Statistical Information Theory

Problem sheet 2

Problems for Lecture 3: Symbol codes

1. **Prefix codes**. Consider two codes, each with 8 codewords, with the following lengths:

$$\begin{aligned} C_1 &: \quad \{1, 2, 3, 3, 4, 4, 4, 4\} \\ C_2 &: \quad \{2, 3, 3, 3, 3, 4, 4, 4\} \\ C_3 &: \quad \{1, 3, 3, 4, 4, 4, 5, 5\} \end{aligned}$$

   (a) Do binary prefix codes with those lengths exist? If so, for each case, give an example code with those lengths.

   (b) Could more codewords be added to any of the codes above? If so, for each case, give an example.

2. **Shannon codes**. Constructing a binary code for a given PMF $p(x)$ can be challenging because $\log p(x)$ may not be an integer. *Shannon coding* is a compression technique that assigns the following codeword lengths:

$$\ell(x) = \lceil -\log p(x) \rceil$$

   (a) Prove that Shannon codes satisfy the Kraft inequality.

   (b) Prove that $H(X) \leq L < H(X) + 1$, where L is the expected length of the code.

   (c) In what cases would you predict Shannon codes to be far from optimal? Give one example PMF for which Shannon codes are optimal, and one for which they are far from optimal.

Programming exercises

See the attached IPython notebook.