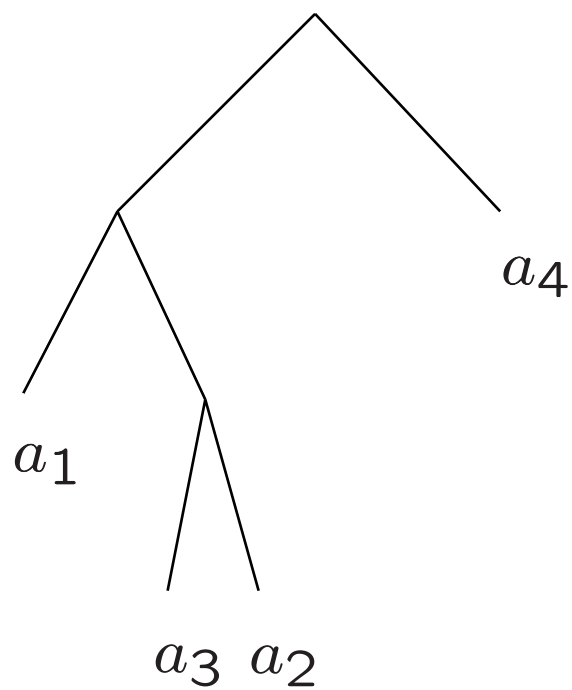
Beispiel: $S = \{a_1, a_2, a_3, a_4\}$

$$k(a_1) := 00,\ k(a_2) := 011,\ k(a_3) := 010,\ k(a_4) := 10.$$



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

$$k(a_1) := 00,\ k(a_2) := 011,\ k(a_3) := 010,\ k(a_4) := 1.$$



Wir formulieren die (bereits oben bewiesenen)
zwei Aussagen um die Fano-Kraft-Ungleichung
jetzt noch einmal für binäre Präfixcodes:

Die Fano-Kraft-Ungleichung für binäre Präfixcodes:

(Teil 1) Für jeden binären Präfixcode k gilt:

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

"Je mehr Buchstaben man codieren will,
um so länger müssen die Codewörter sein."

(Teil 2) Ist $\ell : S \to \mathbb{N}$ eine Abbildung mit $\sum_{a \in S} 2^{-\ell(a)} \leq 1$,

dann gibt es einen binären Präfixcode,

für den $\ell(a)$ die Länge von $k(a)$ ist für alle $a \in S$.

"Jede nicht allzu kurzwertige Abbildung ℓ tritt als
Codewortlängenabbildung auf."