2. Übung Informationstheorie SS18

1. Gegeben Sei die Verteilung

$$P = (0.1, 0.4, 0.05, 0.2, 0.25)$$

   Bestimmen sie den Shannon-Code.

2. Gegeben Sei die Verteilung

$$P = (0.1, 0.4, 0.05, 0.2, 0.25)$$

   Bestimmen sie den Fano-Code.

3. Kanonische Codes: je nachdem, wie der Fragebaum gebaut wird, kommt man zu unterschiedlichen Codewörtern für den Huffman-Code. Um Eindeutigkeit zu erreichen, kann man wie folgt vorgehen (dieses Verfahren lässt sich auf jeden Präfixfreien Code anwenden):

   Zuerst werden mit dem Huffman-Algorithmus nur die Codewortlängen bestimmt. Diese werden austeigend geordnet:

$$l_1 \leq l_2 \leq \ldots \leq l_m.$$

   Die Codewörter (als Binärzahl mit l_i Stellen gelesen, unter Umständen mit führenden Nullen) erhält man durch

$$c_1 = 0,$$
$$c_{i+1} = (c_i + 1)2^{l_{i+1} - l_i}, i = 1, \ldots, m - 1.$$

   Zeigen Sie dass dadurch ein präfixfreier Code erzeugt wird (dieser Code wird der kanonische Code genannt. Ein Vorteil dieses Verfahrens ist, dass zur übertragung des Codes nur die Codewortlängen angegeben werden müssen).

4. Bestimmen Sie für die Verteilung aus Beispiel 1 den kanonischen Huffmancode.

5. X sein eine Zufallsvariable, die nur nichtnegative ganzzahlige Werte annimmt, $m = \mathbb{E}(X) < \infty$. Zeigen Sie

$$H(X) \leq \log_2(m + 1) + m \log_2\left(\frac{m + 1}{m}\right).$$

   (Hinweis: betrachten Sie die Diskrepanz zwischen der Verteilung von X und einer geometrischen Verteilung mit demselben Erwartungswert)

6. Im vorigen Beispiel haben wir mit der Entropie einer Zufallsvariable mit unendlich vielen möglichen Werten gearbeitet. Das ist natürlich möglich, die Definition ist die gleiche, nur ist die Summe jetzt unendlich. Es kann allerdings passieren, dass die Entropie einer solchen Zufallsvariable unendlich ist. Geben Sie ein Beispiel für eine Verteilung, bei der das der Fall ist.

7. Lauflängencodierung: ein Folge $X_1, \ldots, X_n$ von unabhängigen Zufallsvariablen mit $\mathbb{P}(X_i = 1) = p$, $\mathbb{P}(X_i = 0) = 1 - p$ wird codiert, indem zuerst (damit für das Ende keine eigene Regel eingeführt werden muss) $X_{n+1} = 1$ gesetzt wird. Dann werden die Lauflängen

   $$L_i = \tau_i - \tau_{i=1}, i = 1, \ldots, k$$

   mit

   $$k = \sum_{i=1}^{n+1} X_i$$

   und

   $$\tau_0 = 0, \tau_i = \inf\{j > \tau_{i-1} : X_j = i\}, i = 1, \ldots, k$$

   bestimmt. Diese Lauflängen werden mit dem bekannten Verfahren (L wird als Binärzahl mit s Stellen geschrieben, und davor kommen $s - 1$ Nullen). Bestimmen Sie für kleines p (näherungsweise: Sie dürfen so tun als ware der Erwartungswert des Logarithmus gleich dem Logarithmus des Erwartungswerts) die mittlere Anzahl von Codebuchstaben pro Quellzeichen und vergleichen Sie mit der Entropie.

   Was könnte man daran verbessern?