Introduction to Information Theory, Fall 2020

Homework problem set #3 **due November 16, 2020**

Rules: Always explain your solutions carefully. Please hand in the assignment in groups on Canvas. In the werkcollege the TAs can tell you more about how this works.

1. **Entropy and typical sets; LZ algorithm (1 point):** Let P be the probability distribution with three possible outcomes A, B, C and probabilities $P(A) = 1/2$, $P(B) = 1/4$, $P(C) = 1/4$. Let X_1, X_2 be independent and identically distributed (IID) according to P.

   (a) Compute $H(X_1)$, $H(X_2)$, and $H(X_1, X_2)$.
   (b) Make a table that lists the joint probability $P(x_1, x_2)$ and the quantity $\frac{1}{2} \log \frac{1}{P(x_1,x_2)}$ for all possible outcomes x_1 and x_2.
   (c) Compute $H_\delta(X_1, X_2)$ for $\delta = 3/8$.
   (d) Write down all elements of the typical set $T_{2,\varepsilon}(P)$ for $\varepsilon = 0.12345$.

   Now for something else. Imagine running the Lempel-Ziv algorithm on the following string:

   $$\texttt{X00X0X00X0XXXX0000X}\perp$$

   (e) List the distinct phrases that the LZ algorithm splits the string into.
   (f) List the pairs (k, x) generated by the LZ algorithm.
   Notation: Write k as an integer in decimal notation and $x \in \{X, 0, \perp\}$.

2. **Kraft-McMillan inequality (1 point):** Let X be a random variable with distribution P. In class, we discussed that there always exists a prefix code C whose codewords have length $\ell(C(x)) = \lceil \log 1/P(x) \rceil$. Let C_2 be any other uniquely decodable code. Show that, for all k,

   $$\Pr\Big(\ell\big(C_2(X)\big) \leqslant \ell\big(C(X)\big) - k\Big) \leqslant \frac{1}{2^{k-1}}.$$

   This implies no other code can produce much shorter codewords than C most of the time.

   Hint: Write the probability as a sum over all possible x and use the Kraft-McMillan inequality.

3. **Compression algorithms (1 point):**
   In this problem, you will implement two more algorithms: the 'universal' typical-set compression algorithm and the Lempel-Ziv algorithm that we will discuss on Tuesday and Thursday. To get started, open the notebook at https://colab.research.google.com/github/amsqi/iit20-homework/blob/master/03-homework.ipynb and follow the instructions.

   Please submit both the notebook **and** a PDF printout, or provide a link to your solution on Colab. You can achieve the maximum score if your solution produces the correct output. We will only have a closer look at your code in case of problems.

   This programming problem may again be a bit more difficult than last week's; we will grade it gently.