Recitation: review of chapter 2 and chapter 4 concepts

Kanad

10) **Entropy of a disjoint mixture.** Let X_1 and X_2 be discrete random variables drawn according to probability mass functions $p_1(\cdot)$ and $p_2(\cdot)$ over the respective alphabets $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ and $\mathcal{X}_2 = \{m+1, \ldots, n\}$. Let

$$X = \begin{cases} X_1, & \text{with probability } \alpha, \\ X_2, & \text{with probability } 1-\alpha. \end{cases}$$

a) Find $H(X)$ in terms of $H(X_1)$ and $H(X_2)$ and α.
b) Maximize over α to show that $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ and interpret using the notion that $2^{H(X)}$ is the effective alphabet size.

10) *Entropy.* We can do this problem by writing down the definition of entropy and expanding the various terms. Instead, we will use the algebra of entropies for a simpler proof.
Since X_1 and X_2 have disjoint support sets, we can write

$$X = \begin{cases} X_1 & \text{with probability} \quad \alpha \\ X_2 & \text{with probability} \quad 1-\alpha \end{cases}$$

Define a function of X,

$$\theta = f(X) = \begin{cases} 1 & \text{when } X = X_1 \\ 2 & \text{when } X = X_2 \end{cases}$$

Then as in problem 1, we have

$$\begin{aligned} H(X) &= H(X, f(X)) = H(\theta) + H(X|\theta) \\ &= H(\theta) + p(\theta = 1)H(X|\theta = 1) + p(\theta = 2)H(X|\theta = 2) \\ &= H(\alpha) + \alpha H(X_1) + (1-\alpha)H(X_2) \end{aligned}$$

where $H(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$.

29) **Inequalities.** Let X, Y and Z be joint random variables. Prove the following inequalities and find conditions for equality.

a) $H(X,Y|Z) \geq H(X|Z)$.
b) $I(X,Y;Z) \geq I(X;Z)$.
c) $H(X,Y,Z) - H(X,Y) \leq H(X,Z) - H(X)$.
d) $I(X;Z|Y) \geq I(Z;Y|X) - I(Z;Y) + I(X;Z)$.

29) *Inequalities.*

a) Using the chain rule for conditional entropy,

$$H(X,Y \mid Z) = H(X \mid Z) + H(Y \mid X,Z) \geq H(X \mid Z),$$

with equality iff $H(Y \mid X,Z) = 0$, that is, when Y is a function of X and Z.

b) Using the chain rule for mutual information,

$$I(X,Y;Z) = I(X;Z) + I(Y;Z \mid X) \geq I(X;Z),$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

c) Using first the chain rule for entropy and then the definition of conditional mutual information,

$$\begin{aligned} H(X,Y,Z) - H(X,Y) &= H(Z \mid X,Y) = H(Z \mid X) - I(Y;Z \mid X) \\ &\leq H(Z \mid X) = H(X,Z) - H(X), \end{aligned}$$

with equality iff $I(Y;Z \mid X) = 0$, that is, when Y and Z are conditionally independent given X.

d) Using the chain rule for mutual information,

$$I(X;Z \mid Y) + I(Z;Y) = I(X,Y;Z) = I(Z;Y \mid X) + I(X;Z),$$

and therefore

$$I(X;Z \mid Y) = I(Z;Y \mid X) - I(Z;Y) + I(X;Z).$$

We see that this inequality is actually an equality in all cases.

36) **Symmetric relative entropy:** Though, as the previous example shows, $D(p||q) \neq D(q||p)$ in general, there could be distributions for which equality holds. Give an example of two distributions p and q on a binary alphabet such that $D(p||q) = D(q||p)$ (other than the trivial case $p = q$).

36) A simple case for $D((p, 1-p)||(q, 1-q)) = D((q, 1-q)||(p, 1-p))$, i.e., for

$$p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}$$

is when $q = 1-p$.

6) **Monotonicity of entropy per element.** For a stationary stochastic process $X_1, X_2, \ldots, X_n$, show that

a)

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. \tag{190}$$

b)

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \geq H(X_n | X_{n-1}, \ldots, X_1). \tag{191}$$

6) *Monotonicity of entropy per element.*

a) By the chain rule for entropy,

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} = \frac{\sum_{i=1}^{n} H(X_i|X^{i-1})}{n} \tag{246}$$

$$= \frac{H(X_n|X^{n-1}) + \sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n} \tag{247}$$

$$= \frac{H(X_n|X^{n-1}) + H(X_1, X_2, \ldots, X_{n-1})}{n}. \tag{248}$$

From stationarity it follows that for all $1 \leq i \leq n$,

$$H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),$$

which further implies, by averaging both sides, that,

$$H(X_n|X^{n-1}) \leq \frac{\sum_{i=1}^{n-1} H(X_i|X^{i-1})}{n-1} \tag{249}$$

$$= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. \tag{250}$$

Combining (248) and (250) yields,

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \frac{1}{n}\left[\frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1} + H(X_1, X_2, \ldots, X_{n-1})\right]$$

$$= \frac{H(X_1, X_2, \ldots, X_{n-1})}{n-1}. \tag{251}$$

b) By stationarity we have for all $1 \leq i \leq n$,

$$H(X_n|X^{n-1}) \leq H(X_i|X^{i-1}),$$

which implies that

$$\begin{aligned} H(X_n|X^{n-1}) &= \frac{\sum_{i=1}^{n} H(X_n|X^{n-1})}{n} && (252) \\ &\leq \frac{\sum_{i=1}^{n} H(X_i|X^{i-1})}{n} && (253) \\ &= \frac{H(X_1, X_2, \ldots, X_n)}{n}. && (254) \end{aligned}$$

11) **Stationary processes.** Let $\ldots, X_{-1}, X_0, X_1, \ldots$ be a stationary (not necessarily Markov) stochastic process. Which of the following statements are true? Prove or provide a counterexample.

a) $H(X_n|X_0) = H(X_{-n}|X_0)$.

b) $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.

c) $H(X_n|X_1, X_2, \ldots, X_{n-1}, X_{n+1})$ is nonincreasing in n.

d) $H(X_n|X_1, \ldots, X_{n-1}, X_{n+1}, \ldots, X_{2n})$ is non-increasing in n.

11) *Stationary processes.*

a) $H(X_n|X_0) = H(X_{-n}|X_0)$.

This statement is true, since

$$H(X_n|X_0) = H(X_n, X_0) - H(X_0) \quad (269)$$

$$H(X_{-n}|X_0) = H(X_{-n}, X_0) - H(X_0) \quad (270)$$

and $H(X_n, X_0) = H(X_{-n}, X_0)$ by stationarity.

b) $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.

This statement is not true in general, though it is true for first order Markov chains. A simple counterexample is a periodic process with period n. Let $X_0, X_1, X_2, \ldots, X_{n-1}$ be i.i.d. uniformly distributed binary random variables and let $X_k = X_{k-n}$ for $k \geq n$. In this case, $H(X_n|X_0) = 0$ and $H(X_{n-1}|X_0) = 1$, contradicting the statement $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.

c) $H(X_n|X_1^{n-1}, X_{n+1})$ is non-increasing in n.

This statement is true, since by stationarity $H(X_n|X_1^{n-1}, X_{n+1}) = H(X_{n+1}|X_2^n, X_{n+2}) \geq H(X_{n+1}|X_1^n, X_{n+2})$ where the inequality follows from the fact that conditioning reduces entropy.