Problem Solving Session - 2

1.

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) $I(X;Z\,|\,Y) \geq I(Z;Y\,|\,X) - I(Z;Y) + I(X;Z)$.

Inequalities.

a) Using the chain rule for conditional entropy,

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

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

b) Using the chain rule for mutual information,

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

with equality iff $I(Y;Z\,|\,X) = 0$, 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,

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

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

d) Using the chain rule for mutual information,

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

and therefore

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

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

2.

Entropy of a sum. Let X and Y be random variables that take on values $x_1, x_2, \ldots, x_r$ and $y_1, y_2, \ldots, y_s$, respectively. Let $Z = X + Y$.

a) Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent, then $H(Y) \leq H(Z)$ and $H(X) \leq H(Z)$. Thus the addition of *independent* random variables adds uncertainty.
b) Give an example of (necessarily dependent) random variables in which $H(X) > H(Z)$ and $H(Y) > H(Z)$.
c) Under what conditions does $H(Z) = H(X) + H(Y)$?

Entropy of a sum.

a) $Z = X + Y$. Hence $p(Z = z|X = x) = p(Y = z - x|X = x)$.

$$\begin{aligned} H(Z|X) &= \sum p(x) H(Z|X = x) \\ &= -\sum_x p(x) \sum_z p(Z = z|X = x) \log p(Z = z|X = x) \\ &= \sum_x p(x) \sum_y p(Y = z - x|X = x) \log p(Y = z - x|X = x) \\ &= \sum p(x) H(Y|X = x) \\ &= H(Y|X). \end{aligned}$$

If X and Y are independent, then $H(Y|X) = H(Y)$. Since $I(X; Z) \geq 0$, we have $H(Z) \geq H(Z|X) = H(Y|X) = H(Y)$. Similarly we can show that $H(Z) \geq H(X)$.

b) Consider the following joint distribution for X and Y Let

$$X = -Y = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases}$$

Then $H(X) = H(Y) = 1$, but $Z = 0$ with prob. 1 and hence $H(Z) = 0$.

c) We have

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

because Z is a function of (X, Y) and $H(X, Y) = H(X) + H(Y|X) \leq H(X) + H(Y)$. We have equality iff (X, Y) is a function of Z and $H(Y) = H(Y|X)$, i.e., X and Y are independent.

3.

AEP

Let X_i be iid $\sim p(x)$, $x \in \{1, 2, \ldots, m\}$. Let $\mu = EX$, and $H = -\sum p(x) \log p(x)$. Let $A^n = \{x^n \in \mathcal{X}^n : |-\frac{1}{n} \log p(x^n) - H| \leq \epsilon\}$. Let $B^n = \{x^n \in \mathcal{X}^n : |\frac{1}{n} \sum_{i=1}^n X_i - \mu| \leq \epsilon\}$.

a) Does $\Pr\{X^n \in A^n\} \longrightarrow 1$?

b) Does $\Pr\{X^n \in A^n \cap B^n\} \longrightarrow 1$?

c) Show $|A^n \cap B^n| \leq 2^{n(H+\epsilon)}$, for all n.

d) Show $|A^n \cap B^n| \geq (\frac{1}{2}) 2^{n(H-\epsilon)}$, for n sufficiently large.

a) Yes, by the AEP for discrete random variables the probability X^n is typical goes to 1.

b) Yes, by the Strong Law of Large Numbers $Pr(X^n \in B^n) \to 1$. So there exists $\epsilon > 0$ and N_1 such that $Pr(X^n \in A^n) > 1 - \frac{\epsilon}{2}$ for all $n > N_1$, and there exists N_2 such that $Pr(X^n \in B^n) > 1 - \frac{\epsilon}{2}$ for all $n > N_2$. So for all $n > \max(N_1, N_2)$:

$$\begin{aligned} Pr(X^n \in A^n \cap B^n) &= Pr(X^n \in A^n) + Pr(X^n \in B^n) - Pr(X^n \in A^n \cup B^n) \\ &> 1 - \frac{\epsilon}{2} + 1 - \frac{\epsilon}{2} - 1 \\ &= 1 - \epsilon \end{aligned}$$

So for any $\epsilon > 0$ there exists $N = \max(N_1, N_2)$ such that $Pr(X^n \in A^n \cap B^n) > 1 - \epsilon$ for all $n > N$, therefore $Pr(X^n \in A^n \cap B^n) \to 1$.

c) By the law of total probability $\sum_{x^n \in A^n \cap B^n} p(x^n) \leq 1$. Also, for $x^n \in A^n$, from Theorem 3.1.2 in the text, $p(x^n) \geq 2^{-n(H+\epsilon)}$. Combining these two equations gives $1 \geq \sum_{x^n \in A^n \cap B^n} p(x^n) \geq \sum_{x^n \in A^n \cap B^n} 2^{-n(H+\epsilon)} = |A^n \cap B^n| 2^{-n(H+\epsilon)}$. Multiplying through by $2^{n(H+\epsilon)}$ gives the result $|A^n \cap B^n| \leq 2^{n(H+\epsilon)}$.

d) Since from (b) $Pr\{X^n \in A^n \cap B^n\} \to 1$, there exists N such that $Pr\{X^n \in A^n \cap B^n\} \geq \frac{1}{2}$ for all $n > N$. From Theorem 3.1.2 in the text, for $x^n \in A^n$, $p(x^n) \leq 2^{-n(H-\epsilon)}$. So combining these two gives $\frac{1}{2} \leq \sum_{x^n \in A^n \cap B^n} p(x^n) \leq \sum_{x^n \in A^n \cap B^n} 2^{-n(H-\epsilon)} = |A^n \cap B^n| 2^{-n(H-\epsilon)}$. Multiplying through by $2^{n(H-\epsilon)}$ gives the result $|A^n \cap B^n| \geq (\frac{1}{2}) 2^{n(H-\epsilon)}$ for n sufficiently large.