Grundzüge der Informationstheorie (2)

Bisherige Erkenntnisse zum Informationsgehalt:

- Der Informationsgehalt h eines Zeichens (einer Nachricht) ist um so höher, je unwahrscheinlicher sein Auftreten ist:

  $h = \mathrm{ld}\ [\ 1/p\ ] = -\mathrm{ld}\ p$

- Treten die n Zeichen eines Zeichenvorrats mit ungleichen Wahrscheinlichkeiten p_i auf ($i=1,\dots,n$; $\Sigma p_i=1$), so versteht man unter dem mittleren Informationsgehalt H die Summe der mit den individuellen Wahrscheinlichkeiten p_i gewichteten Informationsgehalte h_i der einzelnen Zeichen:

  $H = \Sigma\ (p_i \cdot h_i) = \Sigma\ (p_i \cdot \mathrm{ld}\ [1/p_i]) = -\Sigma\ (\ p_i \cdot \mathrm{ld}\ p_i\ ) > 0$

- H nennt man auch „Entropie“ der Informations- oder Nachrichtenquelle, die die Zeichen entsendet (Parallele zur Thermodynamik); seine Einheit ist bit/Zeichen.

- In einem binären System gilt: $n=2$; $p_1 = p \Rightarrow p_2 = 1 - p$

Beispiel:

Mittlerer Informationsgehalt der Zeichen einer binären Quelle

- bei gleicher Auftretungswahrscheinlichkeit p beider Zeichen:
  $H = h = \mathrm{ld}\ (1/p_i) = \mathrm{ld}\ 2 = 1$ bit
- bei Auftretungswahrscheinlichkeiten p=1/3; (1–p)=2/3:
  $H = \Sigma\ (p_i \cdot \mathrm{ld}[1/p_i]) = 1/3 \cdot (\mathrm{ld}\ 3 - \mathrm{ld}\ 1) + 2/3 \cdot (\mathrm{ld}\ 3 - \mathrm{ld}\ 2)$
  $= 1/3 \cdot (\lg 3 / \lg 2 - 0) + 2/3 \cdot (\lg 3 / \lg 2 - 1)$
  $= \lg 3 / \lg 2 - 2/3 = (0{,}477 / 0{,}301) - 2/3$
  $= 0{,}918$ bit
- bei Auftretungswahrscheinlichkeiten p=1/4; (1–p)=3/4:
  $H = \Sigma\ (p_i \cdot \mathrm{ld}[1/p_i]) = 0{,}25 \cdot (\mathrm{ld}\ 4 - \mathrm{ld}\ 1) + 0{,}75 \cdot (\mathrm{ld}\ 4 - \mathrm{ld}\ 3)$
  $= 0{,}25 \cdot (2 - 0) + 0{,}75 \cdot (2 - \lg 3 / \lg 2)$
  $= 0{,}5 + 0{,}75 \cdot (2 - 0{,}477 / 0{,}301)$
  $= 0{,}811$ bit

Grundzüge der Informationstheorie (2)

- Auftretungswahrscheinlichkeit p, die den mittleren Informationsgehalt einer Nachrichtenquelle maximiert:

  $H = -p \cdot \mathrm{ld}\, p - (1-p) \cdot \mathrm{ld}\,(1-p)$

  $= 1/\ln 2 \,\{ -p \cdot \ln p - \ln(1-p) + p \cdot \ln(1-p) \}$

  Extremwert, wo die 1. Ableitung d/dp =0 wird; hierbei:

