6. Übung aus Informations- und Codierungstheorie SS 2016

1. Sei $P = (p_1, \ldots, p_m)$ mit $p_1 > p_2 \geq \ldots \geq p_m$. Man zeige, dass aus $p_1 > \frac{2}{5}$ folgt, dass das kürzeste Wort im binären Huffman-Code die Länge 1 hat.
   Lösung: Für $m = 2\,, 3$ ist die Aussage trivial.
   $m = 4$: $\quad p_1 > \frac{2}{5} \quad \Rightarrow \quad p_2 + p_3 + p_4 < \frac{3}{5}$, und damit gilt für $p_2 \geq \frac{1}{5}$ $p_3 + p_4 \leq \frac{2}{5} < p_1$ Aber auch für $p_2 < \frac{1}{5}$ gilt wegen $p_2 \geq p_3 \geq p_4$ $p_3 + p_4 \leq \frac{2}{5} < p_1$. Somit werden nach $3\,, 4$ die Ausgänge $2\,, \{3\,, 4\}$ zusammengefasst und die 1-te Frage entscheidet zwischen 1 und diesem Ausgang.
   $m \geq 5$: Wir nehmen an, dass die Behauptung für $m-1$ richtig ist. Klarerweise gilt $\sum\limits_{i=2}^{m} p_i < \frac{3}{5}$. Daraus folgt
   $$2\,(p_{m-1} + p_m) \;\leq\; \lfloor \frac{m-1}{2} \rfloor\,(p_{m-1} + p_m) \leq \sum_{i=2}^{m} p_i < \frac{3}{5}$$
   $$\Rightarrow p_{m-1} + p_m < \frac{3}{10} < p_1\,.$$
   Damit ist p_1 auch in der reduzierten Verteilung die größte Wahrscheinlichkeit und die Behauptung folgt aus der Induktionsannahme.
2. Sei $P = (p_1, \ldots, p_m)$ mit $p_1 > p_2 \geq \ldots \geq p_m$. Man zeige, dass für $p_1 < \frac{1}{3}$ das kürzeste Wort im binären Huffman-Code aus mindestens 2 Zeichen bestehen muss.
   Lösung: Der Huffman-Algorithmus führt nach $m-3$ Schritten zu einer Verteilung $\widetilde{P} = (\widetilde{p}_1, \widetilde{p}_2, \widetilde{p}_3)$ mit $\widetilde{p}_1 \geq \widetilde{p}_2 \geq \widetilde{p}_3$. Gilt $p_1 \neq \widetilde{p}_i\,,\ i = 1, 2, 3\,,$ so wurde p_1 schon in einem der bisherigen Schritte zusammengefasst, sodass gilt $l_1 > 1$. Gilt $p_1 = \widetilde{p}_2$ oder $p_1 = \widetilde{p}_3$, so impliziert das $l_1 = 2$. Aber $p_1 = \widetilde{p}_1$ kann nicht sein wegen $p_1 < \frac{1}{3}$ und $\widetilde{p}_1 + \widetilde{p}_2 + \widetilde{p}_3 = 1$.
3. Man beweise, dass für jede Verteilung $P = (p_1, \ldots, p_m)$, zu der es einen Präfix-Code C mit den Wortlängen $l_1, \ldots, l_m$ auf dem Codealphabet $\{0, \ldots, r-1\}$ gibt, dessen mittlere Wortlänge $l(C)$ mit $H_r(P)$ übereinstimmt, gilt $m \equiv 1 \bmod (r-1)$ und $p_i = r^{-l_i} \quad \forall\, i = 1, \ldots, m$.
   Beweis: Da C ein Präfix-Code ist, gilt $\sum\limits_{i=1}^{m} r^{-l_i} \leq 1$. Mit $Q := (r^{-l_1}, \ldots, r^{-l_m})$ gilt $0 = D(P|Q) = \sum\limits_{i=1}^{m} p_i \,\log_r r^{l_i} - H(P) = \sum\limits_{i=1}^{m} p_i\, l_i - H(P) = l(C) - H(P)$ genau dann, wenn $P = Q$, d.h. $p_i = r^{-l_i} \quad \forall\ 1 \leq i \leq m$. Aber dann gilt auch $\sum\limits_{i=1}^{m} r^{-l_i} = \sum\limits_{i=1}^{m} p_i = 1$, d.h. C ist vollständig. Das impliziert $m \equiv 1 \bmod (r-1)$.

4. Man zeige, dass es zu jedem Präfixcode zum Codealphabet $\{0,\dots,r-1\}$, für dessen Wortlängen $l_1,\dots,l_m$ gilt $\sum\limits_{i=1}^{m} r^{-l_i} < 1$ Codefolgen gibt, die nicht entziffert werden können.
   Beweis: Gilt o.E.d.A. $l_1 \le \dots \le l_m$, so gibt es eine Folge aus $\{0,\dots,r-1\}^{l_m}$, die weder Codewort noch Präfix eines Codewortes ist (wegen $\sum\limits_{i=1}^{m} r^{-l_i} < 1$ ist der Code nicht vollständig). Dieser Folge kann keine Nachricht zugeordnet werden.
5. Ist $\mathcal{P} := \{P := (p_1,\dots,p_m) : \ 0 \le p_i \le 1, \ \sum\limits_{i=1}^{m} p_i = 1\}$ die Menge der Verteilungen auf $\Omega := \{\omega_1,\dots,\omega_m\}$, so zeige man, dass $l(P)$ die mittlere Wortlänge des optimalen Codes zu P eine stetige Funktion auf $\mathcal{P}$ ist.
   Beweis: Ist C ein beliebiger Code mit den Wortlängen $l_1(C),\dots,l_m(C)$, so ist $l(C) := \sum\limits_{i=1}^{m} l_i(C)\, p_i$ natürlich eine stetige Funktion von $p_1,\dots,p_m$. Aber das Huffman-Verfahren liefert einen optimalen Code C_H zum Codealphabet $\{0,\dots,r-1\}$ mit $\max\limits_{1\le i\le m} l_i \le \lceil \frac{m}{r-1} \rceil \le m-1$. Daher gilt auch $l(P) = \min\{l(C) : 1 \le l_i(C) \le m-1 \quad \forall\ 1 \le i \le m\}$, sodass $l(P)$ als Minimum von endlich vielen stetigen Funktionen stetig ist.