Vorlesung 13b

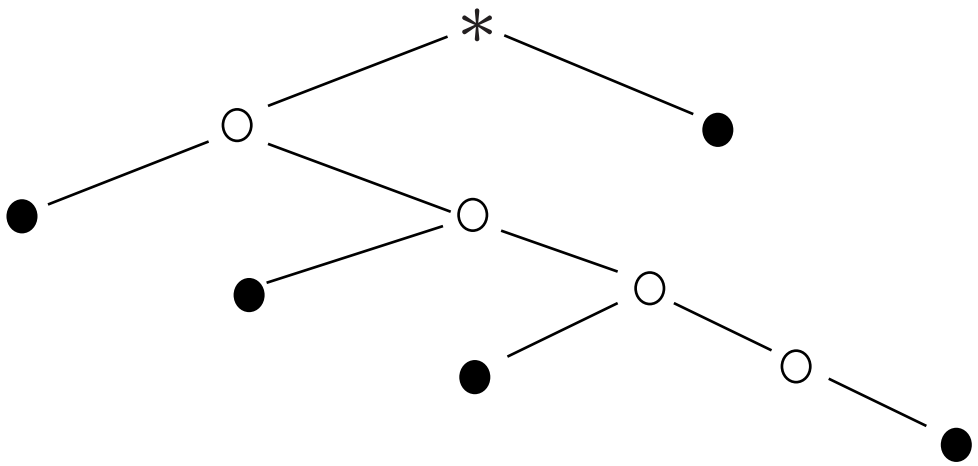
Quellencodieren und Entropie

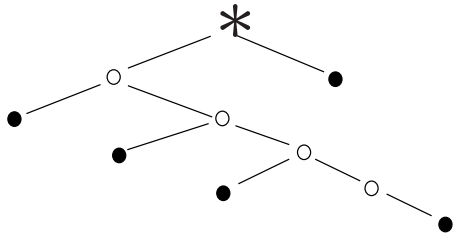
1. Die gewöhnliche Irrfahrt

auf einem vollen Binärbaum

(aufgefasst als gerichteter Graph)

von der Wurzel zu den Blättern





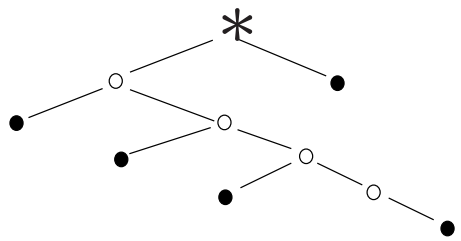
Merkmale eines binären Baumes:

Ein Knoten ist ausgezeichnet als *Wurzel*.

Jeder Knoten (außer der Wurzel) hat genau einen Vorgänger.

Jeder Knoten hat zwei, einen oder keinen Nachfolger.

Die Knoten ohne Nachfolger heißen *Blätter*,
die mit Nachfolger heißen *innere Knoten*.

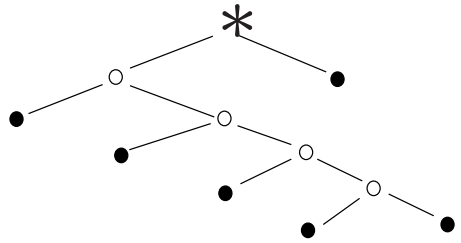


Wir sprechen von einem *vollen Binärbaum,*

wenn jeder innere Knoten zwei Nachfolger hat.

Jeder Binärbaum lässt sich “auf minimale Weise”

zu einem vollen Binärbaum ergänzen.



Wir sprechen von einem *vollen Binärbaum,*

wenn jeder innere Knoten zwei Nachfolger hat.

Jeder Binärbaum lässt sich “auf minimale Weise”

zu einem vollen Binärbaum ergänzen.

Sei T ein voller Binärbaum.

Die gewöhnliche Irrfahrt auf T

- macht von jedem Knoten einen Schritt zu einem zufällig gewählten seiner beiden Nachfolger

- und endet in einem der Blätter.

Wenn nicht anderes vermerkt, nehmen wir immer an,
dass die Irrfahrt in der Wurzel startet.

