TUTORIAL IV

0 Homework 2

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

2. 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 $(p_1, \ldots, p_m)$.

3. 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 $\{\ell_i\}_{1 \le i \le 4}$ and $\{\ell'_i\}_{1 \le i \le 4}$ while both codes minimize the average length $\sum_i p_i \ell_i$.

1 Fixed-length almost lossless compressor: source coding theorem

Recall that a *fixed length compressor* for source $Y \in \mathcal{Y}$ of length ℓ is a function $C : \mathcal{Y} \to \{0,1\}^\ell$. It has error probability at most δ if there exists a decompressor $D : \{0,1\}^\ell \to \mathcal{Y}$ such that $\mathbb{P}[D(C(Y)) = Y] \ge 1 - \delta$. Let define

$$\ell^{opt}(Y, \delta) = \min\{\ell : \text{ there exists a length } \ell \text{ compressor for } Y \text{ with error probability } \delta\}.$$

In class, we stated *Shannon Source Coding Theorem*:
Let $X^n = X_1 \ldots X_n$ be a sequence of independent and distributed as $X \in \mathcal{X}$. For any $\delta \in (0, 1)$,

$$\lim_{n \to \infty} \frac{\ell^{opt}(X^n, \delta)}{n} = H(X). \tag{1}$$

We showed the upper bound in class. In this tutorial, we will first show the lower bound (also called converse) and then give another proof for the upper bound.

1.1 Converse of Shannon source coding theorem

In this section, we will show that for any $\varepsilon > 0$ and for large enough n, we have $H(X) - 2\varepsilon \leq \ell^{opt}(X^n, \delta)/n$.

1. Prove that we are done if we can show that: for every set $S \subseteq \mathcal{X}^n$ such that $\mathbb{P}[X^n \in S] \geq 1 - \delta$ we have $|S| \geq 2^{n(H(X)-2\varepsilon)}$.

2. Now suppose that there is $S \subseteq \mathcal{X}^n$ such that $|S| \leq 2^{\ell n}$ for some ℓ and $\mathbb{P}[X^n \in S] \geq 1 - \delta$. Prove that

$$\mathbb{P}[X^n \in S] \leq \mathbb{P}_{X^n}\left[-\sum_{i=1}^{n} \log \mathbb{P}[X_i] \leq \ell n + \varepsilon n \right] + 2^{-\varepsilon n}.$$

3. Deduce that under the assumption of Question 2, $\ell \geq H(X) - 2\varepsilon$, and so the lower bound holds.

1.2 Achievability using random coding

Recall that in order to prove achievability of the source coding theorem, we chose the set S of correctly encoded symbols to be the set of $x^n \in \mathcal{X}^n$ such that $P_{X^n}(x^n) \geq 2^{-n(H(X)-\varepsilon)}$. We will now show a similar result by choosing the set S at random. In fact, we start by considering a general source (i.e., not necessarily iid) and derive an upper bound on the probability of error and will give us the desired result in the special case of an iid source.

Our objective is to show that for any source X and any integer $l \geq 0$, there exists a compressor with error probability

$$\delta \leq \mathbb{P}[-\log_2(P_X(X)) > l - \tau] + 2^{-\tau}, \ \ \forall \tau > 0. \tag{2}$$

1. Let $\tau > 0$, X be a random variable and C be a length l compressor. Let x_0 be a fixed letter of $\mathcal{X}$.
   Define $D = \{0,1\}^l \to \mathcal{X}$ by

$$D(y) = \begin{cases} x, & \text{if } \exists! x \in \mathcal{X} \text{ s.t. } C(x) = y \text{ and } -\log_2(P_X(x)) \leq l - \tau \\ x_0, & \text{otherwise} \end{cases} \tag{3}$$

   Define also $J(x, C) = \{x' \in \mathcal{X} : C(x) = C(x'), x \neq x', \text{ and } -\log_2(P_X(x')) \leq l - \tau\}$
   Show that

$$\mathbb{P}[D(C(X)) \neq X] \leq \mathbb{P}[-\log_2(P_X(X)) > l - \tau] + \mathbb{P}[J(X, C) \neq \emptyset]$$

2. Let C be a random length-l compressor, that is for each $x \in \mathcal{X}$, $C(x)$ is a random bit string of length l, with each bit choosen independantly and uniformly from $\{0,1\}$. Show that

$$\mathbb{E}_C[\mathbb{P}[J(X, C) \neq \emptyset]] \leq 2^{-\tau}$$

   where we compute the mean on the randomness of C but not on X.
3. Prove Eq. (2)

4. Can you build from the proof a length l compressor with error $\delta \leq \mathbb{P}[-\log_2(P_X(X)) > l - \tau] + 2^{-\tau}$?

5. Use Eq. (2) to give a proof of the upper bound in Shannon source coding theorem (1).

2 Typical sets

Let $X^n = X_1 \dots X_n$ be independent and identically distributed bits with $X_1 \sim \text{Ber}(p)$, i.e., $P_{X_1}(0) = 1 - p$ and $P_{X_1}(1) = p$ (assume that $0 < p < 1/2$) and let $\delta > 0$ with $p + \delta \leq 1/2$.

1. Recall that $h_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$. Show that for $k \leq n/2$ the following inequality holds:
$$1 + \binom{n}{1} + \cdots + \binom{n}{k} \leq 2^{h_2(k/n)n} .$$

2. Using the previous inequality show that there exists a set $S_\delta \subseteq \{0,1\}^n$ with $|S_\delta| \leq 2^{n \cdot h_2(p+\delta)}$ satisfying the property that
$$\lim_{n \to \infty} \mathbf{P}\{X^n \in S_\delta\} = 1 .$$

3 Rényi entropy

The Rényi entropy of order α, where $0 \leq \alpha < 1$, is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \Big(\sum_{i=1}^{n} p_i^\alpha \Big).$$

where X is a discrete random variable taking value in $\{1, 2, \dots, n\}$ each with probability $p_i = \Pr[X = i]$ for $i = 1 \dots n$. To define $\alpha = 0$ setting, we say that $0^0 = 0$.

1. Show that Rényi entropy is non-increasing function of α.

2. What is the value of H_0 and H_1 (here H_1 is defined as Rényi entropy when $\alpha \to 1$).

3. Show that H_α is concave function of the distribution $(p_1, \dots, p_n)$.