Homework Assignment 2

Information Theory, TBSI

October 30, 2017

1. **AEP.** Let $X_1, X_2, \cdots$ be independent identically distributed random variables drawn according to the probability mass function $p(x), x \in \{1, 2, \cdots, m\}$. Thus $p(x_1, x_2, \cdots, x_n) = \prod_i^n p(x_i)$. We know that $-\frac{1}{n} logp(X_1, X_2, \cdots, X_n) \to H(X)$ in probability. Let $q(x_1, x_2, \cdots, x_n) = \prod_i^n q(x_i)$ where q is another probability mass function on $\{1, 2, \cdots, m\}$.

   (a) Evaluate
   $$\lim_{n \to +\infty} \frac{1}{n} logq(x_1, x_2, \cdots, x_n)$$
   where $X_1, X_2, \cdots$ are i.i.d. $\sim p(x)$.

   (b) Now evaluate the limit of the log likelihood ratio $\frac{1}{n} log \frac{q(X_1, X_2, \cdots, X_n)}{p(X_1, X_2, \cdots, X_n)}$ when $X_1, X_2, \cdots$ are i.i.d. $\sim p(x)$. Thus the odds favoring q are exponentially small when p is true.

2. **AEP.** i.i.d. random variables $X_i \sim p(x)$. Let $\mu = EX$, $H = -\sum p(x) logp(x)$, $A^n = \{x^n \in \mathcal{X}^n : | - \frac{1}{n} logp(x^n) - H| \leq \epsilon\}$, $B^n = \{x^n \in \mathcal{X}^n : |\frac{1}{n} \sum_{i=1}^n x_i - \mu| \leq \epsilon\}$. Proof that

   (a) $|A^n \cap B^n| \leq 2^{n(H+\epsilon)}$ for $\forall n$.

   (b) $|A^n \cap B^n| \geq (1/2) 2^{n(H-\epsilon)}$ for n large enough.

3. **AEP and Source Coding.** A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities $p(1) = 0.005$ and $p(0) = 0.995$. The digits are taken 100 at a time and a binary codeword is provided for every sequence of 100 digits containing three or fewer ones.

   (a) Assuming that all codewords are the same length, find the minimum length required to provide codewords for all sequences with three or fewer ones.

   (b) Calculate the probability of observing a source sequence for which no codeword has been assigned.

   (c) Use Chebyshev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (b).

4. **Generating Random Variables.** One wishes to generate a random variable X

$$X = \begin{cases} 0 & \text{with probability p} \\ 1 & \text{with probability 1-p} \end{cases}$$

You are given fair coin flips $Z_1, Z_2, \ldots$. Let N be the (random) number of fips needed to generate X . Find a good way to use $Z_1, Z_2, \ldots$ to generate X . Show that $\mathbb{E}[N] = 2$.

5. **Code's Type.**

   (a) What types are the following codes? (singular/nonsingular/uniquely decodable/instantaneous)

$$C1 = \{00, 01, 0\}$$
$$C2 = \{0, 010, 01, 10\}$$
$$C3 = \{0, 10, 110, 111\}$$
$$C4 = \{0, 10, 101, 100\}$$

   (b) Entropy of coded bits. Code $C : X \to \{0, 1\}^*$ be a nonsingular but nonuniquely decodable code. Let X have entropy $H(X)$. Compare $H(C(X))$ and $H(X)$. Compare $H(C(X^n))$ and $H(X^n)$.

6. **Wrong Code.** $l(x) = \lceil log\frac{1}{q(x)} \rceil$. Proof that expected code length of $p(x)$ satisfies

$$H(p) + D(p\|q) \leq E_p l(X) < H(p) + D(p\|q) + 1$$

7. **Merges.** Companies with values $W_1, W_2, \cdots, W_m$ are merged as follows. The two least valuable companies are merged, thus forming a list of $m?1$ companies. The value of the merge is the sum of the values of the two merged companies. This continues until one supercompany remains. Let V equal the sum of the values of the merges. Thus, V represents the total reported dollar volume of the merges. For example, if $W = (3, 3, 2, 2)$, the merges yield $(3, 3, 2, 2) \to (4, 3, 3) \to (6, 4) \to (10)$ and $V = 4 + 6 + 10 = 20$.

   (a) Argue that V is the minimum volume achievable by sequences of pairwise merges terminating in one supercompany. (Hint: Compare to Huffman coding.)

   (b) Let $W = \sum W_i, \hat{W}_i = W_i/W$, and show that the minimum merge volume V satisfies

$$WH(\hat{W}) \leq V \leq WH(\hat{W}) + W$$

8. **High-Low Game.**

(a) A computer generates a number X according to a known probability mass function $p(x), x \in \{1, 2, \cdots, 100\}$. The player asks a question, "Is $X = i$?" and is told "Yes", "You're too high" or "You're too low". He continues for a total of six questions. If he is right (i.e., he receives the answer "Yes") during this sequence, he receives a prize of value $v(X)$. How should the player proceed to maximize his expected winnings?

(b) The above doesn't have much to do with information theory. Consider the following variation: $X \sim p(x)$, prize $= v(x)$, $p(x)$ known, as before. But arbitrary Yes-No questions are asked sequentially until X is determined. ("Determined" doesn't mean that a "Yes" answer is received.) Questions cost one unit each. How should the player proceed? What is the expected payoff?

(c) Continuing (b), what if $v(x)$ is fixed, but $p(x)$ can be chosen by the computer (and then announced to the player)? The computer wishes to minimize the player's expected return. What should $p(x)$ be? What is the expected return to the player?

9. **Huffman codes.** One bottle of milk sours among six bottles. The probability p_i that milk in the i-th bottle sours is $(p_1, p_2, \cdots, p_6) = (7/26, 5/26, 4/26, 4/26, 3/26, 3/26)$. We can taste the milk to find out exactly which bottle has gone bad. Assume you can only taste one bottle at a time, how to decide the sequence of tasting to find out the sour bottle with minimum average number of tastings. (If the first five bottles in tasting are good, you don't have to taste the soured sixth one.)

   (a) What is the minimum expected number of tasting?

   (b) Which bottle you should taste first?

   (c) Assume you can mix the six bottles for tasting to find out the soured one. How to mix those bottles and taste to achieve the minimum expected number of tasting.

   (d) At the first time, which bottles you should mix?

10. **Huffman codes with costs.** Words such as Run!, Help!, and Fire! are short, not because they are used frequently, but perhaps because time is precious in the situations in which these words are required. Suppose that $X = i$ with probability $p_i, i = 1, 2, \cdots, m$. Let l_i be the number of binary symbols in the codeword associated with $X = i$, and let c_i denote the cost per letter of the codeword when $X = i$. Thus, the average cost C of the description of X is $C = \sum_{i=1}^{m} p_i c_i l_i$.

    (a) Minimize C over all $l_1, l_2, \cdots, l_m$ such that $\sum 2^{-l_i} \leq 1$. Ignore any implied integer constraints on l_i. Exhibit the minimizing $l_1^*, l_2^*, \cdots, l_m^*$ and the associated minimum value C^*.

(b) How would you use the Huffman code procedure to minimize C over all uniquely decodable codes? Let $C_{Huffman}$ denote this minimum.

(c) Can you show that

$$C^* \leq C_{Huffman} < C^* + \sum_{i=1}^{m} p_i c_i$$