Problem Set 4

15-859 Information Theory and Applications in TCS
Name: Shashank Singh
Email: sss1@andrew.cmu.edu
Due: Wednesday, May 1, 2013
Collaborators: Albert Gu

Problem 1

Since each $j \in [n]$ is in at least k of the sets $T_1, \dots, T_m$, for $i \in [m]$, $\Pr(j \in T_i) \geq k/m$. Thus, by Shearer's Lemma, if X and i are chosen uniformly at random from S and $[m]$, respectively,

$$\frac{k}{m} \log_2 |S| = \frac{k}{m} H(X) \leq \mathop{\mathbb{E}}_i [H(X_{T_i})] \leq \mathop{\mathbb{E}}_i [\log_2 n_i] \leq \frac{1}{m} \sum_{i \in [m]} \log_2 n_i$$

(since X_{T_i} is distributed over the n_i values in T_i). Multiplying by m gives

$$|S|^k = 2^{k \log_2 |S|} \leq 2^{\sum_{i \in [m]} \log_2 n_i} = \prod_{i \in [m]} n_i. \quad \blacksquare$$

Problem 2

(a) Let $E_S := \{(x, y) \in E : x, y \in S\}$. For any $X \in \{0,1\}^m, i \in [m], b \in \{0,1\}$, define

$$X^{-i} := (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m) \in \{0,1\}^{m-1}$$

and

$$X^{-i}(b) := (X_1, \dots, X_{i-1}, b, X_{i+1}, \dots, X_m) \in \{0,1\}^m.$$

If X is chosen uniformly at random from S, then, since $X \in \{X^{-i}(0), X^{-i}(1)\}$, $H(X_i \,|\, X^{-i}) = 1$ if and only if the edge $(X^{-i}(0), X^{-i}(1)) \in E_S$, and $H(X_i \,|\, X^{-i}) = 0$ otherwise. Thus,

$$|E_S| = \frac{1}{2} \sum_{Y \in S} \sum_{i=1}^{m} H(X_i \,|\, X^{-i} = Y^{-i}).$$

(the sum counts each edge twice - once for each endpoint). Then, by the Chain Rule and the fact that conditioning cannot increase entropy,

$$\begin{aligned} |E_S| &= \frac{1}{2} \sum_{Y \in S} \sum_{i=1}^{m} H(X_i \,|\, X^{-i} = Y^{-i}) \\ &\leq \frac{1}{2} \sum_{Y \in S} \sum_{i=1}^{m} H(X_i \,|\, X_1 = Y_1, \dots, X_{i-1} = Y_{i-1}) \\ &= \frac{1}{2} \sum_{Y \in S} H(X) = \frac{1}{2} |S| H(X) = \frac{1}{2} |S| \log_2 |S|. \quad \blacksquare \end{aligned}$$

(b) It doesn't seem like the bound can be tightened. Consider, for instance, when S is an embedding of the $\frac{m}{2}$-dimensional hypercube into H (so $|S| = 2^{m/2}$). Then, since an n-dimensional hypercube has $n2^{n-1}$ edges, the bound

$$\frac{|S|\log_2|S|}{2} = \frac{m}{2}2^{m/2-1}$$

on the number of edges with both endpoints in S is tight. ■

Problem 3

We showed in lecture that, for random variables $X \sim P$ and $Y \subseteq V(G)$ independent with $X \in Y$,

$$H(G,P) = \min_{(X,Y)} I(X;Y).$$

Thus, it suffices to give a distribution for Y with

$$I(X;Y) \leq \log_2 \chi(G).$$

Pick a $\chi(G)$-coloring of G, and distribute Y so as to be the set of vertices with the same color as X (clearly, Y is independent and $X \in Y$). Then, since $|\operatorname{supp}(Y)| \leq \chi(G)$,

$$I(X;Y) \leq H(Y) \leq \log_2|\operatorname{supp}(Y)| \leq \log_2 \chi(G). \quad ■$$

Problem 4

(a) Suppose (X,Y) is distributed so as to minimize $I(x;Y)$ with $X \sim P$, $X \in Y \in I(G)$. For all $S \in I(G)$, $\alpha_S := \mathsf{Pr}(Y = S)$, and note that each $\alpha_S \geq 0$ and $\sum_{I \in I(G)} \alpha_S = 1$. Then,

$$(a_1,\ldots,a_n) := \sum_{S\in I(G)} \alpha_S \cdot \chi(S) \in VP(G),$$

and each $a_i = \sum_{i\in S\in I(G)} \alpha_S = \sum_{i\in S\in I(G)} \mathsf{Pr}(Y = S)$. By Jensen's Inequality and concavity of $\log_2$,

