This is intended to provide a sketch of a solution to every problem. There are often many possible solutions to each problem that receive credit. This is not intended to be a fully rigorous "model" for submissions - you should be able to reconstruct the full solution from this sketch, but keep in mind we expect a fully elaborated solution for your actual submissions.

1: Problem 1 [25 pts]

(a) 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 $H(X), H(Y), H(Y|X)$, and $H(X|Y)$.

Solution

Calculate:

$$\Pr(Y=3)=\frac{1}{4}, \Pr(Y=4)=\frac{3}{8}, \Pr(Y=5)=\frac{3}{8}$$

Therefore

$$H(Y)=\frac{1}{4}\lg 4+2\cdot\frac{3}{8}\lg\frac{8}{3}=\frac{11}{4}-\frac{3}{4}\lg 3$$

Obviously $H(Y|X)=0$. Now calculate

$$H(X|Y)=\sum_{y=3}^{5}\Pr(Y=y)H(X|Y=y)$$

$$H(X|Y)=\frac{1}{4}\lg 2+\frac{3}{8}\lg 6+\frac{3}{8}\lg 12=\frac{11}{8}+\frac{3}{4}\lg 3$$

Finally, $H(X)=H(X|Y)+H(Y)=\frac{33}{8}$ (since $H(Y|X)=0$).

(b) Let X, Y be integer-valued random variables and let $Z=X+Y$. Prove that $H(Z|X)=H(Y|X)$. (Hint: Expand $H(Z|X)$ using the definition of conditional entropy.)

Solution

$$H(Z|X)=\mathbb{E}_x[H(X+Y|X=x)]$$

$$H(Z|X)=-\mathbb{E}_x[\sum_{z\in\mathbb{Z}}\Pr(X+Y=z|X=x)\lg\Pr(X+Y=z|X=x)]$$

$$H(Z|X)=-\mathbb{E}_x[\sum_{z-x\in\mathbb{Z}}\Pr(Y=z-x|X=x)\lg\Pr(Y=z-x|X=x)]$$

Change variables $y:=z-x$ and observe the expression inside the expectation is $H(Y|X=x)$.

(b.1) Let X, Y, Z be as defined in (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.

Solution

$$H(Z)\geq H(Z|X)=H(Y|X)=H(Y)$$

Likewise for X.

(b.2) Let X, Y, Z be as defined in (b). Give an example of random variables X, Y for which $H(Z) < \min\{H(X), H(Y)\}$

Solution

Many possible solutions, for example $X \sim Bern(\frac{1}{2}), Y = -X$

(b.3) State and prove a necessary and sufficient condition for when the entropy of the sum equals the sum of the entropies, i.e., $H(Z) = H(X) + H(Y)$.

Solution

$$H(X, Y, Z) = H(X, Y) + H(Z|X, Y) = H(X, Y|Z) + H(Z)$$

Obviously $H(Z|X, Y) = 0$. A useful identity is $H(X, Y) = H(X) + H(Y) - I(X; Y)$.

$$H(X) + H(Y) - I(X; Y) = H(X, Y|Z) + H(Z)$$

$$H(Z) = H(X) + H(Y) \iff I(X; Y) + H(X, Y|Z) = 0$$

Entropy and mutual information are always nonnegative so the necessary/sufficient conditions are

(1) $I(X; Y) = 0 \implies X, Y$ must be independent

(2) $H(X, Y|Z) = 0 \implies$ any value of Z uniquely determines **both** X and Y. E.g. $X \in_R \{1, 2, 3\}, Y \in_R \{100, 200, 300\}$

2: Problem 2 [20 pts]

In this exercise, we will prove "Fano's Inequality", which informally states that a random variable $\hat{X}$ that predicts X with high probability, must also "sip" almost all of the entropy out of X.

More formally, let X be an arbitrary random variable that takes values in $[n] = \{1, 2, ..., n\}$, and suppose that $\hat{X}$ is a random variable satisfying:

$$\Pr(\hat{X} = X) \geq 1 - \epsilon$$

Prove that in this case, $H(X|\hat{X}) \leq H(\epsilon) + \epsilon \cdot \log(n - 1)$, where $H(\epsilon)$ is the binary entropy of ϵ.

