HW 2: Shannon entropy and data compression

(due October 2nd, before tutorial)

1. Show that $H(X|Y) = 0$ implies that X is a (deterministic) function of Y.

2. We showed in class that the optimal $H(X) - \log_2(1 + \lfloor \log_2 |\mathcal{X}| \rfloor) \leq \mathbf{E}\{|C^*(X)|\} \leq H(X)$. Show that there is a distribution P_X such that the lower bound holds with equality. (We want a nontrivial example, i.e., $|\mathcal{X}| > 1$.)

3. Huffman's algorithm constructs a prefix code C_{H} given a distribution $(p_1, \ldots, p_m)$ on the symbols $\{1, \ldots, m\}$. The objective of this problem is to show that the expected length $L(C_{\mathrm{H}})$ is minimum among all the prefix codes. Huffman's algorithm constructs a binary tree as follows. The algorithm starts with independent nodes labeled by the elements $1, \ldots, m$ and the corresponding probability. At the beginning, all the nodes or marked unvisited. At each step, we choose the two unvisited nodes u, v with minimum value of p_u, p_v. We create a new node w with an assigned probability $p_w = p_u + p_v$ which is the parent of u and v. w is marked as unvisited and u, v are marked as visited. The step is repeated $m - 1$ times until we have one unvisited node (the root) with an assigned probability 1. To every path from the root to a leaf of the tree, we assign a bitstring where a "left" edge is read as 0 and a "right" edge is read as 1. The obtained tree defines a code in the following way: for any $x \in \{1, \ldots, m\}$, $C_{\mathrm{H}}(x)$ is the bitstring corresponding to the path from the root to x.

   (a) Show that for any optimal code, it can be transformed to one with the following property: the two longest codewords correspond to the two least likely symbols, and they have the same length and they only differ in the last bit.

   (b) Conclude that C_{H} achieves the optimal expected length for $(p_1, \ldots, p_m)$.

4. Find a distribution (p_1, p_2, p_3, p_4) on elements $\{1, 2, 3, 4\}$ such that there are two codes with different encoding lengths $\{\ell_i\}_{1 \leq i \leq 4}$ and $\{\ell'_i\}_{1 \leq i \leq 4}$ while both codes minimize the average length $\sum_i p_i \ell_i$.