Sei T ein voller Binärbaum auf T

und Y die gewöhnliche Irrfahrt auf T

Dann gilt für jedes Blatt b von T mit Tiefe $t(b)$:

$$\mathbf{P}(Y \text{ endet in } b) = 2^{-t(b)}$$

(denn auf dem Weg von der Wurzel zu b muss $t(b)$-mal

das “richtige” Münzwurfergebnis kommen)

Folgerung:

Für jeden vollen Binärbaum T mit Blattmenge B gilt:

$$\sum_{b \in B} 2^{-t(b)} = 1.$$

Ist T nicht voll, dann gilt die entsprechende Gleichheit

für jede "Vervollständigung" T' von T.

Bei T' kommen dann noch Blätter dazu,

also ist dann

$$\sum_{b \in B} 2^{-t(b)} < 1.$$

Dies fassen wir zusammen in der

Bemerkung:

Für jeden Binärbaum T

mit Blattmenge B und Blatttiefen $t(b)$ gilt:

$$\sum_{b \in B} 2^{-t(b)} \leq 1,$$

mit Gleichheit genau dann wenn T voll ist.

Wir fragen:

Welche Bedingung an Zahlen $l_1, \ldots, l_m$

ist notwendig und hinreichend dafür,

dass diese Zahlen

als Blatttiefen eines binären Baumes auftreten?

Die Antwort gibt die

Fano-Kraft Ungleichung für binäre Bäume:

$l_1, \ldots, l_m$ seien natürliche Zahlen.

Genau dann gibt es einen binären Baum T,

für den die $l_1, \ldots, l_m$ die Tiefen seiner Blätter sind,

wenn gilt:

$$(*) \quad \sum_{i=1}^{m} 2^{-l_i} \leq 1.$$

Beweis:

" $\Longrightarrow$" siehe die Bemerkung auf Folie 10.

" $\Longleftarrow$": mit Induktion: Für $m = 1$ ist die Behauptung klar.
Schritt vom m zu $m + 1$:

Seien $l_1 \leq l_2 \leq \cdots \leq l_{m+1}$ natürliche Zahlen mit $\sum_{i=1}^{m+1} 2^{-l_i} \leq 1$.

Nach Induktionsvoraussetzung gibt es einen Binärbaum T_m mit Blattiefen $l_1, \ldots, l_m$. Nach der Bemerkung auf Folie 10 ist T_m nicht voll, besitzt also einen inneren Knoten (in Tiefe $< \ell_m$) mit nur einem Nachfolger.
Zu diesem Knoten fügen wir einen zweiten Nachfolger hinzu und setzen (falls $\ell_{m+1} > \ell_m$) seine "Nachkommenslinie" fort, bis wir bei der Tiefe ℓ_{m+1} angelangt sind.
Damit ist ein Binärbaum mit den Blatttiefen $l_1, \ldots, l_{m+1}$ konstruiert. □

2. Binäre Präfixcodes

als Blattbeschriftungen von binären Bäumen

S sei eine endliche (oder abzählbar unendliche) Menge

(ein “Alphabet”).

Die Elemente von S nennen wir *Buchstaben*.

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

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

Dabei soll so etwas ausgeschlossen sein:

$$k(a) = 001,\ k(b) = 00101.$$

Definition:

Eine Abbildung

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

heißt **(binärer) Präfixcode**,

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

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

$$\ell(a) = l$$

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

des *Codeworts* $k(a)$.

Binäre Präfixcodes kann man sich auch als

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

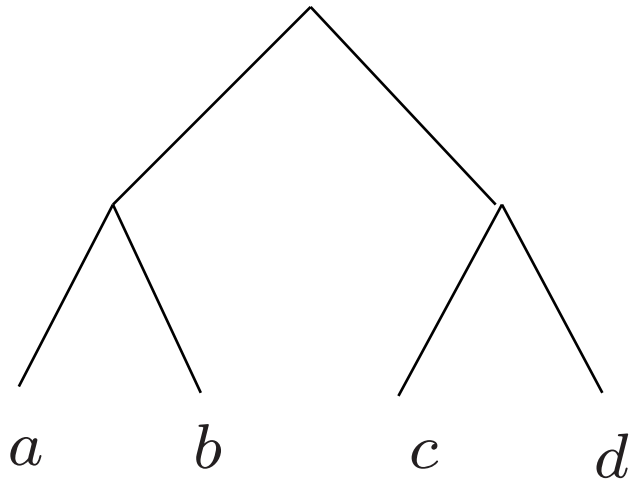
binäre Bäume vorstellen,

deren Blätter bijektiv

mit den Buchstaben des Alphabets beschriftet sind.

