Solution to Task 3.1 (Typical Sequences)

a) $H(X) = h(0.4) = -0.4 \log_2 0.4 - 0.6 \log_2 0.6 = 0.9710$

b)
- A sequence $(X_1, \dots, X_n) =: \mathbf{X}^n$ with the probability $f_{X_1 \dots X_n}(x_1, \dots, x_n) =: P(\mathbf{X}^n)$ is in the typical set A_ϵ^n if
$$2^{-n(H(X)+\epsilon)} \leq P(\mathbf{X}^n) \leq 2^{-n(H(X)-\epsilon)}.$$
- With $n = 25, \epsilon = 0.1$ we have $2^{-n(H(X)+\epsilon)} = 8.7155 \cdot 10^{-9}$ and $2^{-n(H(X)-\epsilon)} = 2.7890 \cdot 10^{-7}$.
- $P(\mathbf{X}^n)$ depends only on the number of ones contained in the sequence $\mathbf{X}^n$.
- The probability of one specific sequence that contains i ones and $n - i$ zeros is
$$q_i = 0.6^i \, 0.4^{n-i},$$
where $\binom{n}{i}$ such sequences exist and q_i is monotonically increasing in i ($0 \leq i \leq n$).
- By calculating q_i for all values of i we find
$$q_{10} = 6.4925 \cdot 10^{-9} \quad < 2^{-n(H(X)+\epsilon)} \quad < q_{11} = 9.7388 \cdot 10^{-9} \tag{1}$$
$$q_{19} = 2.4959 \cdot 10^{-7} \quad < 2^{-n(H(X)-\epsilon)} \quad < q_{20} = 3.7439 \cdot 10^{-7} \tag{2}$$
Thus, all sequences containing at least 11 and at most 19 ones are in the set $A_{0.1}^{25}$
$$P(A_{0.1}^{25}) = \sum_{i=11}^{19} \binom{n}{i} 0.6^i \, 0.4^{n-i} = 0.9362$$

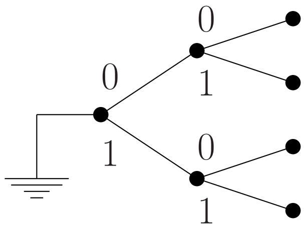
c) $|A_{0.1}^{25}| = \sum_{i=11}^{19} \binom{n}{i} = 26\,366\,510$

Solution to Task 3.2 (Uniquely decodable and prefix-free codes)

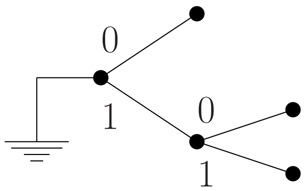
a) A prefix-free code is always uniquely decodable, because for every finite sequence one can start decoding at the beginning of the sequence and read as many symbols as necessary to obtain a valid codeword. This step is repeated until the end of the sequence. As no codeword is prefix of another codeword, this always results in one unique sequence of codewords.
Therefore, prefix-free codes are also called instantaneously decodable codes.

b) Note: A code can always be represented by a code tree. The code is prefix-free if only the leafs of this tree are codewords, where the tree does not have to be symmetric.

i) $\{00, 01, 10, 11\}$ is a block-code and therefore prefix-free. As all codewords have the same length, a codeword can only be prefix of another one, if they are identical. However, this is not the case here.



ii) $\{0, 10, 11\}$ is prefix-free, because all codewords are leafs of the code tree.



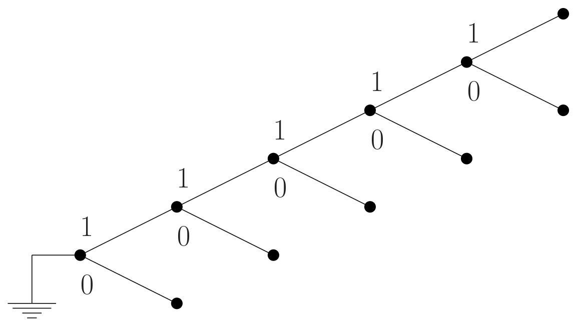
iii) $\{0, 01, 11\}$ is not prefix-free, since 0 is a prefix of 01. However, this code is uniquely decodable as it is suffix-free. A suffix-free code can always be decoded from the end of a sequence. However, it is not instantaneously decodable as one has to wait until the end of a sequence.

iv) $\{1, 101\}$ is not prefix-free, since 1 is a prefix of 101. However, this code is uniquely decodable as all 0s in the sequence can be uniquely determined.

v) $\{0, 1, 01\}$ is neither prefix-free nor uniquely decodable. For instance 01 and $0|1$ are two valid interpretations of the same sequence.

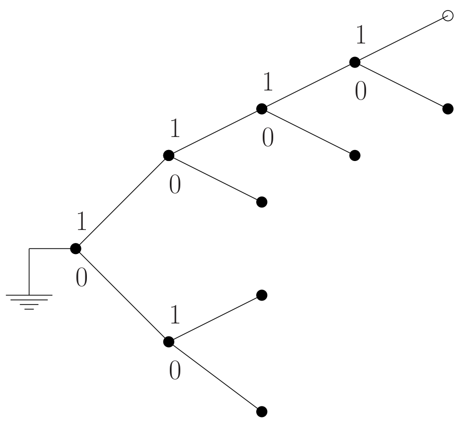
Solution to Task 3.3 (Existence of prefix-free codes)

a) $2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-5} + 2^{-5} = 1 \quad \Rightarrow \quad$ Code exists.



