

Homework 1

Due: October 28, 2014

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).** 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.** n people, say A_1, \dots, 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.** 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 U , and let Q be the uniform distribution on U . Then

$$D(P||Q) = \log |U| - H(X).$$

- (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.** There is no 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.** 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

$$\rho(Y|X) = \mathbb{E}_x [\|P(Y|X = x) - P(Y)\|_1],$$

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

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

6. **Matching families.** Recall that a perfect matching on $2n$ vertices is a collection of n edges such that each vertex has exactly one edge incident to it. Let \mathcal{F} be a family of graphs on $2n$ vertices such that for any two graphs $G_1, G_2 \in \mathcal{F}$, $G_1 \cap G_2$ contains a perfect matching.

(a) Prove that $|\mathcal{F}| \leq 2^{\binom{2n}{2} - n}$.

(b) Prove that the above bound is tight.

7. **Energy-aware Kraft's inequality.** Suppose we have a channel where a 0 takes 1 unit of energy to transmit and 1 takes 2 units of energy transmit. Suppose there exists a prefix-free code for a universe $U = \{a_1, \dots, a_n\}$ such that the codeword for a_i takes e_i units of energy to transmit. Show that

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