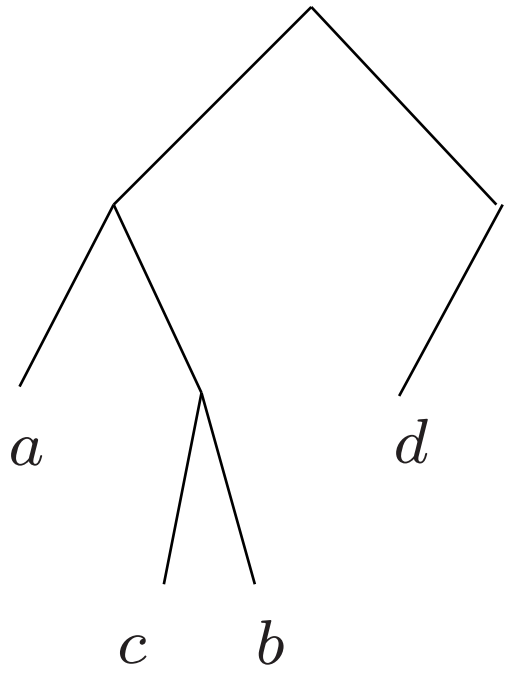
Beispiel: $S = \{a, b, c, d\}$

$$k(a) := 00, \;\; k(b) := 01, \;\; k(c) := 10, \;\; k(d) := 11.$$



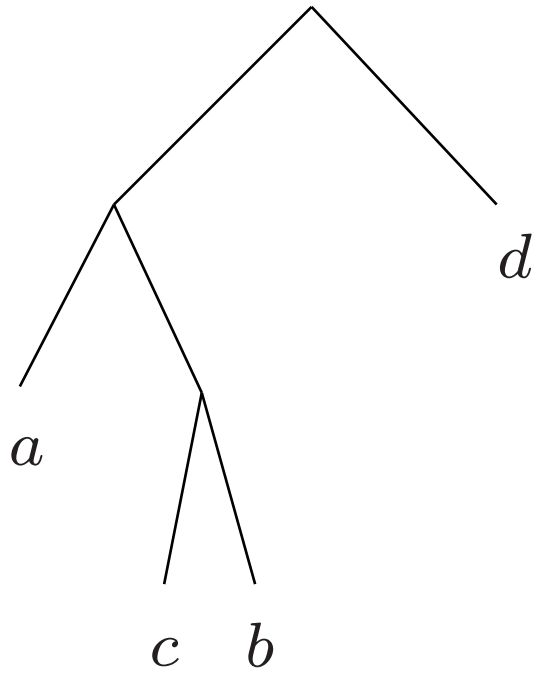
Beispiel: $S = \{a, b, c, d\}$

$$k(a) := 00, \quad k(b) := 011, \quad k(c) := 010, \quad k(d) := 10.$$



Beispiel: $S = \{a, b, c, d\}$

$$k(a) := 00, \quad k(b) := 011, \quad k(c) := 010, \quad k(d) := 1.$$



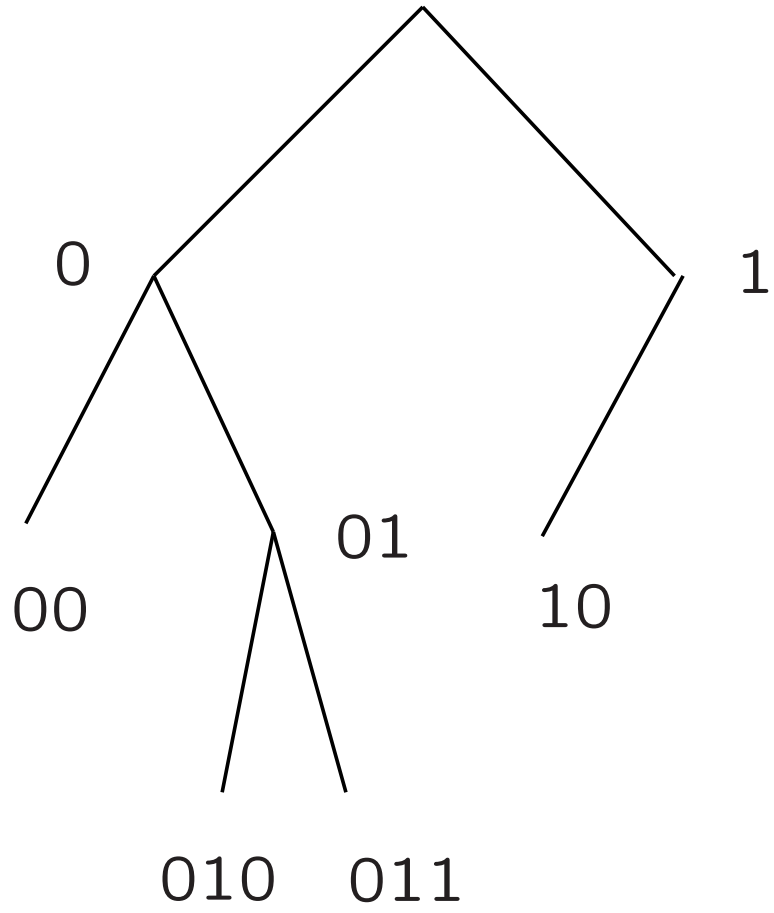
Regeln für die Beschriftung der Knoten mit 01-Wörtern:

Die Wurzel trägt das Wort der Länge 0 (das “leere Wort”).

Das Wort an einem Knoten (außer der Wurzel)
setzt das Wort am Vorgängerknoten um ein Bit fort
(die Wörter in der Tiefe l haben somit die Länge l).

Verschiedene Knoten tragen verschiedene Wörter.

Beispiel:



Wir formulieren die (bereits oben bewiesenen)
zwei Aussagen um die Fano-Kraft-Ungleichung
jetzt noch einmal für binäre Präfixcodes
und beweisen sie nochmal “am Stück”:

Die Fano-Kraft-Ungleichung (Teil 1)

Für jeden binären Präfixcode k gilt:

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

Beweis durch ein Gedankenexperiment:

Erzeuge per Münzwurf 01-Folgen und stoppe,

sobald eines der Codewörter vollendet ist.

Wegen der Präfixeigenschaft

schließen sich diese Ereignisse paarweise aus. Also gilt:

$$1 \geq \mathbf{P}(\text{man trifft schließlich auf ein Codewort})$$

$$= \sum_{a \in S} 2^{-\ell(a)}. \quad \square$$

Die Fano-Kraft-Ungleichung (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 $a \in S$.

Beweis: Wir ordnen die $a_1, a_2, \ldots$ so, dass

$$\ell(a_1) \leq \ell(a_2) \leq \cdots .$$

Angenommen $a_1, \ldots, a_m$ ist schon codiert,

aber a_{m+1} noch nicht. Dann ist $\sum_{i=1}^{m} 2^{-\ell(a_i)} < 1$, also

$$1 - \sum_{i=1}^{m} 2^{-\ell(a_i)} \geq 2^{-\ell(a_m)} \geq 2^{-\ell(a_{m+1})}.$$

Daraus folgt: Bei einem Münzwurf der Länge $\ell(a_{m+1})$

wirft man mit W'keit $2^{-\ell(a_{m+1})}$ ein Wort w,

das keines der $k(a_1), \ldots, k(a_m)$ als Anfangsstück enthält.

Setze dann $k(a_{m+1}) := w$. $\square$

Die Ungleichung von Fano und Kraft
nochmal formuliert für binäre Bäume:

S sei eine abzählbare Menge
und $\ell(a),\ a \in S$, seien natürliche Zahlen.

Genau dann gibt es einen binären Baum, dessen Blätter
bijektiv mit den Elementen $a \in S$ beschriftet sind
und Tiefen $\ell(a)$ haben,

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

3. Sparsames Codieren zufälliger Buchstaben.

Sei X eine S-wertige Zufallsvariable

(ein “zufälliger Buchstabe”)