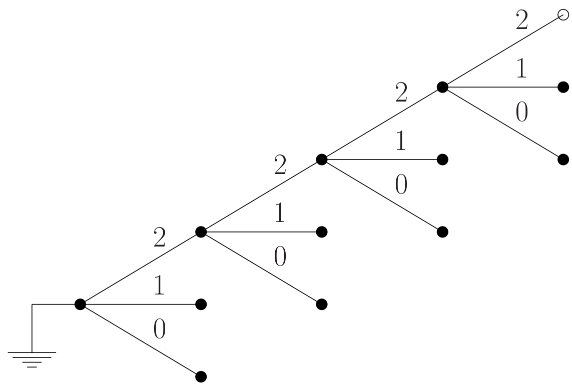
b) $2^{-1} + 2^{-2} + 2^{-3} + 2^{-4} + 2^{-4} + 2^{-5} = \frac{33}{32} > 1 \quad \Rightarrow$ Code does not exist.

c) $2^{-2} + 2^{-2} + 2^{-2} + 2^{-3} + 2^{-4} = \frac{15}{16} < 1 \quad \Rightarrow$ Code exists.

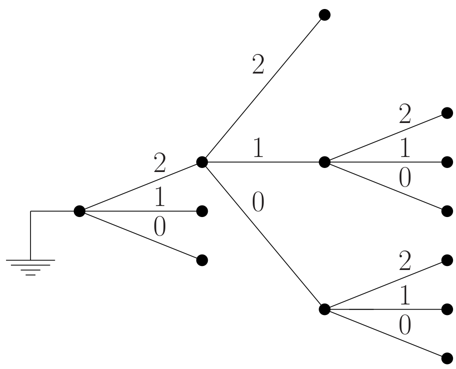


d) $3^{-1} + 3^{-1} + 3 \cdot 3^{-2} + 2 \cdot 3^{-3} = \frac{29}{27} > 1 \quad \Rightarrow$ Code does not exist.

e) $2 \cdot 3^{-1} + 2 \cdot 3^{-2} + 2 \cdot 3^{-3} + 2 \cdot 3^{-4} = \frac{80}{81} < 1 \quad \Rightarrow$ Code exists.



f) $2 \cdot 3^{-1} + 3^{-2} + 6 \cdot 3^{-3} = 1 \quad \Rightarrow$ Code exists.



Solution to Task 3.4 (Probability tree)

a) Average path-length:

$$\begin{aligned} \mathrm{E}[W] &= 1/4 \cdot 2 + 1/8 \cdot 3 + 1/8 \cdot 3 + 1/6 \cdot 2 + 1/6 \cdot 2 + 1/18 \cdot 3 + 1/18 \cdot 3 + 1/18 \cdot 3 \\ &= \frac{29}{12} \approx 2,42 \end{aligned}$$

Entropy H_{tree}:

$$\begin{aligned} H_{\text{tree}} &= -1/4 \log_2 1/4 - 2 \cdot 1/8 \log_2 1/8 - 2 \cdot 1/6 \log_2 1/6 - 3 \cdot 1/18 \log_2 1/18 \\ &= \frac{7}{4} + \frac{2}{3} \log_2 3 \approx 2,81 \end{aligned}$$

Observe, that the lower bound for $\mathrm{E}[W]$ is given by

$$\frac{H_{\text{tree}}}{\log_2 3} \approx 1,77,$$

because it is a ternary code.

b) Node probability:

$$P_i = 1, \quad P_{ii} = 1/2, \quad P_{iii} = 1/2, \quad P_{iv} = 1/4, \quad P_v = 1/6$$

c) Alternative method for average path length:

$$\mathrm{E}[W] = \sum_j P_j = 1 + 1/2 + 1/2 + 1/4 + 1/6 = \frac{29}{12} \approx 2,42$$

d) Branch-entropy $H_k = -\sum_j \frac{q_{kj}}{P_k} \log \frac{q_{kj}}{P_k}$, where q_{kj} are the probabilities of the corresponding outgoing branches.

$$\begin{aligned} H_i &= -1/2 \log_2 1/2 - 1/2 \log_2 1/2 = 1 \\ H_{ii} &= -\frac{1/4}{1/2} \log_2 \frac{1/4}{1/2} - \frac{1/4}{1/2} \log_2 \frac{1/4}{1/2} = 1 \\ H_{iii} &= -3 \cdot \left(\frac{1/6}{1/2} \log_2 \frac{1/6}{1/2} \right) = \log_2 3 \\ H_{iv} &= -2 \cdot \left(\frac{1/8}{1/4} \log_2 \frac{1/8}{1/4} \right) = 1 \\ H_v &= -3 \cdot \left(\frac{1/18}{1/6} \log_2 \frac{1/18}{1/6} \right) = \log_2 3 \end{aligned}$$

Alternative calculation for the entropy H_{tree}

$$H_{\text{tree}} = \sum_j P_j H_j = 1 \cdot 1 + 1/2 \cdot 1 + 1/2 \log_2 3 + 1/4 \cdot 1 + 1/6 \log_2 3 \approx 2,81$$

Solution to Task 3.5 (Optimal prefix-free codes)

Huffman-Algorithm for the construction of optimal prefix-free codes: