COMP2610/COMP6261 – Information Theory

Tutorial 5: Fundamental and Probabilistic inequalities (soln.)

(Week 5, Semester 2, 2023)

1. Let X, Y and Z be joint random variables. Prove the following inequalities and find conditions for equality.

   a) $H(X, Y | Z) \geq H(X|Z)$.

   Soln. By using Chain rule for Conditional Entropy, $\boldsymbol{H(X, Y | Z) = H(X|Z) + H(Y | X, Z) \geq H(X|Z)}$ (1)

   In equation (1), equality can be achieved if $\boldsymbol{H(Y | X, Z) = 0}$ ***i***.e., when Y will be the function of X and Z.

   b) $I(X, Y; Z) \geq I(X; Z)$.

   Soln. By using Chain rule for Mutual Information, $\boldsymbol{I(X, Y; Z) = I(X; Z) + I(Y; Z/X) \geq I(X; Z)}$ (2)

   In equation (2), equality can be achieved if $\boldsymbol{I(Y; Z/X) = 0}$ ***i***.e., when Y and Z are conditionally independent given X.

   c) $H(X, Y, Z) - H(X, Y) \leq H(X, Z) - H(X)$.

   Soln. Using first the chain rule for entropy and then the definition of conditional mutual information,

$$
\begin{aligned}
H(X, Y, Z) - H(X, Y) &= H(Z|X, Y) = H(Z|X) - I(Y; Z|X) \\
&\leq H(Z|X) = H(X, Z) - H(X) \qquad \textbf{(3)}
\end{aligned}
$$

   In equation (3), equality can be achieved if $\boldsymbol{I(Y; Z/X) = 0}$ ***i***.e., when Y and Z are conditionally independent given X.

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

   Soln. By using Chain rule for mutual information,

$$\boldsymbol{I(X; Z|Y) + I(Z; Y) = I(X, Y; Z) = I(Z; Y | X) + I(X; Z)} \qquad \textbf{(4)}$$

   Therefore, $\boldsymbol{I(X; Z|Y) = I(Z; Y | X) - I(Z; Y) + I(X; Z)}$ **(5)**

   In equation (5), there will always be an equality in all cases.

2. Which of the following inequalities are generally ≥, =, ≤? Label each with ≥, =, or ≤.

   a) $H(5X)$ vs. $H(X)$

   Soln. $X \rightarrow 5X$ is a one to one mapping, and hence $H(X) = H(5X)$.

   b) $I(g(X); Y)$ vs. $I(X; Y)$

   Soln. By data processing inequality, $I(g(X); Y) \leq I(X; Y)$.

c) H $(X_0|X_{-1})$ vs. H $(X_0|X_{-1}, X_1)$

Soln. Because conditioning reduces entropy, ***$H(X_0|X_{-1}) \geq H(X_0|X_{-1}, X_1)$.***

d) H (X, Y) / (H(X) + H (Y)) vs. 1

Soln. Since, ***H (X, Y) ≤ H(X) + H (Y).*** Hence, ***H (X, Y)/(H(X) + H (Y)) ≤ 1***.

3. (Markov's inequality for probabilities) Let p(x) be a probability mass function.

   Prove, for all d ≥ 0, Pr {p(X) ≤ d} log $\left(\frac{1}{d}\right)$ ≤ H(X)**.**

   Soln. By applying Markov's inequality to Entropy,

$$\begin{aligned} P(p(X) < d) \log \frac{1}{d} &= \sum_{x:p(x)<d} p(x) \log \frac{1}{d} \\ &\leq \sum_{x:p(x)<d} p(x) \log \frac{1}{p(x)} \\ &\leq \sum_{x} p(x) \log \frac{1}{p(x)} \\ &= H(X) \end{aligned}$$

4. (Asymptotic Equipartition Property) A discrete memoryless source emits a sequence of statistically independent binary digits with probabilities p (1) = 0.005 and p (0) = 0.995. The digits are taken 100 at a time and a binary code word is provided for every sequence of 100 digits containing three or fewer ones.

   (a) Assuming that all code words are the same length, find the minimum length required to provide code words for all sequences with three or fewer ones.

   Soln. The number of 100-bit binary sequences with three or fewer ones is,

   $^{100}C_0 + {}^{100}C_1 + {}^{100}C_2 + {}^{100}C_3 = 1 + 100 + 4950 + 161700 = \mathbf{166751}$
   => Minimum length required = $\log_2$ (166751) = 17.3473

   (b) Calculate the probability of observing a source sequence for which no code word has been assigned.

   Soln. The probability that a 100-bit sequence has three or fewer ones is,

$$\sum_{i=0}^{3} \binom{100}{i} * (p(1))^{i} * (p(0))^{100-i} = \sum_{i=0}^{3} \binom{100}{i} * (0.005)^{i} * (0.995)^{100-i}$$
$$= 0.60577 + 0.30441 + 0.07572 + 0.01243$$
$$= 0.99833$$

   Thus the probability that the sequence that is generated cannot be encoded is 1 – 0.99833 = **0.00167**.

(c) Use Chebyshev's inequality to bound the probability of observing a source sequence for which no codeword has been assigned. Compare this bound with the actual probability computed in part (b).

Soln. For a random variable S_n, the **Chebyshev's inequality** states that the sum of n i.i.d. random variables X1, X2, . . ., Xn is given by, $\mathbf{Pr}(|\mathbf{Sn} - \mathbf{n}\mu| \geq \epsilon) \leq \frac{\mathbf{n}\sigma^2}{\epsilon^2}$ **(6)**

where μ and σ^2 are the mean and variance of X_i

Therefore $\mathbf{n}\mu$ and $\mathbf{n}\sigma^2$ will be the mean and variance of S_n.

⇨ n = 100, $\mu = 0.005$, and $\sigma^2 = (0.005) \times (0.995)$ **(7)**

By Chebyshev's inequality, $S_{100} \geq 4$ if and only if $|S_{100} - 100(0.005)| \geq 3.5$, **(8)**

$$\Pr(S_{100} \geq 4) \leq \frac{100(0.005)(0.995)}{(3.5)^2} \approx 0.04061\,.$$

Thus this bound is much larger than the actual probability **0.00167.**