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

Gefragt ist nach einem binären Präfixcode,

dessen erwartete Codelänge

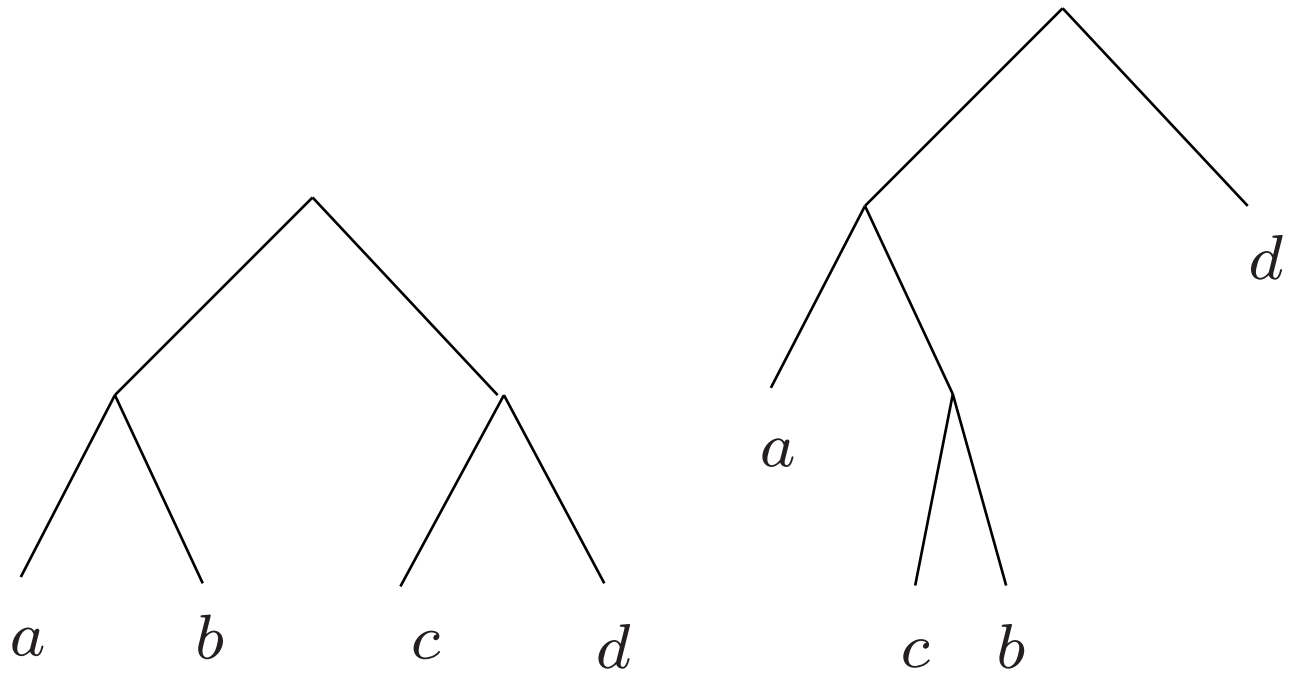
$$\mathbf{E}[\ell(X)] = \sum_{a \in S} \ell(a)\rho(a)$$

möglichst klein ist.

Beispiel: $S = \{a, b, c, d\}$.

Sind die vier Ausgänge gleich wahrscheinlich,

dann ist der Code links günstiger als der rechts.



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

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

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

dann ist es am besten, alle 2^l Buchstaben

mit den 01-Folgen der Länge l zu codieren.

In diesem Fall gilt

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

4. Shannon-Codes und der Quellencodierungssatz

Ein *Shannon-Code* ist ein Präfixcode,

bei dem jeder Buchstabe a 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.

Wir werden zeigen:

Shannon-Codes

verfehlen die minimal mögliche erwartete Codelänge

um höchstens ein Bit.

Sei dazu

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

mit $0 \log 0 := 0$.

Satz

(Quellencodierungssatz von Shannon)

a) Für jeden binären Präfixcode gilt

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

b) Für binäre Shannon-Codes gilt außerdem

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

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 b) des Satzes.

Jetzt zu Teil a):

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

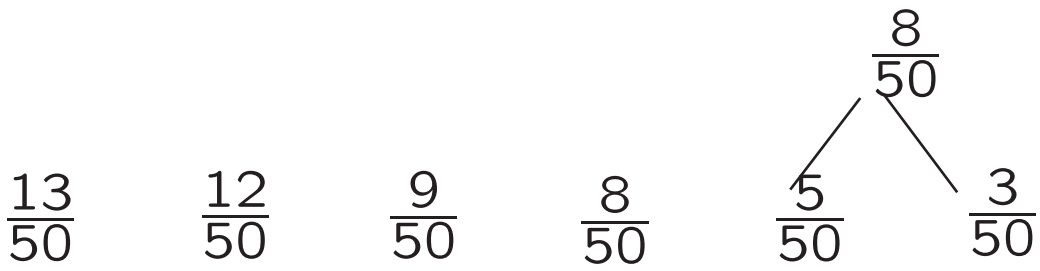
$$\mathbf{H}_2[X] - \mathbf{E}[\ell(X)] \leq 0. \quad \square$$