Solution

Consider the 'error indicator' r.v. $E := \mathbf{1}(X \neq \hat{X})$. WLOG assume $\varepsilon \leq \frac{1}{2}$, then $H(E) \leq H(\varepsilon)$. We have

$$H(X|\hat{X}) = H(E, X|\hat{X}) = H(X|E, \hat{X}) + H(E|\hat{X})$$

$$H(X|E, \hat{X}) = \Pr(E = 1)H(X|E = 1, \hat{X}) + \Pr(E = 0)H(X|E = 0, \hat{X})$$

$$H(E|\hat{X}) \leq H(E) \leq H(\varepsilon)$$

We know $\Pr(E = 1) \leq \varepsilon, H(X|E = 0, \hat{X}) = 0$. Putting it all together,

$$H(X|\hat{X}) = H(E, X|\hat{X}) \leq \varepsilon H(X|E = 1, \hat{X}) + H(\varepsilon)$$

Finally, use the support bound to conclude $H(X|E = 1, \hat{X}) \leq \lg(n - 1)$ (every value **other** than the correct value) and hence:

$$H(X|\hat{X}) \leq \varepsilon \log(n - 1) + H(\varepsilon)$$

3: Problem 3 [20 pts]

For $\tau \in (0, \frac{1}{2})$, 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 must satisfy $|C| \geq 2^{(1-H(\tau))n}$, where $H(\tau)$ denotes the binary entropy function with parameter τ. (Hint: Use the inequality we proved in class: $\sum_{j=0}^{\tau n} \binom{n}{j} \leq 2^{nH(\tau)}$).

Solution

Suppose C is the smallest set that τ-covers $\{0,1\}^n$. Each element $y \in C$ τ-covers $\sum_{j=0}^{\tau n} \binom{n}{j} \leq 2^{nH(\tau)}$ elements in $\{0,1\}^n$. Since there are 2^n elements in $\{0,1\}^n$, we need at least

$$|C| \geq \frac{2^n}{2^{nH(\tau)}} = 2^{(1-H(\tau))n}$$

such elements $y \in C$ to cover the whole set $\{0,1\}^n$.

(b) Prove that for any $\tau \in (0, \frac{1}{4})$ and large enough n, a random subset of $\{0,1\}^n$ of size $n^3 \cdot 2^{(1-H(\tau))n}$ is τ-covering with probability at least $1 - 2^{-\Omega(n)}$. (Hint: You may use without proof the inequality $\binom{n}{\tau n} \geq 2^{H(\tau)n}/n$. You can also use without a proof the Chernoff bound: If $X_1, \ldots, X_n$ are i.i.d s.t $X_i \sim Ber(p)$, then $Pr[\sum_i X_i \notin (1 \pm \varepsilon)pn] \leq 2^{-\varepsilon^2 pn/4}$).

Solution

Choose C uniformly at random from all subsets of $\{0,1\}^n$ of size $n^3 2^{(1-H(\tau))n}$. Define $h(x,y)$ as the Hamming distance between $x, y \in \{0,1\}^n$. Fix some $x \in_R \{0,1\}^n$ and consider the τn-ball $B_x^{\tau n}$ centered at x. As above, there are $|B_x^{\tau n}| = \sum_{j=0}^{\tau n} \binom{n}{j} \leq 2^{H(\tau)n}/n$ such vectors. For each element x_i in $B_x^{\tau n}$, define the indicators $X_i := \mathbf{1}\{x_i \in C\}, X := \sum_i X_i$. Then each X_i is i.i.d. Bernoulli for some p and

$$np = \mathbb{E}[X] \geq \frac{|C|}{2^n} \cdot \frac{2^{H(\tau)n}}{n} = n^2$$

Now apply the Chernoff bound with $\varepsilon = 1 - \frac{1}{np} = 1 - \frac{1}{n^2}$,

$$\Pr(\text{no } x_i \in B_x^{\tau n} \text{ in } C) = \Pr(X < 1) \leq 2^{-(1-\frac{1}{n^2})^2 \cdot \frac{n^2}{4}} = 2^{-\Omega(n^2)}$$

Finally, apply the union bound

$$\Pr(\exists x. \text{no } x_i \in B_x^{\tau n} \text{ in } C) = \sum_x \Pr(\text{no } x_i \in B_x^{\tau n} \text{ in } C) \leq 2^n \cdot 2^{-\Omega(n^2)} = 2^{-\Omega(n^2)}$$

But $\Pr(\exists x. \text{no } x_i \in B_x^{\tau n} \text{ in } C) \equiv \Pr(\text{C does not } \tau\text{-cover } \{0,1\}^n)$ so
$\Pr(\text{C } \tau\text{-covers } \{0,1\}^n) = 1 - 2^{-\Omega(n^2)}$

4: Problem 4 [35 pts]

Let X be a random variable taking values in an alphabet $\{a_1, a_2, ..., a_n\}$ with the probability of $X = a_i$ being p_i for $i = 1, 2, ..., n$. 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 $k \in \{1, 2, ..., n1\}$ 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 $a_1, ..., a_k$, and 1 for the first bit of the encoding for source symbols $a_{k+1}, ..., a_n$. Repeat the process recursively for each of the two subsets $\{a_1, ..., a_k\}$ and $\{a_{k+1}, ..., a_n\}$. By this recursive procedure, we obtain a prefix-free code for the symbols $a_1, a_2, ..., a_n$.

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 $\{a_i, a_{i+1}, ..., a_j\}$ 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.

Solution

By induction.

