Carnegie Mellon University

15-859: Information Theory and Applications in TCS — Spring 2013

PROBLEM SET 1
Due by Thursday, February 7

INSTRUCTIONS

- You are allowed to collaborate with up to two other students taking the class in solving problem sets. But here are some rules concerning such collaboration:
  1. You should think about each problem by yourself for at least 30 minutes before commencing any collaboration.
  2. Collaboration is defined as discussion of the lecture material and solution approaches to the problems. Please note that *you are not allowed to share any written material and you must write up solutions on your own without any "collaboration notes" as an aid.*
  3. You must clearly acknowledge your collaborator(s) in the write-up of your solutions.
  4. Of course, if you prefer, you can also work alone (see the last bullet item for some "credit" for doing so).
- Solutions typeset in LaTeX are strongly preferred.
- You should *not* search for solutions on the web. More generally, you are urged to try and solve the problems without consulting *any* reference material other than the course notes and what we cover in class. If for some reason you feel the need to consult some source, *please acknowledge the source* and try to articulate the difficulty you couldn't overcome before consulting the source and how it helped you overcome that difficulty. Alternatively, before turning to any such material, we encourage you to ask us for hints or clarifications.
- Please start work on the problem set early. The problem set has **six** problems and is worth a total of 100 points. As a rather rough guess/estimate, scoring around 80% of the points, or 70% of the points if you work by yourself, might suffice for an A on this problem set.

---

1. (15 points) Two teams A and B play a best-of-five series that terminates as soon as one of the teams wins three games. Let X be the random variable that represents the outcome of the series written as a string of who won the individual games — possible values of X are AAA, $BAAA$, $ABABB$, etc. Let Y be the number of games played before the series ends. Assuming that A and B are equally matched and the outcomes of different games in the series are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, $H(X|Y)$, and $I(X;Y)$.

2. (15 points)

   (a) Consider a sequence of n random variables $B_1, B_2, \ldots, B_n$, each taking values in $\{0, 1\}$, such that sequences with an odd number of 1s have probability $2^{-(n-1)}$ each, and those with an even number of 1s have probability 0. Calculate the mutual informations

$$I(B_1; B_2), \;\; I(B_2; B_3 \mid B_1), \;\; \cdots \;\; , I(B_{n-1}; B_n \mid B_1, B_2, \ldots, B_{n-2}) \; .$$

(b) For joint random variables X, Y, Z does the "submodular" inequality

$$H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$$

always hold? When, if at all, does equality hold?

3. (15 points) Let X, Y be integer-valued random variables and let $Z = X + Y$.

(a) Prove that $H(Z \mid X) = H(Y \mid X)$. (Hint: Expand $H(Z \mid X)$ using the definition of conditional entropy.)

(b) Prove that if X, Y are independent, then $H(Z) \geq \max\{H(X), H(Y)\}$. That is, addition of independent random variables increases entropy.

(c) Give an example of random variables X, Y for which $H(Z) < \min\{H(X), H(Y)\}$.

(d) State and prove a necessary and sufficient condition for when the entropy of the sum equals the sum of the entropies, i.e., $H(Z) = H(X) + H(Y)$.

4. (15 points) For $\tau \in (0, 1/2)$, define a subset $C \subset \{0,1\}^n$ to be τ-*covering* if every $\mathbf{r} \in \{0,1\}^n$ is within Hamming distance τn from some element C.

(a) Prove, using the language of entropy and conditional entropy, that the size of such a τ-covering code must satisfy $|C| \geq 2^{(1-h(\tau))n}$. Here $h(\tau)$ denotes the binary entropy function $h(\tau) = \tau \log_2(1/\tau) + (1-\tau)\log_2(1/(1-\tau))$. (You may use without proof the inequality $\sum_{j=0}^{\tau n} \binom{n}{j} \leq 2^{h(\tau)n}$.)

(b) Prove that for large enough n, a random subset of $\{0,1\}^n$ of size size $n^3 \cdot 2^{(1-h(\tau))n}$ is τ-covering with probability at least $1 - 2^{-\Omega(n)}$. (You may use without proof the inequality $\binom{n}{\tau n} \geq 2^{h(\tau)n}/n$ for large n.)