$$\begin{aligned}
I(X;Y) &= \sum_{i\in[n]} \sum_{i\in S\in I(G)} \mathsf{Pr}(X=i)\,\mathsf{Pr}(Y=S\,|\,X=i)\log_2\left(\frac{\mathsf{Pr}(Y=S\,|\,X=i)}{\mathsf{Pr}(Y=S)}\right)\\
&= -\sum_{i\in[n]} p_i \sum_{i\in S\in I(G)} \mathsf{Pr}(Y=S\,|\,X=i)\log_2\left(\frac{\alpha_S}{\mathsf{Pr}(Y=S\,|\,X=i)}\right)\\
&\geq -\sum_{i\in[n]} p_i \log_2\left(\sum_{i\in S\in I(G)} \mathsf{Pr}(Y=S\,|\,X=i)\cdot\frac{\alpha_S}{\mathsf{Pr}(Y=S\,|\,X=i)}\right)\\
&= -\sum_{i\in[n]} p_i \log_2\left(\sum_{i\in S\in I(G)} \alpha_S\right) = -\sum_{i\in[n]} p_i \log_2 a_i = \sum_{i\in[n]} p_i \log_2\left(\frac{1}{a_i}\right),
\end{aligned}$$

and the desired result follows. ■

(b) Suppose

$$(a_1, \dots, a_n) = \sum_{S \in I(G)} \alpha_S \cdot \chi(S) \in VP(G),$$

minimizes $\sum_{i \in [n]} p_i \log_2 \left(\frac{1}{a_i}\right)$ (with each $\alpha_S \geq 0$ and $\sum_{S \in I(G)} \alpha_S = 1$). Define the distribution of (X, Y) by $X \sim P$ and

$$\mathsf{Pr}(Y = S \,|\, X = i) = \frac{\alpha_S}{a_i}$$

(it can be checked that this distribution is well-defined, and $\mathsf{Pr}(Y = S) = \sum_{i \in [n]} p_i \alpha_S / a_i$). Then,

$$\begin{aligned}
I(X;Y) &= \sum_{i \in [n]} p_i \sum_{i \in S \in I(G)} \mathsf{Pr}(Y = S \,|\, X = i) \log_2 \left(\frac{\mathsf{Pr}(Y = S \,|\, X = i)}{\mathsf{Pr}(Y = S)} \right) \\
&= - \sum_{i \in [n]} p_i \sum_{i \in S \in I(G)} \frac{\alpha_S}{a_i} \log_2 \left(\frac{\sum_{j \in [n]} p_j \alpha_S / a_j}{\alpha_S / a_i} \right) \\
&= - \sum_{i \in [n]} p_i \log_2 \left(a_i \sum_{j \in [n]} p_j / a_j \right) & \left(\sum_{S \in I(G)} \frac{\alpha_S}{a_i} = 1 \right) \\
&\leq - \sum_{i \in [n]} p_i \log_2 \left(a_i \sum_{j \in [n]} p_j \right) = \sum_{i \in [n]} p_i \log_2 \left(\frac{1}{a_i} \right),
\end{aligned}$$

where the inequality occurs because each $a_j \leq 1$. The desired result follows. ■

(c) Note that χ is monotone in the sense that, if $S \subseteq T \in I(G)$, then $(\chi(S))_i \leq (\chi(T))_i, \forall i \in [n]$. Consider any family $\{\alpha_S\}_{S \in I(G)}$ with each $\alpha_S \geq 0$ and $\sum_{S \in I(G)} \alpha_S = 1$, if $\alpha_S > 0$ for some $S \in I(G)$. Then, choosing $T \in I(G)$ maximal with $S \subseteq T$ (it is easy to construct such a T from S), and defining, for each $R \in I(G)$,

$$\beta_R = \begin{cases} \alpha_T + \alpha_S & : \text{ if } R = T \\ 0 & : \text{ if } R = S \\ \alpha_R & : \text{ else} \end{cases},$$

it is clear that each $\beta_R \geq 0$, $\sum_{R \in I(G)} \beta_R = 1$, and, moreover, by monotonicity of χ,

$$\left(\sum_{R \in I(G)} \alpha_R \cdot \chi(R) \right)_i \leq \left(\sum_{R \in I(G)} \beta_R \cdot \chi(R) \right)_i, \forall i \in [n],$$

so that all components are maximized by α_S supported on maximal independent sets. Therefore, since $\log_2$ is non-decreasing, the sum

$$\sum_{i \in [n]} p_i \log_2 \left(\frac{1}{a_i} \right)$$

is non-increasing in each a_i and minimized by α_S supported on maximal independent sets. ■

Problem 5

(a) Let $\varepsilon > 0$, and let Π be a protocol with error at most ε. Π partitions $X \times Y$ into rectangles; let $\mathcal{L}$ be the set of such rectangles R with $\mu(R) \geq \rho$, and let $\mathcal{S}$ be the remaining set of "small" rectangles. Let $E := \{(x, y) \in X \times Y : \Pi(x, y) \neq f(x, y)\}$. If $R \in \mathcal{L}$, then

$$\mu(E \cap R) \geq \min\{\mu(f^{-1}(0) \cap R), \mu(f^{-1}(1) \cap R)\} \geq \alpha\mu(f^{-1}(0) \cap R)),$$

since $\alpha \in (0, 1)$ and R is large. Thus, since the rectangles are disjoint,

