Exercises for **Applied Information Theory**

Prof. Dr.-Ing. Martin Bossert
M.Sc. Cornelia Ott / M.Sc. Jiongyue Xing

Exercise sheet 3

Task 3.1 (Typical Sequences)

The typical set A_ϵ^n is defined as the set of sequences $X_1, X_2, \ldots, X_n$ with the property (see Cover & Thomas, chapter 3)

$$2^{-n(H(X)+\epsilon)} \leq f_{X_1 \ldots X_n}(x_1, \ldots, x_n) \leq 2^{-n(H(X)-\epsilon)}$$

To clarify the notion of a typical set A_ϵ^n, we will calculate the set for a simple example. Consider a sequence of independent binary random variables, $X_1, X_2, \ldots, X_n$, where the probability that $X_i = 1$ is 0.6 (and therefore the probability that $X_i = 0$ is 0.4).

a) Calculate $H(X)$.

b) With $n = 25$ and $\epsilon = 0.1$, which sequences fall in the typical set A_ϵ^n ? What is the probability of the typical set?

c) How many elements are there in the typical set? (This involves computation of a table of probabilities for sequences with k 1's, $0 \leq k \leq 25$, and finding those sequences that are in the typical set.)

Task 3.2 (Uniquely decodable and prefix-free codes)

For source coding, only *uniquely decodable codes* are used in practice. They have the property, that each finite sequence with the corresponding symbol alphabet can be uniquely mapped to a sequence of codewords. Although there might be different interpretations of a sequence, those can never result in a sequence of valid codewords.

a) Explain why each *prefix-free code* is also a uniquely decodable code.

b) Determine for each of the following codes, whether it is uniquely decodable and whether it is prefix-free. If the code is not uniquely decodable, give an example of a sequence, for which more than one valid interpretation as a sequence of codewords exists.

   i) $\{00, 10, 01, 11\}$
   ii) $\{0, 10, 11\}$
   iii) $\{0, 01, 11\}$
   iv) $\{1, 101\}$
   v) $\{0, 1, 01\}$

Task 3.3 (Existence of prefix-free codes)

Exists a binary prefix-free code for following codeword-lengths? If yes, then construct a corresponding code.

a) $(1, 2, 3, 4, 5, 5)$

b) $(1, 2, 3, 4, 4, 5)$

c) $(2, 2, 2, 3, 4)$

Exists a ternary prefix-free code for following codeword-lengths? If yes, then construct a corresponding code.

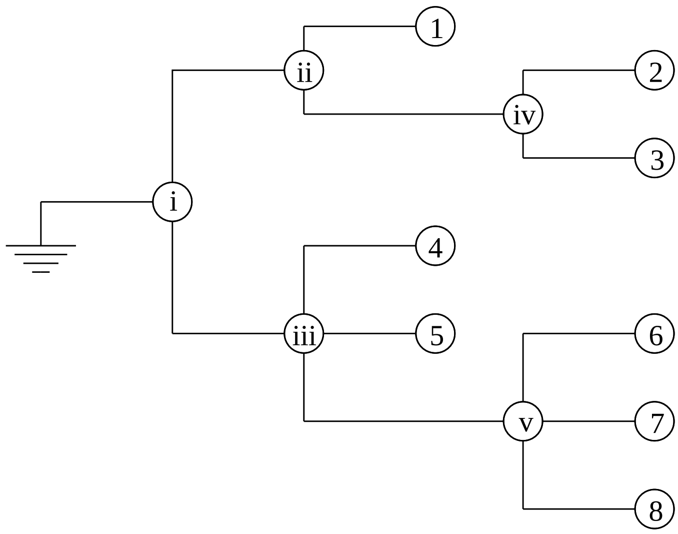
d) $(1, 2, 3, 1, 2, 3, 2)$

e) $(1, 2, 3, 4, 1, 2, 3, 4)$

f) $(3, 3, 3, 3, 3, 3, 2, 1, 1)$

Task 3.4 (Probability tree)

Following *probability tree is given*:



The nodes are enumerated with Roman numbers, the leafs with Arabic numbers. For the probability of the leafs, it holds: $p_1 = 1/4$, $p_2 = p_3 = 1/8$, $p_4 = p_5 = 1/6$ and $p_6 = p_7 = p_8 = 1/18$.

a) Give the average path-length $\mathrm{E}[W]$ from the root to the leafs and the entropy defined in the lecture $H_{\text{tree}} = -\sum_{j=1}^{8} p_j \log p_j$.

b) Calculate the probabilities P_i, P_{ii}, P_{iii}, P_{iv} and P_v of the nodes in the tree?

c) Give $\mathrm{E}[W]$ with the help of the *path length lemma*: $\mathrm{E}[W] = \sum_{j=i}^{v} P_j$.

d) Give the *branch–entropy* H_k for $k = i, ii, iii, iv, v$. Calculate the entropy H_{tree} on a different way than in a).

Task 3.5 (Optimal prefix-free codes)

Show that fundamentally different prefix-free codes can exist for a given random variable. Fundamentally different codes differ in the list of codeword lengths. How many fundamentally different optimal binary codes exist for the random variable, which yields different symbols with probabilities 0.3, 0.2, 0.2, 0.1, 0.1, 0.05 and 0.05.