(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

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

Solution

Fix a codeword A. Then the length of A is simply

$$L(A) = \sum_{[i,j] \in \mathcal{I}} \mathbf{1}\{\text{node } [i, j] \text{ in } A\}$$

Take the expectation of $L(A)$ over all codewords A; by linearity of expectation, conclude

$$L = \mathbb{E}_A[L(A)] = \sum_{[i,j] \in \mathcal{I}} q_{[i,j]}$$

(c) Prove that

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

where $k, i \leq k < j$ is such that $[i, k]$ and $[k + 1, j]$ are the left and right children of internal node $[i, j]$ and $H(p)$ is the binary entropy function.

Solution [credit to Natania Wolansky]

Let d be the depth of the binary tree, and let X_ℓ be the ℓth node from root to leaf X. Then by the chain rule,

$$H(X) = \sum_{\ell=1}^{d} H(X_\ell | X_{<\ell})$$

We need to calculate $H(X_\ell|X_{<\ell}) = \mathbb{E}_I\{H(X_\ell|X_{<\ell} = I)\}$. Let $I = [i,j]$ split into $[i,k],[k+1,j]$, then

$$H(X_\ell|X_{<\ell} = I) = H(\frac{q_{[i,k]}}{q_{[i,j]}})$$

Hence at level ℓ of the tree

$$H(X_\ell|X_{<\ell}) = \mathbb{E}_I\{H(X_\ell|X_{<\ell} = I)\} = \sum_{[i,j]\in\mathcal{I}_\ell} q_{[i,j]} H(\frac{q_{[i,k]}}{q_{[i,j]}})$$

Hence

$$H(X) = \sum_{\ell=1}^{d} \sum_{[i,j]\in\mathcal{I}_\ell} q_{[i,j]} H(\frac{q_{[i,k]}}{q_{[i,j]}}) = \sum_{[i,j]\in\mathcal{I}} q_{[i,j]} H(\frac{q_{[i,k]}}{q_{[i,j]}})$$

(d) Using the equality $H(p) \geq 2p$ for $p \in [0, \frac{1}{2}]$, deduce that:

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

Solution

From (b) and (c):

$$L - H(X) = \sum_{[i,j]\in\mathcal{I}} q_{[i,j]}(1 - H(\frac{q_{[i,k]}}{q_{[i,j]}}))$$

$$L - H(X) \leq \sum_{[i,j]\in\mathcal{I}} q_{[i,j]}(1 - 2\min(\frac{q_{[i,k]}}{q_{[i,j]}}, \frac{q_{[k+1,j]}}{q_{[i,j]}}))$$

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

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

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

Solution

By contradiction.

(f) Finally, put parts (d) and (e) together to show that $L \leq H(X) + 2$.

Solution

From (d) and (e):

$$L - H(X) \leq \sum_{[i,j]\in\mathcal{I}} |q_{[i,k]} - q_{[k+1,j]}| \leq \sum_{k=0}^{n} \max\{p_k, p_{k+1}\}$$

Define $p_0, p_{n+1} = 0$ for consistency. Then:

$$L - H(X) \leq \sum_{k=0}^{n} \max\{p_k, p_{k+1}\} \leq \sum_{k=0}^{n} p_k + p_{k+1} = 2$$

5: Problem 5 [20 pts]

Let $a < b$ be any two integers. Prove that in any undirected graph G,

$$(b!n_b)^a \leq (a!n_a)^b$$

where n_b denotes the number of cliques of size b in G, and n_a denote the number of cliques of size a in G (where permutations count as distinct copies of a subgraph).

Solution

Choose a b-clique $B = (B_1, ..., B_b)$ uniformly at random from G. Then $H(B) = \lg(b!n_b)$. Choose u.a.r. a subset S from $[b]$, $|S| = a$. Each $i \in [b]$ is in S with probability $\frac{a}{b}$ by symmetry. By Shearer's lemma,

$$\frac{a}{b}H(B) \leq \mathbb{E}_S\{H(B_S)\}$$

Each $H(B_S) \leq \lg|\text{supp}(B_S)|$ so this must also be true of $\mathbb{E}_S\{H(B_S)\}$. Furthermore $\text{supp}(B_S) \subseteq \text{supp}(A) \implies |\text{supp}(B_S)| \leq |\text{supp}(A)|(*)$, where A is an a-clique chosen u.a.r. from G. Putting it all together, we have

$$\frac{a}{b}\lg(b!n_b) \leq \mathbb{E}_S\{H(B_S)\} \leq \lg|\text{supp}(B_S)| \leq \lg(a!n_a)$$

Hence $(b!n_b)^a \leq (a!n_a)^b \square$

Note: Many submissions missed the step $(*)$. Remember, B_S is an a-clique chosen from within the b-clique you've already chosen - it doesn't necessarily have the same support as A since there could be a-cliques not part of any b-clique. Recall elements in the support of a distribution must have **non-zero** probability.

Acknowledgement. Problems 1,3 and 4 are borrowed from Vankat Guruswami's problem sets.