  $d(p \cdot \ln p) / dp = d(p)/dp \cdot \ln p + p \cdot d(\ln p)/dp$

  $= \ln p + 1$

  d(ln x)/dx = 1/x

  $d(\ln(1-p)) / dp = d(\ln(1-p)) / d(1-p) \cdot d(1-p)/dp$

  $= 1 / (1-p) \cdot (-1)$

  $= -1 / (1-p)$

  $d(p \cdot \ln(1-p)) / dp = d(p)/dp \cdot \ln(1-p) + p \cdot d(\ln(1-p))/dp$

  $= \ln(1-p) - p/(1-p)$

  $dH/dp = (1/\ln 2) \cdot \{ -\ln p - 1 + 1/(1-p) + \ln(1-p) - p/(1-p) \}$

  $= (1/\ln 2) \cdot \{ \ln(1-p) - \ln p \}$

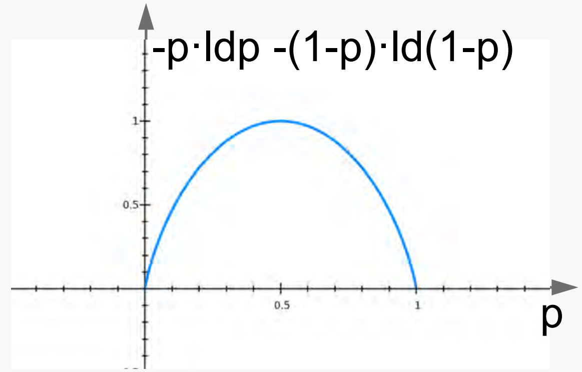
  $= (1/\ln 2) \cdot \ln [(1-p) / p]$

Grundzüge der Informationstheorie (2)

Erste Ableitung =0. wenn:

$dH/dp = (1/\ln 2) \cdot \ln [(1-p) / p] = 0$, bzw. $\ln [(1-p) / p] = 0$

d.h., wenn $(1-p) / p = 1$ oder wenn $p=½$



Zweite Ableitung ist < 0, denn:

$$d^2H/dp^2 = (1/\ln 2) \cdot d \{ \ln(1-p) - \ln p \} / dp$$
$$= (1/\ln 2) \cdot \{-1/(1-p) - 1/p \}$$
$$= (1/\ln 2) \cdot 1/ [p \cdot (p - 1)] < 0 \text{ (weil } 0 < p \leq 1)$$

Das heißt:

Der mittlere Informationsgehalt einer Binärquelle ist maximal, wenn beide Zeichen gleich wahrscheinlich sind.

Auch bei Quellen mit mehreren Zeichen ist H maximal, wenn alle verfügbaren Zeichen gleich wahrscheinlich sind (hier ohne Beweis).

Grundzüge der Informationstheorie (2)

Ungleiche Auftretungswahrscheinlichkeiten binär codierter (z.B.: Schrift-) Zeichen können technisch genutzt werden.

Beispiel:

Eine Quelle erzeuge 4 Zeichen: a, b, c, d, mit verschiedenen Wahrscheinlichkeiten in zwei Anwendungen α und β:

Fall α: $p_\alpha (a) = 1/4$; $p_\alpha (b) = 1/4$; $p_\alpha (c) = 1/4$; $p_\alpha (d) = 1/4$

Fall β: $p_\beta (a) = 1/2$; $p_\beta (b) = 1/4$; $p_\beta (c) = 1/8$; $p_\beta (d) = 1/8$

Mittlerer Informationsgehalt der Zeichen in den zwei Fällen:

- Fall α:
  $H_\alpha = \Sigma\,(p_i \cdot \mathrm{ld}[1/p_i]) = 4\,(1/4 \cdot \mathrm{ld}4) = \mathrm{ld}\,4 = \underline{2\text{ bit/Zeichen}}$
- Fall β:
  $H_\beta = \Sigma\,(p_i \cdot \mathrm{ld}[1/p_i]) = 1/2 \cdot \mathrm{ld}\,2 + 1/4 \cdot \mathrm{ld}\,4 + 2/8 \cdot \mathrm{ld}\,8$
  $= 1/2 + 1/2 + 3 \cdot 2/8 = \underline{1{,}75\text{ bit/Zeichen}}$

Grundzüge der Informationstheorie (2)