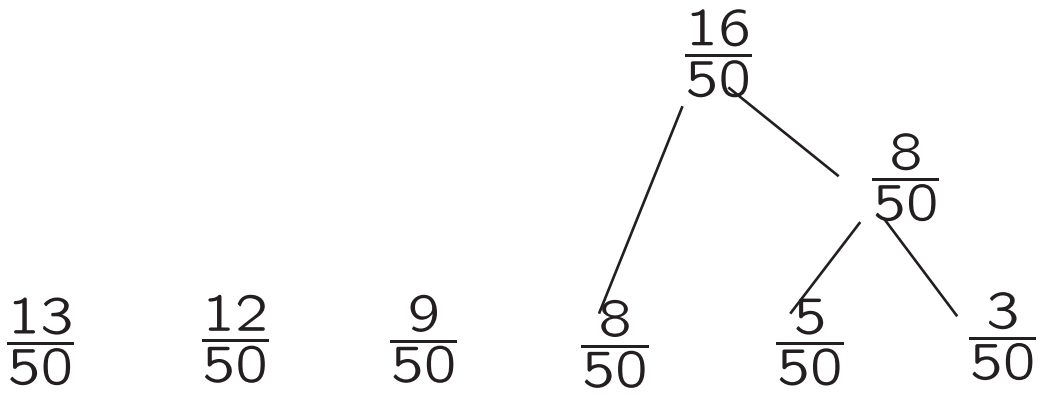
5. Huffman-Codes

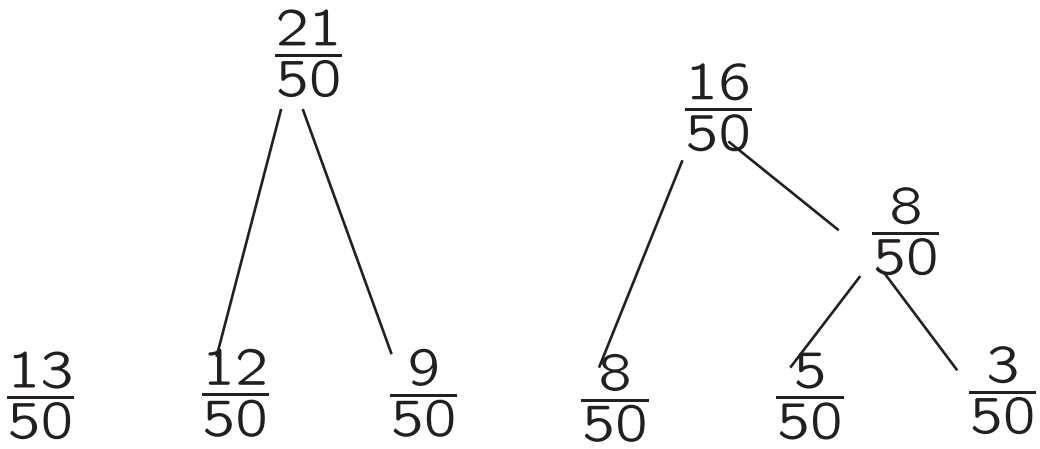
Shannon-Codes sind wegweisend für den
fundamentalen Begriff der Entropie (siehe Abschnitt 6).

