0 Homework 3

1. (Typical sets) Let $X^n = X_1 \ldots X_n$ 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 $0 < p < 1/2$). Let $\delta > 0$ with $p + \delta \leq 1/2$, show that there exists a set $S_\delta \subseteq \{0, 1\}^n$ with $|S_\delta| \leq 2^{n \cdot h_2(p + \delta)}$ where $h_2(p) = -p \log_2 p - (1 - p) \log_2 (1 - p)$ satisfying the property that

$$
\lim_{n \to \infty} P\{X^n \in S_\delta\} = 1.
$$

You may assume the following inequality for $k \leq n/2$ without proof:

$$
1 + \binom{n}{1} + \ldots + \binom{n}{k} \leq 2^{h_2(k/n)n}.
$$

Remark: We actually proved a more general version of this in class. You are asked here to produce an elementary self-contained proof.

2. (Bonus) Prove inequality (1).

3. Consider a source given by $X^n = X_1 \ldots X_n$ with $X_i$ independent and identically distributed bits with $P\{X_i = 1\} = 1/4$. Describe the distribution of the random variable $h_{X^n}(X^n) = -\log_2 P_{X^n}(X^n)$. How many values does it take? What is the probability for each different value? What is the expectation?

4. We showed in class that taking a sequence $X^n = X_1 \ldots X_n$ of independent copies of $X$, we have $h_{X^n}(X^n)$ converges weakly to $H(X)$. Now for random variables $X, Y$ we define $i_{X,Y}(X : Y) = \log_2 \frac{P_{XY}(X,Y)}{P_X(X)P_Y(Y)}$. If $X^n$ is $n$ independent copies of $X$ and $Y^n$ is $n$ independent copies of $Y$. What can you say on the random variable $\frac{i_{X^n,Y^n}(X^n : Y^n)}{n}$ as $n \to \infty$?

1 Midterm 2016-2017, Problem 4

[See Tutorial 4]

2 From fair coins to any discrete distributions

Given a random variable $X$ following a specific discrete distribution $p$, we want to know how many fair coins does it take to generate $X$. We want to minimize the average number of tosses we have to make. More formally: we are given a sequence of fair tosses $Z_1, Z_2, \ldots$, and wish to generate a discrete random variable $X \in \mathcal{X} = \{1, \ldots, m\}$, with a distribution $p = (p_1, \ldots, p_m)$. Let $T$ be the random variable denoting the number of coins flips used in the algorithm.

We can describe the algorithm using a tree: the leaves are marked by output symbols $X$, and the path to the leaves is given by the sequence of bits produced by the fair coin. We moreover assume that the tree satisfies some properties:
The tree should be complete (i.e. every node is either a leaf or has two descendants)

The probability of a leaf at depth \( k \) is \( 2^{-k} \). Many leaves may be labeled with the same output symbol – the total probability of all these leaves should be the one corresponding to this output symbol in the distribution \( p \).

In this representation, the average number of tosses is the expected depth of the tree. We want to find a tree with such properties that minimize its expected depth.

1. Consider the following distribution for \( X \):

\[
X = \begin{cases} 
  a & \text{with probability } \frac{1}{3} \\
  b & \text{with probability } \frac{2}{3} \\
  c & \text{with probability } \frac{1}{3}
\end{cases}
\]

Find the minimal average number of fair bits (tosses) needed to generate \( X \). Compare this value with \( H(X) \).

2. Given a complete tree, we denote by \( \mathcal{Y} \) the set of the leaves. Consider a distribution \( Y \) on the leaves such that the probability of a leaf at depth \( k \) is \( 2^{-k} \). Show that the expected depth of the tree is equal to the entropy of such a distribution.

3. Show that for any algorithm generating \( X \), the expected number of fair bits used is greater than the entropy, i.e. that: \( ET \geq H(X) \).

4. Show that if all the \( p_i \)'s are dyadic (i.e. \( p_i = 2^{-l_i} \)), one can achieve \( ET = H(X) \) with a finite algorithm.

5. Now we want to extend this result for non-diadic distributions. We will assume that this result holds even in the infinite case: i.e. for a dyadic distribution over an infinite set \( \mathcal{Y} \), we still can find an (infinite) algorithm \( T \) that achieves \( ET = H(Y) \).

(a) Let’s begin with an example: give an infinite tree that generate a random variable \( X \) with a distribution \( \left( \frac{1}{3}, \frac{2}{3} \right) \). What is its expected height? Compare this value with \( H(X) \).

(b) Given a non-dyadic distribution \( p = (p_1, \ldots, p_m) \), we split it into dyadic atoms, for example \( p_1 \to (p_1^{(1)}, p_1^{(2)}, \ldots) \), and so on. We take the tree \( T \) that achieves \( H(Y) = ET \), and want to show that it achieves the following inequalities:

\[
H(X) \leq ET < H(X) + 2
\]

We already proved the first inequality in a previous question. Show that the second inequality is equivalent to \( H(Y|X) < 2 \).

(c) Expanding the entropy of \( Y \), we have:

\[
H(Y) = - \sum_{i=1}^{m} \sum_{j \geq 1} p_i^{(j)} \log p_i^{(j)} = \sum_{i=1}^{m} \sum_{j: p_i^{(j)}>0} j 2^{-j}
\]

For \( i \in [1; m] \), we denote the corresponding term in the expansion by \( T_i \), i.e.:

\[
T_i = \sum_{j: p_i^{(j)}>0} j 2^{-j}
\]

Show that in order to prove the upper bound, it’s enough to prove that for all \( i \), \( T_i < -p_i \log p_i + 2p_i \).

(d) Denote by \( n \) the only integer such that: \( 2^{-(n-1)} > p_i \geq 2^{-n} \), so we can rewrite \( \sum_{j: p_i^{(j)}>0} \) into \( \sum_{j \geq n, p_i^{(j)}>0} \). Using the fact that \( p_i = \sum_{j} p_i^{(j)} \), show that \( T_i + p_i \log p_i - 2p_i < 0 \). Conclude.