Information and Coding Theory **Autumn 2022**

Homework 1

Due: October 16, 2022

Note: *You may discuss these problems in groups. However, you must write up your own solutions and mention the names of the people in your group. Also, please do mention any books, papers or other sources you refer to. It is recommended that you typeset your solutions in LaTeX.*

1. **World Series (Problem 2.18 from the book).** **[4 points]**

   Two teams A and B play a series of up to 5 games, in which the team to win 3 games wins the series. Let X be a random variable which is a sequence of letters corresponding to the winners of each of the games played - possible values for X then include AAA, $ABBAB$ etc. Let Y be the number of games played (the teams play till the series winner is decided). Calculate $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$ and $I(X;Y)$. Assume both teams are equally likely to win each game independent of any previous games.

2. **Lost in transmission.** **[4 points]**

   n people, say $A_1, \ldots, A_n$ (sitting in a circle) play a game in which A_1 gives a message to A_2, A_2 passes it to A_3, A_3 to A_4 and so on. Finally, A_n passes the message she received back to A_1. Let us assume for simplicity that the message passed by A_1 is a random variable X_1 which is 0 or 1 with equal probability. Let X_i be the message passed by the person A_i: assume that person A_i pass the message they received correctly ($X_i = X_{i-1}$) with probability $1 - \varepsilon$ and get confused and pass the opposite message ($X_i = \overline{X_{i-1}}$) with probability ε. Calculate $I(X_1; X_n)$.

3. **Entropy and friends.** **[3+3 = 6 points]**

   Prove the following basic identities about the quantities we have studied so far:

   (a) Let X be a random variable distributed according to the distribution P on a finite universe $\mathcal{X}$, and let Q be the uniform distribution on $\mathcal{X}$. Then

   $$D(P||Q) \;=\; \log|\mathcal{X}| - H(X)\,.$$

   (b) Let X, Y be random variables jointly distributed according to the distribution $P(X,Y)$. Let $P(X)$ and $P(Y)$ denote the marginal distributions for the variables X and Y. Then

   $$I(X;Y) \;=\; D(P(X,Y)||P(X)P(Y))\,.$$

4. **Three's a crowd.** **[2+4 = 6 points]**

   Currently, there is no known good notion of the mutual information between three random variables X, Y and Z. One possible definition is given as follows: thinking of entropy of a variable $H(X)$ as the "single variable mutual information" $I(X)$, we can write the two-variable mutual information $I(X;Y)$ as $I(X;Y) = I(X) - I(X|Y)$. We extend this to define

   $$I(X;Y;Z) \;=\; I(X;Y) - I(X;Y|Z)\,.$$

   (a) Show that $I(X;Y;Z)$ is symmetric in X, Y, Z. In particular:

   $$I(X;Y;Z) = H(XYZ) - H(XY) - H(YZ) - H(ZX) + H(X) + H(Y) + H(Z)\,.$$

   (b) Give an example of three random variables X, Y, Z such that $I(X;Y;Z) < 0$.

5. **Measures of independence.** **[4 points]**

   We have seen $I(X;Y)$ is a measure of how much the distribution of Y is affected by conditioning on X. Let $P(X,Y)$ be the joint distribution of X and Y. Consider the following quantity, which is the expected distance between the original distribution of Y and the one obtained conditioning on X

   $$\rho(Y|X) \;=\; \mathop{\mathbb{E}}_{x}\left[\|P(Y|X=x) - P(Y)\|_1\right]\,,$$

   where the expectation over X is according to the marginal distribution $P(X)$. Prove that

   $$\rho(Y|X) \le \sqrt{2\ln 2 \cdot I(X;Y)}\,.$$

6. **Energy-aware Kraft's inequality.** **[6 points]**

   Suppose we have a channel where a 0 takes 1 unit of energy to transmit and 1 takes 2 units of energy to transmit. Suppose there exists a prefix-free code for a universe $\mathcal{X} = \{a_1, \ldots, a_n\}$ such that the codeword for a_i takes e_i units of energy to transmit. Show that

   $$\sum_{i=1}^{n} \left(\frac{\sqrt{5}-1}{2}\right)^{e_i} \;\le\; 1\,.$$

7. **Extra problem (no need to submit): Counting homomorphisms.**

   In this problem, we will see that Shearer's lemma can be used to give a tight bound for the maximum number of ways of embedding a graph G (of constant size) into a graph H with at most m edges. This was originally proved by Alon and the proof outlined here is due to Friedgut and Kahn.

   Let $G = (V_G, E_G)$ be a given undirected graph of constant size. For an undirected graph $H = (V_H, E_H)$, let $\mathcal{E}(G,H)$ denote the number of embeddings of G in H i.e., the number of maps $f : V_G \to V_H$ such that for all $(i,j) \in E_G$, we have $(f(i), f(j)) \in E_H$. Let $\mathcal{E}(G,m)$ denote the maximum of $\mathcal{E}(G,H)$ over all graphs H with at most m edges. We will show that

$$c_1 \cdot m^{\alpha^*(G)} \;\leq\; \mathcal{E}(G,m) \;\leq\; c_2 \cdot m^{\alpha^*(G)},$$

where c_1 and c_2 are constants depending on the graph G, and $\alpha^*(G)$ is a parameter known as the fractional independent set number of the graph.

(a) For a graph G, the quantity $\alpha^*(G)$ is defined to the optimal solution to the following linear programming relaxation for the maximum independent set (largest set of vertices not containing any edges).

$$\begin{array}{lll} \text{maximize:} & \displaystyle\sum_{i \in V_G} x_i & \\ \text{subject to:} & x_i + x_j \leq 1 & \forall e = (i,j) \in E_G \\ & x_i \in [0,1] & \forall i \in V_G \end{array}$$

Show that if we restrict x_i to only take values 0 or 1, then the above is the same as the maximum independent set problem. Show that the following linear program is the dual of the above, and hence also has optimum value equal to $\alpha^*(G)$ by LP duality. You may assume that all vertices in G have degree at least 1 (why?) to simplify the above LP before writing it's dual.

$$\begin{array}{lll} \text{minimize:} & \displaystyle\sum_{e \in E_G} y_e & \\ \text{subject to:} & \displaystyle\sum_{e \ni i} y_e \geq 1 & \forall i \in V_G \\ & y_e \in [0,1] & \forall e \in V_G \end{array}$$

(b) Let x be an optimal solution to the first LP. We will use it to prove a lower bound on $\mathcal{E}(G,m)$ by constructing a graph H. Let $|V_G| = k$ and $|E_G| = \ell$. The vertices of H will consist of k disjoint sets $V_1, \ldots, V_k$ of sizes

$$|V_i| \;=\; \left(\frac{m}{|E_G|}\right)^{x_i}.$$

We add a complete bipartite graph between V_i and V_j whenever $(i,j) \in E_G$. Show that the graph H constructed as above has at most m edges. Also show that

$$\mathcal{E}(G,H) \;\geq\; c_1 \cdot m^{\alpha^*(G)},$$

where c_1 is a constant depending on G.

(c) Finally, we upper bound $\mathcal{E}(G,m)$. For any graph H with at most m edges, let $(F(1),\ldots.F(k))$ denote a random embedding of G in H. Let y be an optimal solution to the dual LP. Use y to construct an appropriate distribution over pairs of random variables $(F(i),F(j))$ and use Shearer's lemma to bound $H(F(1),\ldots.F(k))$. Show that this gives

$$\mathcal{E}(G,m) \;\leq\; c_2 \cdot m^{\alpha^*(G)}\,,$$

where c_2 is a constant depending on G.