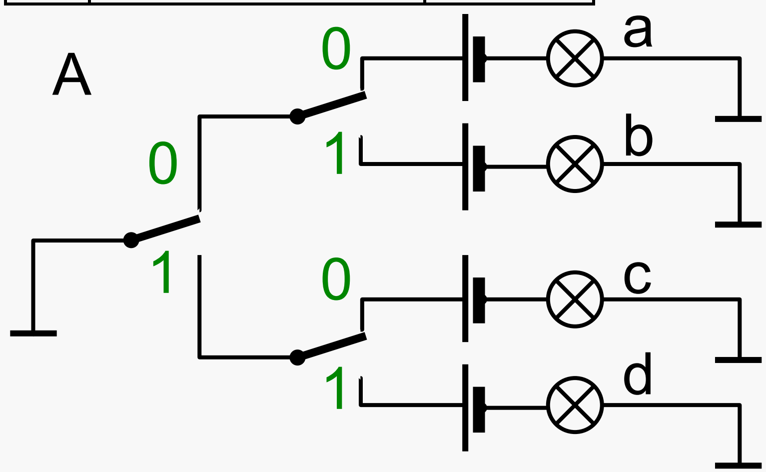
Fall α: p_α (a) = 1/4; p_α (b) = 1/4; p_α (c) = 1/4; p_α (d) = 1/4

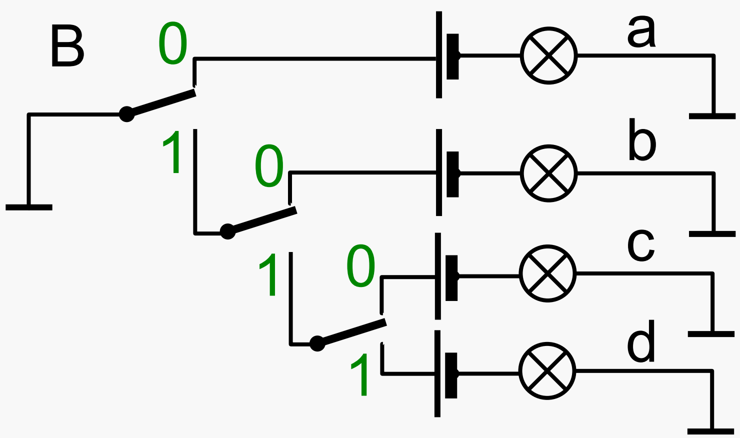
Fall β: p_β (a) = 1/2; p_β (b) = 1/4; p_β (c) = 1/8; p_β (d) = 1/8

Prinzipiell Schaltungen (Binärbäume) mit ungleicher Anzahl von Binärstellen (Schaltvorgängen) *m* je Zeichen möglich:

A:	Binärstellen	Code
a	2	0 0
b	2	0 1
c	2	1 0
d	2	1 1

B:	Binärstellen	Code
a	1	0
b	2	1 0
c	3	1 1 0
d	3	1 1 1





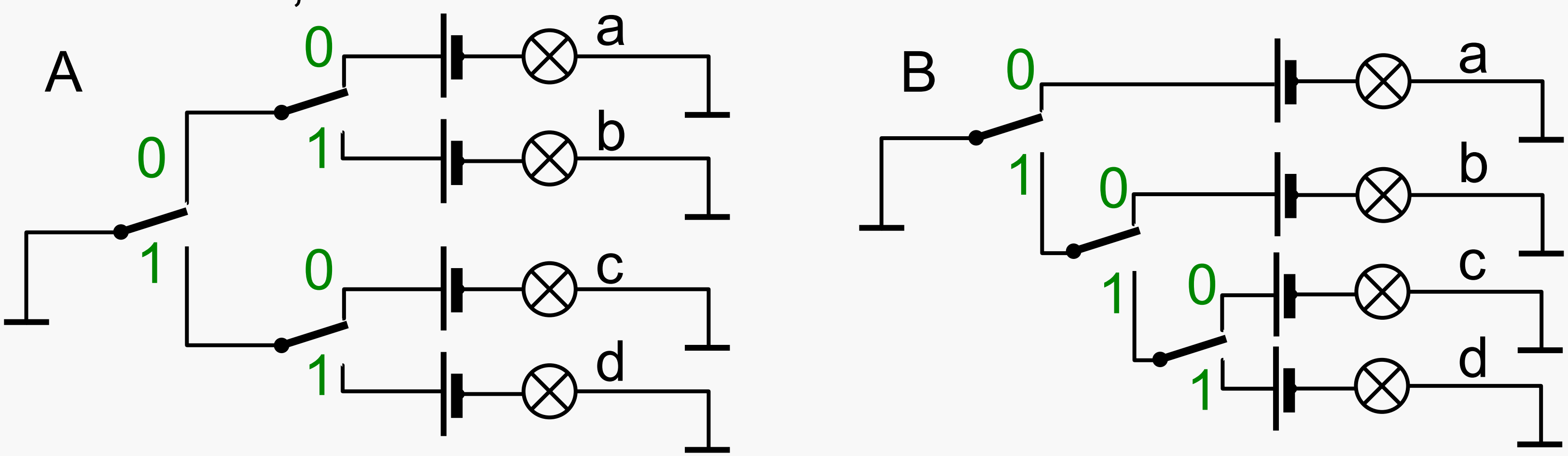
Grundzüge der Informationstheorie (2)