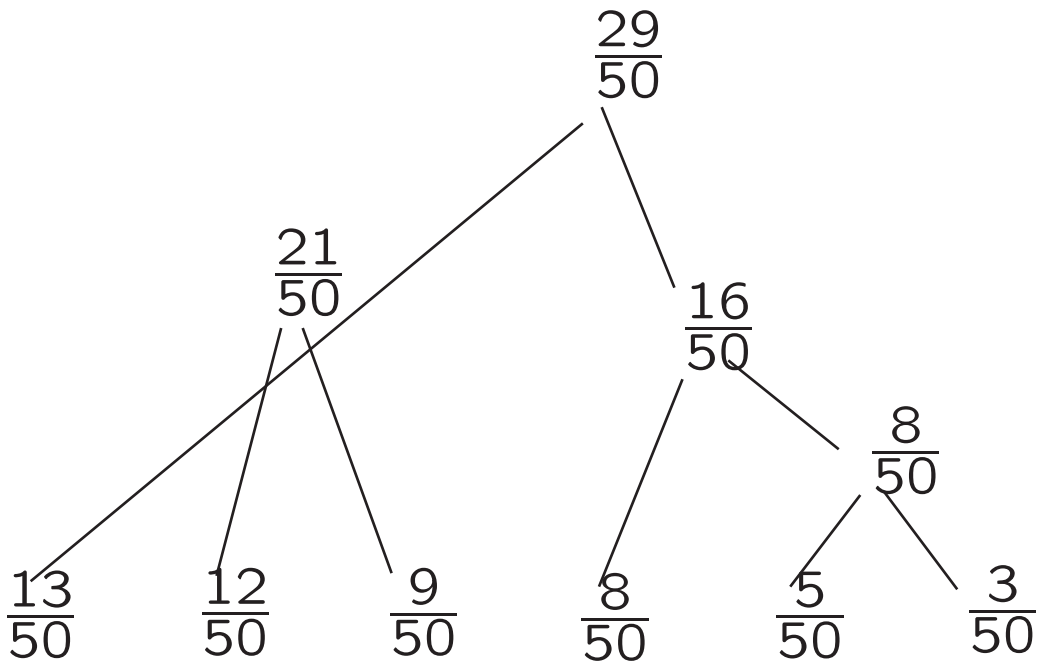
Demgegenüber bieten die
(konzeptionell weniger tiefgreifenden) *Huffman-Codes*
eine einfache Möglichkeit,
Codes mit kürzester erwartet Wortlänge
“von den Blättern zur Wurzel” zu *konstruieren*,
durch sukzessives Verschmelzen von Paaren kleinster W’keit
(siehe Buch S. 145).
Wir geben hier nur ein Beispiel:

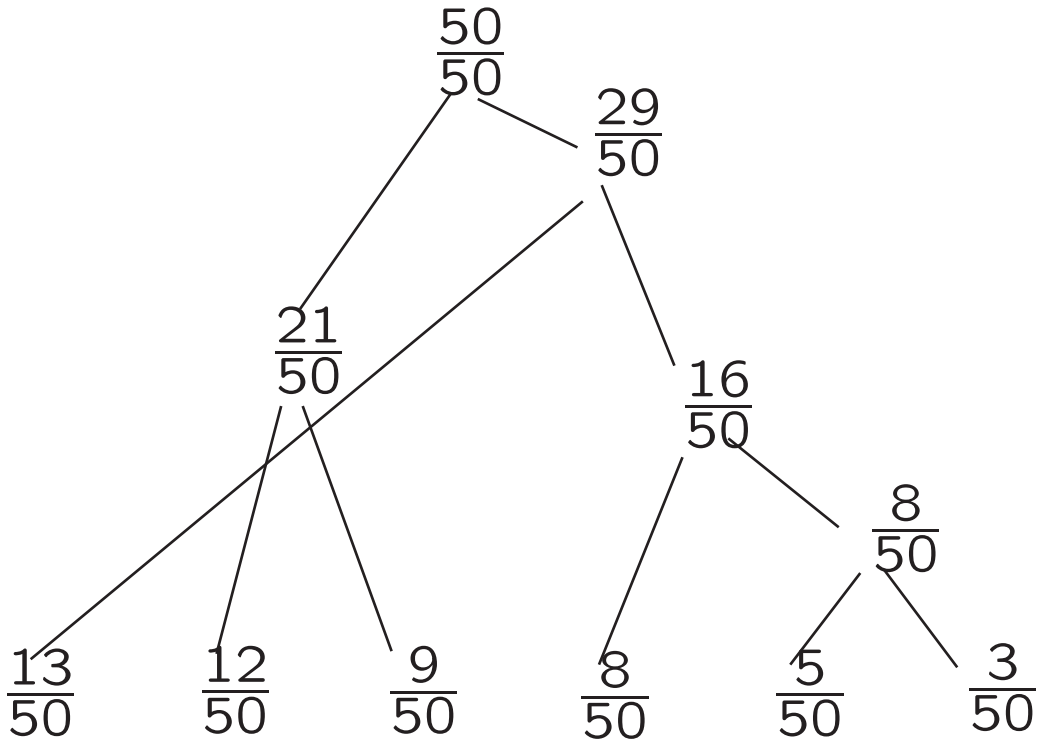
$\frac{13}{50}$ $\frac{12}{50}$ $\frac{9}{50}$ $\frac{8}{50}$ $\frac{5}{50}$ $\frac{3}{50}$

