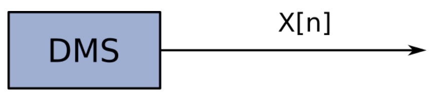
Dr. Jürg M. Stettbacher
Neugutstrasse 54
CH-8600 Dübendorf

Telefon: +41 43 299 57 23
Email: dsp@stettbacher.ch

Übung

Huffman Codierung



Eine diskrete gedächtnislose Quelle (DMS[1]) erzeugt zu jedem diskreten Zeitpunkt $n = 0,\ 1,\ 2,\ \ldots$ eine Zufallsvariable $X[n]$. Das heisst, es entsteht mit der Zeit eine Sequenz $X[.] = \{X[0]\ X[1]\ X[2]\ \ldots\}$ von derartigen Zufallsvariablen. Jede Zufallsvariable $X[n]$ kann einen der Werte A, B oder C annehmen. Dabei gilt:

$$P(A) = \frac{1}{10} \qquad P(B) = \frac{3}{10} \qquad P(C) = \frac{6}{10}$$

1. Wie gross ist die Entropie $H(X)$ der Quelle?
2. Entwerfen Sie einen Huffman Code für die Symbole A, B und C.
3. Wie gross ist die mittlere Codewortlänge L_1 und die Redundanz R_1 des Huffman Codes?
4. In der Sequenz $X[.]$ werden nun immer zwei aufeinander folgende Zufallsvariablen zusammen gefasst, so dass Doppelsymbole aus A, B und C entstehen. Wie gross ist die Entropie $H(XX)$ dieser Doppelsymbole?
5. Entwerfen Sie den Huffman Code für die Doppelsymbole.
6. Wie gross ist die mittlere Codewortlänge L_2 und die Redundaz R_2 des zweiten Huffman Codes bezogen auf ein ursprüngliches Symbol?
7. Wir betrachten nun einen anderen Code für die Symbole A, B und C:

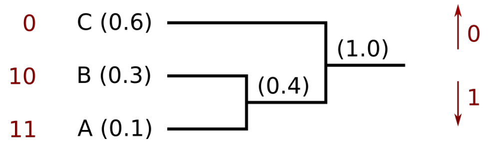
$$\begin{aligned} A &= 01 \\ B &= 1 \\ C &= 0 \end{aligned}$$

   Was ist das Problem dieses Codes?

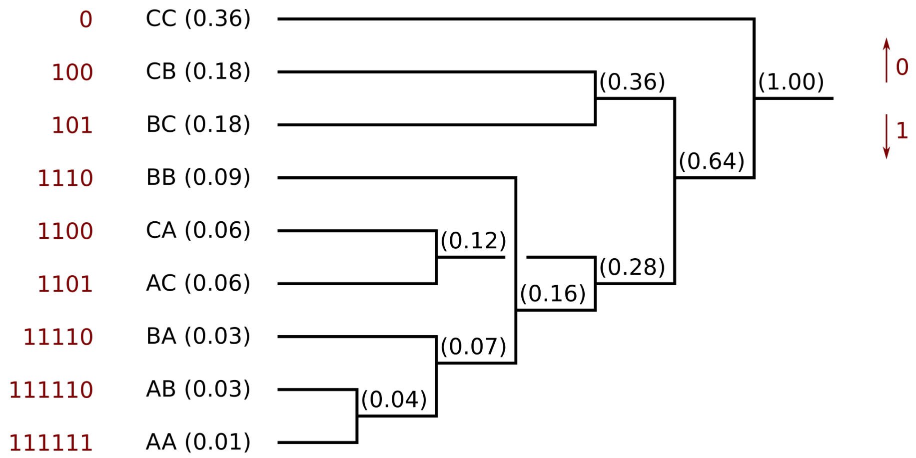
[1] Englisch: Discrete Memoryless Source.

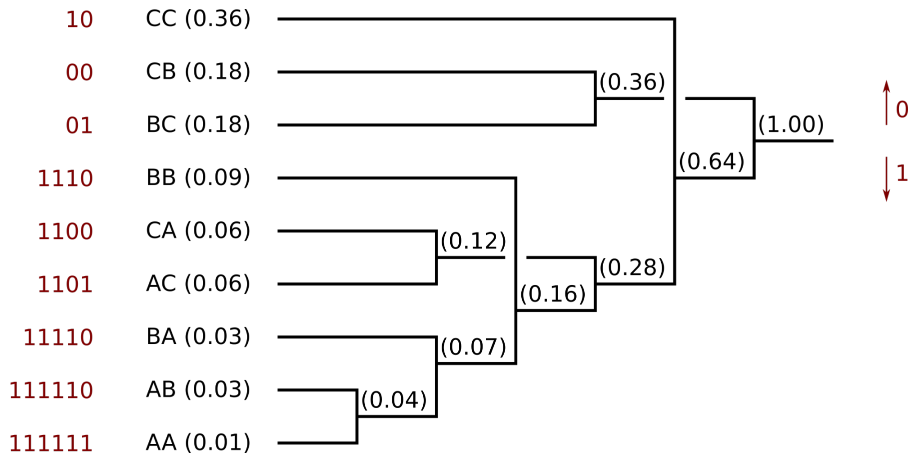
Antworten

1. Entropie: $H(X) = 1.30 \quad \text{Bit/Symbol}$.
2. Huffman Baum:



3. Mittlere Codewortlänge: $L_1 = 1.40 \quad \text{Bit/Symbol}$.
   Redundanz: $R_1 = 0.10 \quad \text{Bit/Symbol}$.
4. Entropie: $H(XX) = 2 \cdot H(X) = 2.60 \quad \text{Bit/Symbol}$.
5. Es gibt zwei Möglichkeiten für den Huffman Baum:





6. Für beide Huffman Bäume gilt:
   Mittlere Codewortlänge: $L_2 = 2.67/2 = 1.335 \quad \text{Bit/Symbol}$.
   Redundanz: $R_2 = 0.07/2 = 0.035 \quad \text{Bit/Symbol}$.
7. Der Code ist nicht präfixfrei. Werden also die einzelnen Codeworte in einen Bitstrom eingereiht, und betrachten wir darin den Ausschnitt (... 0 1 ...), so können wir nicht entscheiden, ob das für (... A ...) oder für (... C B ...) steht.