- Fall α: $p_\alpha(a) = 1/4$; $p_\alpha(b) = 1/4$; $p_\alpha(c) = 1/4$; $p_\alpha(d) = 1/4$
  Mittlere Anzahl von Binärstellen jeder Realisierung:

A: $m_{\alpha A} = p_\alpha(a) \cdot m_A(a) + p_\alpha(b) \cdot m_A(b) + p_\alpha(c) \cdot m_A(c) + p_\alpha(d) \cdot m_A(d)$
$= 4 \cdot (1/4 \cdot 2)$
$= 2$ bit / Zeichen $(= H_\alpha)$

Realisierung A günstiger für α !

B: $m_{\alpha B} = p_\alpha(a) \cdot m_B(a) + p_\alpha(b) \cdot m_B(b) + p_\alpha(c) \cdot m_B(c) + p_\alpha(d) \cdot m_B(d)$
$= 1/4 \cdot 1 + 1/4 \cdot 2 + 2 \cdot 1/4 \cdot 3$
$= 2{,}25$ bit / Zeichen

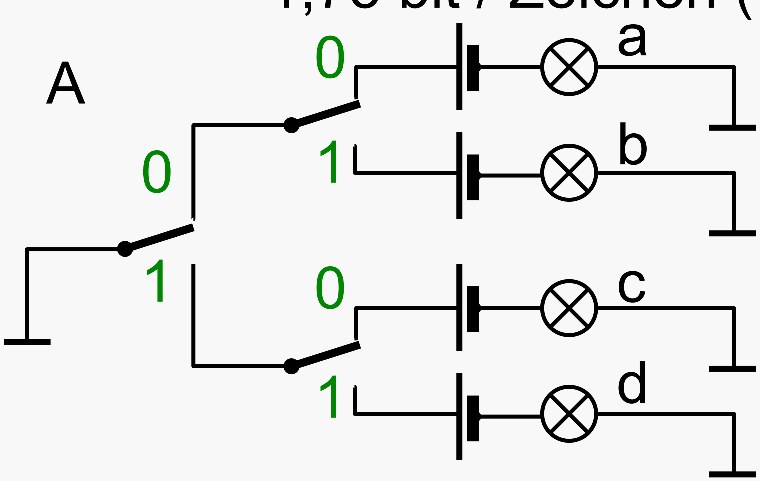