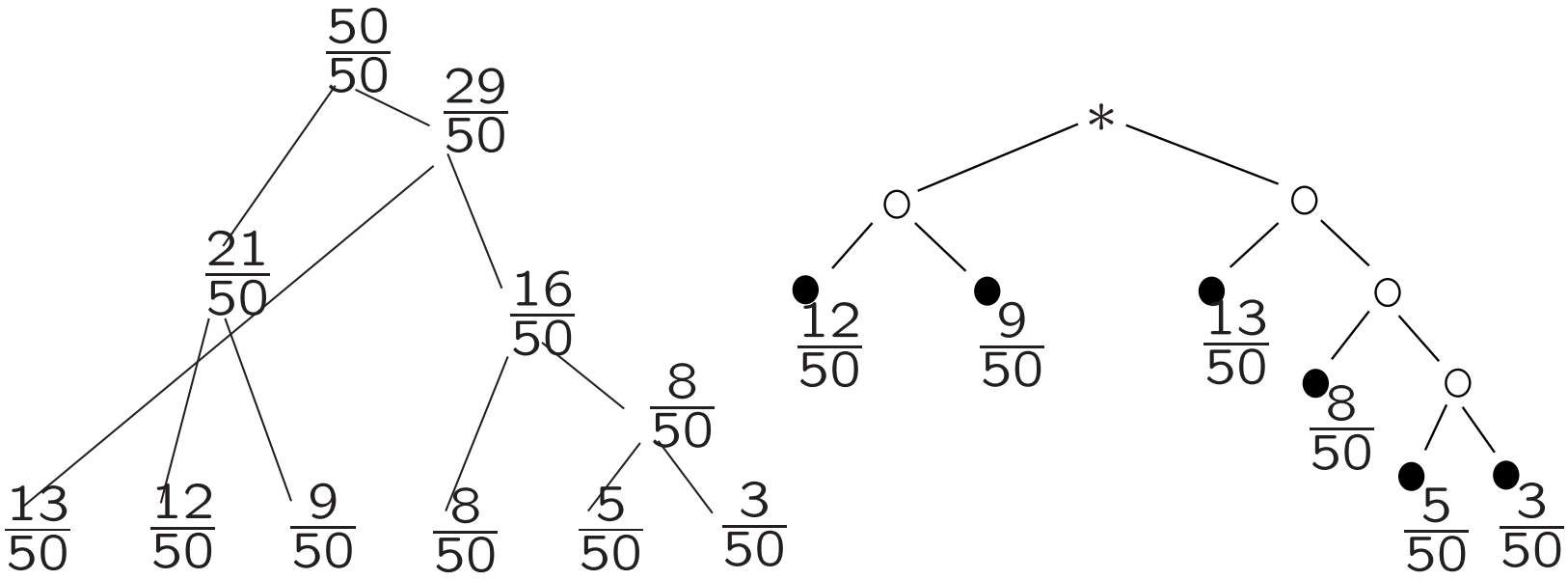

13/50
12/50
9/50
8/50
8/50
5/50
3/50


16/50
8/50
13/50
12/50
9/50
8/50
5/50
3/50


21/50
13/50
12/50
9/50
16/50
8/50
8/50
5/50
3/50


29/50
21/50
16/50
8/50
13/50
12/50
9/50
8/50
5/50
3/50


50/50
29/50
21/50
16/50
8/50
13/50
12/50
9/50
8/50
5/50
3/50


50/50
29/50
21/50
16/50
8/50
13/50
12/50
9/50
8/50
5/50
3/50
*
12/50
9/50
13/50
8/50
5/50
3/50

6. Die Entropie

Die im Quellencodierungssatz auftretende Größe

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

heißt die **Entropie** von X (zur Basis 2).

Es ist sinnvoll, auch andere Basen zu erlauben:

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

Wir schreiben für beliebige, fest gewählte Basis b:

$$\log := \log_b, \quad \mathbf{H}[X] := \mathbf{H}_b[X]$$

$$\boxed{\mathbf{H}[X] = -\sum_{a \in S} \rho(a) \log \rho(a)}$$

Die Entropie ist *der* fundamentale Begriff

der Informationstheorie.

Nach dem Quellenkodierungssatz gibt die 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.