1: Problem 1 (Shearer for Mutual Information), 25 pts

Let $X = X_1, \ldots, X_n$ be (correlated) random variables, and let $S \subset_R [n]$ s.t $\Pr[i \in S] \geq \alpha \ \forall i \in [n]$. Recall that Shearer's lemma asserts that $\mathbb{E}_S[H(X_S)] \geq \alpha H(X)$. This exercise examines the possibility of extending Shearer's lemma to mutual information.

(a) With the above assumption, is it also true that for any random variable T, $\mathbb{E}_S[I(X_S;T)] \geq \alpha I(X;T)$? Prove or find a counter example.

(b) Show that if further: (i) all X_i's are independent of each other conditioned on T (i.e., $I(X_i; X_{<i}|T) = 0$), and (ii) $\Pr[i \in S] = \alpha \ \forall i \in [n]$, then the "Shearer inequality" from (a) in fact holds. (note that requiring equality and not just a lower bound of α is necessary).

2: Problem 2 (Properties of KL Divergence), 25 pts

Prove the following claims (left as exercises from class):

(a) (Chain Rule for KL) Prove that

$$\mathsf{D}\left(\frac{X_1, \ldots, X_n}{Y_1, \ldots, Y_n}\right) = \sum_i \mathbb{E}_{(v_{<i}) \sim X_{<i}}\left[\mathsf{D}\left(\frac{X_i | X_{<i} = v_{<i})}{Y_i | Y_{<i} = v_{<i}}\right)\right].$$

(b) (Convexity of KL) Let $\{\mu_i\}_{i=1}^n, \{\nu_i\}_{i=1}^n$ be distributions on the same universe, and define the two (convex combinations) distributions $\mu := \sum_i \alpha_i \mu_i$, $\nu = \sum_i \alpha_i \nu_i$. Show that

$$\sum_i \alpha_i \mathsf{D}\left(\frac{\mu_i}{\nu_i}\right) \geq \mathsf{D}\left(\frac{\mu}{\nu}\right).$$

(Hint: Use the so-called Log-Sum inequality: For any nonnegative real numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$, it holds that $\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq (\sum_{i=1}^n a_i) \log\left(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\right)$).

3: Problem 3 (Entropy vs. Expectation), 25 pts

Let X be an $\mathbb{N}$-valued random variable. Show that $H(X) \leq O(\mathbb{E}[\lg X])$.

(Hint: Define the random variable Y s.t $\Pr[Y = i] = 1/ci^2$ for any $i \in \mathbb{N}$ using the fact that the series $\sum_{i=1}^\infty 1/i^2$ converges (to $c = \pi^2/6$). Now use nonnegativity of KL divergence).

(5 pt Bonus) Can you think of a counterexample when $Supp(X) \not\subseteq \mathbb{N}$?

4: Problem 4 (Statistical distance vs. Mutual Information), 20 pts

Let (X, M) be jointly distributed r.vs. Prove that

$$\mathbb{E}_x\left[\|(M|x) - M\|_1\right] \leq \sqrt{(2\ln 2)I(X;M)},$$

where $(M|x)$ denotes the distribution of M conditioned on $X = x$. Conclude that one can approximately sample the message $M = M(X)$ even without knowing X, so long that X and M are not very correlated. (Hint: Pinsker's Inequality).

5: Problem 5 (Fooling Set vs. Rank lower bound), 30 pts

Let $f(x, y)$ be a two-party Boolean function. Recall that a *Fooling-Set* of f of size k is a set of input pairs $\{(x_i, y_i)\}_{i=1}^k$ such that $f(x_i, y_i) = 1$ but for any $i \neq j \in [k]$, either $f(x_i, y_j) = 0$ or $f(x_j, y_i) = 0$. Let $FS(M_f)$ denote the size of the largest fooling set of f.

(a) Let $GT_n(x, y) = 1 \Leftrightarrow \ x \geq y$. Use the Fooling-Set technique to show that $D(GT_n) = \Omega(n)$.

(b) Call a Fooling-Set S of f "strong" if it has the stronger property that $f(x_i, y_i) \equiv 1$ but for any $i \neq j \in [|S|]$, *both* $f(x_i, y_j) = 0$ *and* $f(x_j, y_i) = 0$. Denote by $\overline{FS(M_f)}$ the size of the largest strong fooling set of f. Prove that

$$rk(M_f) \geq \overline{FS(M_f)},$$

where $rk(M_f)$ is the rank of the communication matrix M_f (over the reals), recalling that $M_f(x, y) := f(x, y)$). (Hint: What can you say about the set of rows $r_{x_1}, \ldots, r_{x_{|S|}}$ in M_f?)

We note that this argument can be adapted to show that $rk(M_f) \geq FS(M_f)/2$ for any f.