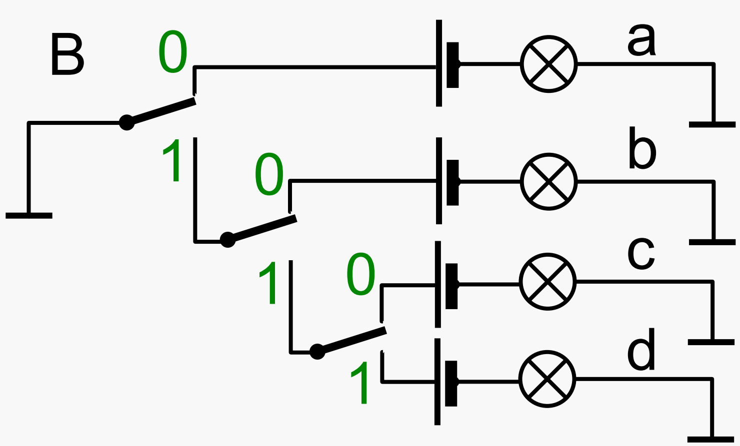
- Fall β: p_β (a) = 1/2; p_β (b) = 1/4; p_β (c) = 1/8; p_β (d) = 1/8
  Mittlere Anzahl von Binärstellen jeder Realisierung:

A: $m_{\beta A} = p_\beta(a)\cdot m_A(a)+p_\beta(b)\cdot m_A(b)+p_\beta(c)\cdot m_A(c)+p_\beta(d)\cdot m_A(d)$
$= (1/2 + 1/4 + 2/8) \cdot 2$
$=$ 2 bit / Zeichen

B: $m_{\beta B} = p_\beta(a)\cdot m_B(a)+p_\beta(b)\cdot m_B(b)+p_\beta(c)\cdot m_B(c) + p_\beta(d)\cdot m_B(d)$
$= 1/2 \cdot 1 + 1/4 \cdot 2 + 2 \cdot 1/8 \cdot 3$
$=$ 1,75 bit / Zeichen (= H_β)

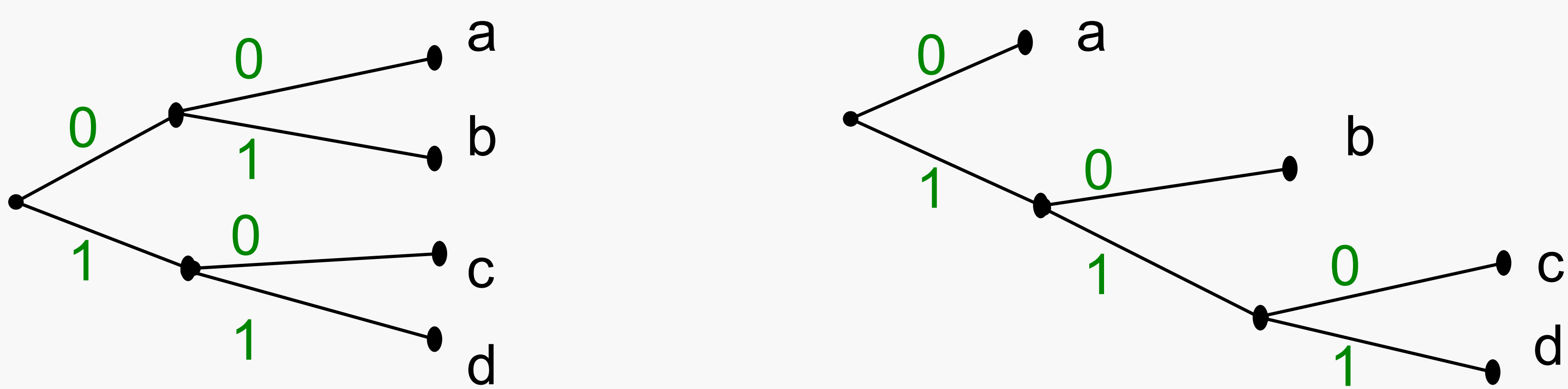
Realisierung B günstiger für β !

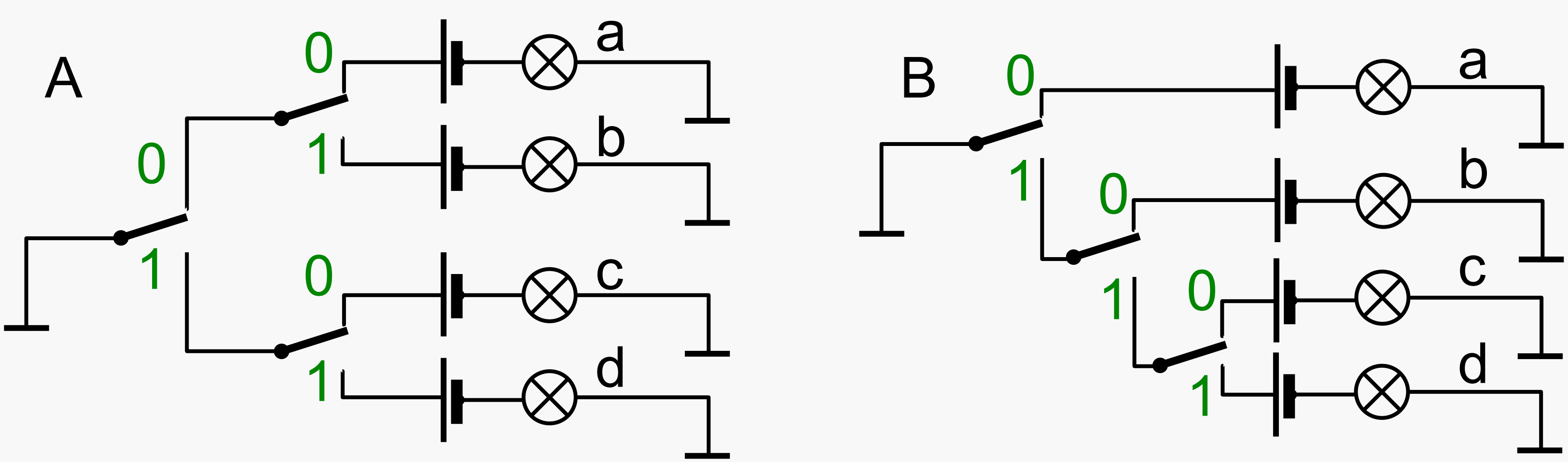




Grundzüge der Informationstheorie (2)

Die bislang genutzten Schalterstellungen („Relaispyramiden“) lassen sich auch abstrakt als binäre Entscheidungsbäume, sog. „Codebäume“ darstellen. Die beiden Darstellungen sind äquivalent.





Grundzüge der Informationstheorie (2)

- Alle Codierungsverfahren benötigen eine mittlere Anzahl von Binärstellen m, die gleich oder größer dem mittleren Informationsgehalt H ist.
- Bestrebung: Entwurf von Codierungsverfahren zur Erlangung einer mittleren Anzahl m von Binärstellen, die den mittleren Informationsgehalt nicht um mehr als eine kleine, vorgegebene Zahl ε übersteigt:

$$H \leq m \leq H + \varepsilon \qquad \text{(„Codierungssatz“)}$$

- Als Gütekriterium gilt die (absolute) Redundanz

$$R = m - H$$

- Besonders interessant sind die Codierungsverfahren

  nach Robert Fano (Italien/USA, geb. 1917),

  nach David Albert Huffman (USA, 1925 – 1999)

  [Eine Weiterentwicklung der Huffman-Codierung wird zur Bild-Kompression beim JPEG- und MPEG-Verfahren eingesetzt.]

Grundzüge der Informationstheorie (2)

Codierung nach Fano:

Auflistung in benachbarten Spalten

- aller zulässigen Zeichen (nach abnehmender Wahrscheinlichkeit)
- ihrer Auftrittswahrscheinlichkeit
- der Teilsummen ab Wahrscheinlichkeit 0 bis zum jeweiligen Zeichen.

Beispiel: Fall β (s.o.)

x_i	$p(x_i)$	$\Sigma p(x_i)$	Code	$m(x_i)$
a	0,5	1,0	0	1
b	0,25	0,5	1 0	2
c	0,125	0,25	1 1 0	3
d	0,125	0,125	1 1 1	3

- Sukzessive Aufteilung der Tabelle möglichst genau bei den Teilsummen $1/2^n$ (je 50%) und Zuordnung (v.l.n.r.) von Nullen (0) den oberen (wahrscheinlicheren) und Einsen (1) den unteren Nachrichten, bis die Hälften je nur ein Zeichen enthalten.