$$\begin{aligned}\varepsilon \geq \mu(E) \geq \sum_{R \in \mathcal{L}} \mu(R \cap E) \geq \sum_{R \in \mathcal{L}} \alpha\mu\left(f^{-1}(0) \cap R\right) &= \alpha\mu\left(f^{-1}(0) \cap \bigcup_{R \in \mathcal{L}} R\right) \\ &= \alpha\mu\left(f^{-1}(0)\right) - \alpha\mu\left(f^{-1}(0) \cap \bigcup_{R \in \mathcal{S}} R\right),\end{aligned}$$

and so

$$\rho|S| \geq \sum_{R \in \mathcal{S}} \mu(R) = \mu\left(\bigcup_{R \in \mathcal{S}} R\right) \geq \mu\left(f^{-1}(0) \cap \bigcup_{R \in \mathcal{S}} R\right) \geq \mu\left(f^{-1}(0)\right) - \frac{\varepsilon}{\alpha}.$$

Recalling that the number of rectangles is at most $2^{R(f)}$, we have

$$\rho 2^{R(f)} \geq \rho|S| \geq \mu\left(f^{-1}(0)\right) - \frac{\varepsilon}{\alpha},$$

giving the desired result when $\varepsilon = 1/3$. ■

(b) Let μ be the uniform distribution over $\{0, 1\}^{2n}$. Note that, if $x \in \{0, 1\}^n$ is non-zero, then $IP(x, y) \neq 0$ for exactly half of the $y \in \{0, 1\}^n$. Therefore, if $\rho := \frac{8}{2^n}$ and R is a rectangle with $\mu(R) \geq \rho$, since R has elements corresponding to least 7 non-zero input x values,

$$\mu(f^{-1}(1) \cap R) \geq \frac{\frac{1}{2} \cdot 7 \cdot 2^n}{2^{2n}} \quad \text{and} \quad \mu(f^{-1}(0) \cap R) \leq \frac{2^n + \frac{1}{2} \cdot 7 \cdot 2^n}{2^{2n}},$$

and so

$$\frac{\mu(f^{-1}(1) \cap R)}{\mu(f^{-1}(0) \cap R)} \geq \frac{\frac{7}{2}}{1 + \frac{7}{2}} = 7/9 =: \alpha.$$

Therefore, noting that $f^{-1}(0) \geq \frac{1}{2}$, by the result of part (a),

$$2^{R(IP)} > \frac{1}{2^{3-n}}\left(\frac{1}{2} - \frac{1}{3 \cdot 7/9}\right) = 2^{n-3} \cdot \frac{1}{14} = 2^{n-3-\log_2 14}$$

and so $R(IP) \in n - O(1)$. ■

Problem 6

As in the proof for the original Indexing Problem, we distribute $X = (X_1, \ldots, X_n)$ and i uniformly on $\{0,1\}^n$ and $[n]$, respectively, and note

$$CC(\Pi) \geq I(X_1, \ldots, X_n \,|\, M = H(X_1, \ldots, X_n) - H(X_1, \ldots, X_n \,|\, M) = n - H(X_1, \ldots, X_n \,|\, M).$$

By the Chain Rule,

$$H(X_1, \ldots, X_n \,|\, M) = \sum_{i=1}^{n} H(X_i \,|\, M, X_1, \ldots, X_{i-1}).$$

By the error guarantee of the protocol, if $P_e^{m,i}$ is the probability Bob makes an error,

$$\mathop{\mathbb{E}}_{m,i}\left[P_e^{m,i}\right] \leq \frac{1}{3}.$$

By Fano's Inequality, since Bob's estimate of X_i is a function of M and $X_1, \ldots, X_{i-1}$, $h(P_e^{m,i}) \geq H(X_i \,|\, M = m, X_1, \ldots,)$. Therefore,

$$\begin{aligned}
\sum_{i=1}^{n} H(X_i \,|\, M, X_1, \ldots, X_{i-1}) &= n \mathop{\mathbb{E}}_{i}\left[H(X_i \,|\, M, X_1, \ldots, X_{i-1})\right] \\
&= n \mathop{\mathbb{E}}_{i}\left[\mathop{\mathbb{E}}_{m}\left[H(X_i \,|\, M = m, X_1, \ldots, X_i)\right]\right] \\
&\leq n \mathop{\mathbb{E}}_{m,i}\left[h(P_e^{m,i})\right].
\end{aligned}$$

Thus, by concavity of h,

$$\sum_{i=1}^{n} H(X_i \,|\, M, X_1, \ldots, X_{i-1}) \leq n\,\mathbb{E}[h(P_e^{m,i})] \leq nh(\mathbb{E}[P_e^{m,i}]) \leq nh\left(\frac{1}{3}\right),$$

and hence

$$CC(\Pi) \geq n - \sum_{i=1} H(X_i \,|\, M, X_1, \ldots, X_{i-1}) \geq n\left(1 - h\left(\frac{1}{3}\right)\right) \in \Omega(n). \quad \blacksquare$$