5. (30 points) Let X be a random variable taking values in an alphabet $\{a_1, a_2, \ldots, a_n\}$ with the probability of $X = a_i$ being p_i for $i = 1, 2, \ldots, n$. Assume that the probabilities are sorted $0 < p_1 \leq p_2 \leq \cdots \leq p_n$. Consider the following natural procedure to build a prefix-free code for these n symbols. Choose a $k \in \{1, 2, \ldots, n-1\}$ such that $|\sum_{i=1}^{k} p_i - \sum_{i=k+1}^{n} p_i|$ is minimized. Assign 0 for the first bit of the encoding for source symbols $a_1, \ldots, a_k$, and 1 for the first bit of the encoding for source symbols $a_{k+1}, \ldots, a_n$. Repeat the process recursively for each of the two subsets $\{a_1, \ldots, a_k\}$ and $\{a_{k+1}, \ldots, a_n\}$. By this recursive procedure, we obtain a prefix-free code for the symbols $a_1, a_2, \ldots, a_n$.

The goal of this exercise is to prove the expected length L of the resulting source code is close to $H(X)$. To this end, we will view the prefix-free code naturally as a binary tree, with the symbols at the n leaves, as described in lecture.

(a) Argue that in the above construction, the leaves in the subtree rooted at any internal node will consist of a consecutive subset $\{a_i, a_{i+1}, \ldots, a_j\}$ of symbols for some $1 \leq i < j \leq n$.

We will denote such an internal node as $[i, j]$, and use the shorthand $q_{[i,j]} = p_i + p_{i+1} + \cdots + p_j$ for the total probability of leaves in its subtree. Note that $[i, i]$ is just the leaf with symbol a_i.

(b) Let $\mathcal{I}$ denote the set of internal nodes of the tree. Prove that the expected length L of the above source code is

$$L = \sum_{[i,j] \in \mathcal{I}} q_{[i,j]} \, .$$

(c) Prove that

$$H(X) = \sum_{[i,j]\in\mathcal{I}} q_{[i,j]} h\Big(\frac{q_{[i,k]}}{q_{[i,j]}}\Big)$$

where k, $i \le k < j$, is such that $[i,k]$ and $[k+1,j]$ are the left and right children of internal node $[i,j]$, and $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the entropy function on the interval $[0,1]$.

(d) Using the inequality $h(x) \ge 2x$ for $x \in [0, 1/2]$, deduce that

$$L - H(X) \le \sum_{[i,j]\in\mathcal{I}} \Big| q_{[i,k]} - q_{[k+1,j]} \Big| \, ,$$

where again $[i,k]$ and $[k+1,j]$ are the left and right children of internal node $[i,j]$.

(e) So far what we have said applies for arbitrary choices of k, $i \le k < j$, to branch at each internal node $[i,j]$. In order to analyze the effect of making the most balanced split, prove that if k, $i \le k < j$ minimizes $\big|q_{[i,k]} - q_{[k+1,j]}\big|$, then this minimum is in fact at most $\max\{p_k, p_{k+1}\}$. Or more formally,

$$\min_{\ell: i\le \ell < j} \Big| q_{[i,\ell]} - q_{[\ell+1,j]} \Big| \le \max\{p_k, p_{k+1}\}$$

where $k = \arg\min_{\ell: i\le\ell<j} \Big| q_{[i,\ell]} - q_{[\ell+1,j]} \Big|$.

(f) Finally, put the previous two parts together and show that $L \le H(X) + 2$.

6. (10 points) Let X, Y be random variables taking values in sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. For $x \in \mathcal{X}$, define

$$d(x) = \sum_{y\in\mathcal{Y}} \Big| \Pr[Y = y \mid X = x] - \Pr[Y = y] \Big|$$

which intuitively measures how much the distribution of Y is "affected" given the knowledge of $X = x$. It is thus reasonable to expect that $d(x)$ is small on average when X does not reveal much information about Y. Prove that this is indeed the case by establishing the inequality

$$\mathop{\mathbb{E}}_{x\leftarrow X} [d(x)] \le \sqrt{2\, I(X;Y)} \, .$$

Hint: You may use *Pinsker's inequality* which states that for any two distributions p, q defined on a set A:

$$D(p||q) \ge \frac{1}{2}\Big(\sum_{a\in A} |p(a) - q(a)|\Big)^2 \, .$$