  Die Länge der Codewörter nimmt von oben nach unten zu.

Codierung nach Huffman:

- Auflistung aller zulässigen Zeichen nach abnehmender Wahrscheinlichkeit
- Sukzessive Zusammenfassung der zwei Zeichen mit den jeweils niedrigsten Auftretungswahrscheinlichkeiten zu einem neuen Zeichen
- Aufstellung einer neuen Liste ohne die eliminierten Zeichen unter Einreihung des neuen Zeichens mit der aufsummierten Wahrscheinlichkeit
- Wiederholung des Vorgangs, bis die letzte Liste aus einem einzigen Eintrag besteht (mit $p(x_i)=1{,}0$)
- Bildung des Codebaums in umgekehrter Richtung (von den Blättern zu der Wurzel) anhand der zusammengefaßten Zeichen (bei frei wählbarer Verteilung v. 0 u. 1).

Codierung nach Huffman:

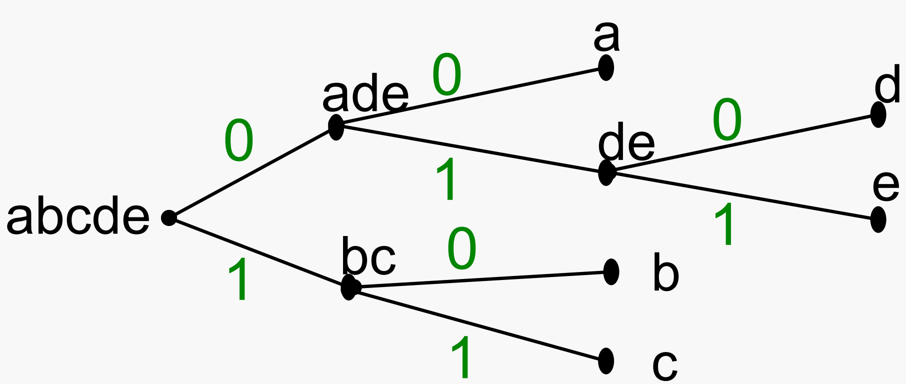
x_i	$p(x_i)$
a	0,30
b	0,24
c	0,20
d	0,15
e	0,11

x_i	$p(x_i)$
a	0,30
de	0,26
b	0,24
c	0,20

x_i	$p(x_i)$
bc	0,44
a	0,30
de	0,26

x_i	$p(x_i)$
ade	0,56
bc	0,44

x_i	$p(x_i)$
abcde	1,0



Codierung nach Huffman:

- Auflistung zulässiger Zeichen
- Zusammenfassung der zwei niedrigsten Wahrsch/keiten zu neuem Zeichen
- Neue Liste inkl. Zeichen mit aufsummierter Wahrsch/keit
- Wiederholung bis $p(x_i)$=1,0
- Bildung des Codebaums von den Blättern zu der Wurzel bei frei wählbaren 0 und 1.

Grundzüge der Informationstheorie (2)

Anmerkungen und Erkenntnisse (ohne Beweis):

- Die Verfahren von Fano und Huffman liefern Codierungen, bei denen kein kurzes Codewort dem Anfang eines längeren Codewortes gleich ist (Bedingung von Fano).
- Sind die auftretenden Wahrscheinlichkeiten von der Form 2^{-n} ($n \in \mathbb{N}$), so liefern die Codierungsverfahren nach Fano und Huffman redundanzfreie Codierungsergebnisse.
- „Codierungssatz“:

  Zur Codierung der Zeichen einer Informationsquelle der Entropie H bit/Zeichen läßt sich stets ein Code finden, dessen Codewörter im Mittel m Binärstellen pro Zeichen enthalten mit $H \leq m \leq H + \varepsilon$. Mit entsprechend hohem Codierungsaufwand läßt sich ε beliebig klein vorgeben.