COMP2610/COMP6261 – Information Theory

Tutorial 6

Zhifeng Tang (zhifeng.tang@anu.edu.au)

Question 1. Entropy

Let $p = (p_1, p_2, \cdots, p_m)$ be a probability distribution on m elements, i.e, $p_i \geq 0$, and $\sum_{i=1}^{m} p_i = 1$. Define a new distribution q on $m-1$ elements as $q_1 = p_1$, $q_2 = p_2$, $\cdots$, $q_{m-2} = p_{m-2}$, and $q_{m-1} = p_{m-1} + p_m$, i.e., the distribution q is the same as p on any $i \in \{1, 2, \cdots, m-2\}$, and the probability of the last element in q is the sum of the last two probabilities of p. Show that

$$H(p) = H(q) + (p_{m-1} + p_m) H\left(\frac{p_{m-1}}{p_{m-1} + p_m}, \frac{p_m}{p_{m-1} + p_m}\right).$$

Solution:

$$\begin{aligned}
H(p) &= -\sum_{i=1}^{m} p_i \log p_i \\
&= -\sum_{i=1}^{m-2} p_i \log p_i - p_{m-1} \log p_{m-1} - p_m \log p_m \\
&= -\sum_{i=1}^{m-2} p_i \log p_i - p_{m-1} \log \frac{p_{m-1}}{p_{m-1} + p_m} - p_m \log \frac{p_m}{p_{m-1} + p_m} - (p_{m-1} + p_m) \log(p_{m-1} + p_m) \\
&= H(q) - p_{m-1} \log \frac{p_{m-1}}{p_{m-1} + p_m} - p_m \log \frac{p_m}{p_{m-1} + p_m} \\
&= H(q) - (p_{m-1} + p_m) \left(\frac{p_{m-1}}{p_{m-1} + p_m} \log \frac{p_{m-1}}{p_{m-1} + p_m} + \frac{p_m}{p_{m-1} + p_m} \log \frac{p_m}{p_{m-1} + p_m} \right) \\
&= H(q) + (p_{m-1} + p_m) H\left(\frac{p_{m-1}}{p_{m-1} + p_m}, \frac{p_m}{p_{m-1} + p_m}\right).
\end{aligned}$$

Question 2. Mutual Information and Relative Entropy

Let X, Y, Z be three random variables with a joint probability mass function $p(X,Y,Z)$.
(a). Show that

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

(b). The relative entropy between the joint distribution and the product of the marginals is $D(p(x,y,z)||p(x)p(y)p(z))$. Show that

$$D(p(x,y,z)||p(x)p(y)p(z)) = I(X;Y) + I(X,Y;Z).$$

Solution:
(a). Note that

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

We obtain

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

Hence, it can be proved that

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

(b).

$$\begin{aligned}
&D(p(x,y,z)||p(x)p(y)p(z)) = \sum_{x,y,z} p(x,y,z)\log\frac{p(x,y,z)}{p(x)p(y)p(z)}\\
&= \sum_{x,y,z} p(x,y,z)\log p(x,y,z) - \sum_{x,y,z} p(x,y,z)\log p(x) - \sum_{x,y,z} p(x,y,z)\log p(y) - \sum_{x,y,z} p(x,y,z)\log p(z)\\
&= \sum_{x,y,z} p(x,y,z)\log p(x,y,z) - \sum_{x} p(x)\log p(x) - \sum_{y} p(y)\log p(y) - \sum_{z} p(z)\log p(z)\\
&= -H(X,Y,Z) + H(X) + H(Y) + H(Z)\\
&= I(X,Y;Z) + I(X;Y).
\end{aligned}$$

Question 3. Markov Chain

Suppose a Markov chain $X_1 \to X_2 \to X_3$, starts in one of n states, i.e., $X_1 \in \{1,2,\cdots,n\}$. Suppose X_2 will go down to $k < n$ states, i.e., $X_2 \in \{1,2,\cdots,k\}$. Then X_3 go back to $m > k$ states, i.e., $X_3 \in \{1,2,\cdots,m\}$.
(a). What is the upper bound of $I(X_1;X_3)$?
(b). Evaluate $I(X1;X3)$ for $k = 1$, and conclude that no dependence can survive.

Solution:
(a). From the data processing inequality, and the fact that entropy is maximum for a uniform distribution, we get

$$\begin{aligned} I(X1;X3) &\leq I(X1;X2) \\ &= H(X2) - H(X2|X1) \leq H(X2) \\ &\leq \log|\mathcal{X}_2| = \log k, \quad (1) \\ I(X1;X3) &\leq H(X_1) \leq \log|\mathcal{X}_1| = \log n, \quad (2) \\ I(X1;X3) &\leq H(X_3) \leq \log|\mathcal{X}_3| = \log m. \quad (3) \end{aligned}$$

Hence, $I(X1;X3) \leq \log k$.

(b). For $k = 1$, $I(X_1;X_3) \leq \log 1 = 0$. Since $I(X_1;X_3) \geq 0$, we obtain $I(X_1;X_3) = 0$. Thus, for $k = 1$, X_1 and X_3 are independent.

Question 4. Inequalities

A coin is known to land heads with probability $\frac{1}{5}$. The coin is flipped N times for some even integer N.
(a). Using Markov's inequality, provide a bound on the probability of observing $\frac{N}{2}$ or more heads.
(b). Using Chebyshev's inequality, provide a bound on the probability of observing $\frac{N}{2}$ or more heads. Express your answer in terms of N.

Solution:
(a). Let's denote the flip result of a coin as X with head as 1 and tail as 0, i.e., $P(X=1)=\frac{1}{5}$ and $P(X=0)=\frac{4}{5}$. Then we obtain

$$\mathbb{E}[X] = P(X=1)\times 1 + P(X=0)\times 0 = \frac{1}{5}.$$

Based on Markov inequality, we obtain

$$p\left(X^N \geq \frac{N}{2}\right) \leq \frac{\mathbb{E}[X^N]}{N/2} = \frac{2}{5}.$$

(b). Based on Chebyshev's inequality, we obtain

$$p\left(\left|X^N - \mathbb{E}[X^N]\right| \geq \frac{3N}{10}\right) \leq \frac{V[X^N]}{(3N/10)^2} = \frac{NV[X]}{(3N/10)^2} = \frac{16}{9N}.$$

Note that $p\left(\left|X^N - \mathbb{E}[X^N]\right| \geq \frac{3N}{10}\right) = p\left(X^N - \mathbb{E}[X^N] \geq \frac{3N}{10}\right) + p\left(\mathbb{E}[X^N] - X^N \geq \frac{3N}{10}\right) = p\left(X^N - \mathbb{E}[X^N] \geq \frac{3N}{10}\right)$.
Hence, we obtain

$$p\left(X^N \geq \frac{N}{2}\right) \leq \frac{16}{9N}$$