NTIN 100 Intro to Info Transmission and Processing summer 2019/2020

2nd homework assignment - Coding and compression

turn in by April 7, 2020.

Problem 1.   For each of the following codes decide whether it is uniquely decodable. For each uniquely decodable code find an infinite string which can be decoded in two different ways if there is such a string. Show that prefix-free codes do not have such a sequence.

a) $C_1 = \{0, 10, 11\}$, $C_2 = \{0, 01, 11\}$

b) $C_3 = \{0, 01, 10\}$, $C_4 = \{1, 10\}$

c) $C_5 = \{110, 11, 10\}$, $C_6 = \{110, 11, 100, 00, 10\}$

Problem 2.   *20 questions.* Consider n objects, each of them being independently *good* or *bad*. The probability of the i-th item being good is p_i, where $p_1 > p_2 > \cdots > p_n > 1/2$. You should determine which items are good and which items are bad. You can ask arbitrary YES/NO questions.

a) Give a good lower bound on the minimum average number of questions required.

c) Give an upper bound (within ± 1) on the minimum average number of questions required.

Problem 3.   Consider prefix-free code over binary alphabet with codeword lengths $\ell_1, \ell_2, \ldots, \ell_k$ such that

$$\sum_{i=1}^{k} 2^{-\ell_i} < 1.$$

Show that there are arbitrary long strings in $\{0,1\}^*$ that cannot be split into a sequence of codewords.

Problem 4.   Give an example of a random variable X and an arbitrary code $C : X \to \{0,1\}^*$ such that $E[\|C(X)\|] < H(X)$. For any random variable X and a code C, show a lower bound on $E[\|C(X)\|]$ in terms of $H(X)$ which is as good as you can come-up with.