Wolfgang K. Seiler
Tel. 2515

B6, B4.08

September 28, 2023

Mathematics and Information, Exercise sheet 3

Problem 1: (*4 points*)
Let $\mathcal{X} = (X_t)_{t\in\mathbb{N}}$ be a sequence of independent random variables with values in $A = \{a, b\}$, where each X_i takes the value a with a probability of $2/3$. Determine for $\varepsilon = \frac{1}{2}$ and $\varepsilon = \frac{1}{4}$ which percentage of the elements of A^{10} lies in A^{10}_ε !

Problem 2: (*4 points*)
A given source produces sequences of zeroes and ones with $p(1) = 1/200$.

a) Compute the entropy of this source!

b) The sequences produced by this source are divided into blocks of length one hundred. For each sequence containing at most three ones, a code word is chosen, and there is also one singe word for all sequences containing more than three ones. All code words have the same length. Determine the smallest possible length of these code words and the probability of the word for sequences with more than three ones!

c) How many blocks are encoded by this word?

Problem 3: (*4 points*)
Let X be a random variable with values in $A = \{a, b, c, d, e\}$ with probabilities $p(a) = \frac{1}{11}$, $p(b) = p(c) = \frac{2}{11}$, and $p(d) = p(e) = \frac{3}{11}$.

a) Compute the entropy of X !

b) Now encode A over $\mathcal{D} = \{0, 1\}$. Which is the smallest average code length permitted by reasons of number theory?

c) Find a distribution of code lengths permitted by the inequality of Kraft and McMillan leading to this value!

d) Find a corresponding encoding!

Problem 4: (*4 points*)
The alphabet $\{a, b, c, d, e, f, g, h\}$ is to be binary coded.

a) Show that the assignment $\ell(a) = 3$, $\ell(b) = \ell(c) = 4$, $\ell(d) = 3$,$\ell(e) = 2$, $\ell(f) = 5$, $\ell(g) = 4$ and $\ell(h) = 5$ satisfies the inequality of Kraft and McMillan!

b) Find an binary prefix code with these lengths!

c) Which code words can be shortened in such a way that the code remains a prefix code?

d) Show that no code word in the resulting code can be replaced by a shorter one without loosing unique decodability!

Problem 5: (*4 points*)

a) Let m be a positive integer which is not a power of two, and consider a source emitting the letters form an alphabet A of cardinality m with equal probability. Sequences of r letters from this alphabet should be coded by sequences of bits, where the lengths of the longest and the shortest code word differ by at most one, in such a way that the average code length is as small as possible. Use the inequality of Kraft and McMillan and its converse to determine, how many code words of which lengths should be used in such a code!

b) Compute these numbers for the case $m = 3$ and $1 \le r \le 5$!

c) Show, in the general case, that the average code length converges to the entropy of the source for $r \to \infty$!

Can be submittel till Wednesday, October 4, 2023 at 15.30h