CIE 6020 Assignment 2

WU, Chenhao 117010285

February 18, 2019

1. Let X, Y, Z be three random variables with a joint probability mass function $p(x, y, z)$. The relative entropy between the joint distribution and the product of the marginal is

$$D(p(x,y,z)||p(x)p(y)p(z)) = E[\log \frac{p(x,y,z)}{p(x)p(y)p(z)}]$$

Expand this in terms of entropies. When is this quantity zero?

Answer

$$\begin{aligned} E[\log \frac{p(x,y,z)}{p(x)p(y)p(z)}] &= E[\log p(x,y,z) - \log p(x) - \log p(y) - \log p(z)] \\ &= E[\log p(x,y,z)] - E[\log p(x)] - E[\log p(y)] - E[\log p(z)] \\ &= -H(X,Y,Z) + H(X) + H(Y) + H(Z) \end{aligned}$$

in which the quantity is zero iff X, Y and Z are mutually independent
i.e. $p(x,y,z) = p(x)p(y)p(z)$.

2. Let the random variable X have three possible outcomes a, b, c. Consider two distributions on this random variable:

symbol	$p(x)$	$p(y)$
a	$\frac{1}{2}$	$\frac{1}{3}$
b	$\frac{1}{4}$	$\frac{1}{3}$
c	$\frac{1}{4}$	$\frac{1}{3}$

Calculate $H(p)$, $H(q)$, $D(p||q)$ and $D(q||p)$. Verify that in this case, $D(p||q) \neq D(q||p)$

Answer

$$H(p) = -(\frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4} + \frac{1}{4}\log\frac{1}{4}) = \frac{3}{2}$$
$$H(q) = -(\frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3} + \frac{1}{3}\log\frac{1}{3}) = \log 3$$
$$D(p||q) = \frac{1}{2}\log\frac{\frac{1}{2}}{\frac{1}{3}} + \frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{3}} + \frac{1}{4}\log\frac{\frac{1}{4}}{\frac{1}{3}} = \log 3 - \frac{3}{2}$$
$$D(q||p) = \frac{1}{3}\log\frac{\frac{1}{3}}{\frac{1}{2}} + \frac{1}{3}\log\frac{\frac{1}{3}}{\frac{1}{4}} + \frac{1}{3}\log\frac{\frac{1}{3}}{\frac{1}{4}} = -\log 3 + \frac{5}{3}$$

3. Show that $lnx \geq 1 - \frac{1}{x}$ for $x > 0$, where the equality holds when $x = 1$.

Proof

Let $f(x) = \ln x + \frac{1}{x} - 1$, and $f(1) = 0$.

The derivative of $f(x)$ is $\frac{d}{dx}f(x) = \frac{1}{x} - \frac{1}{x^2}$.

(1). For $0 < x < 1$, $\frac{d}{dx}f(x) = \frac{x-1}{x^2} < 0$. Therefore $f(x)$ is monotonically decreasing over $(0, 1)$

(2). For $x > 1$, $\frac{d}{dx}f(x) = \frac{x-1}{x^2} > 0$. Therefore $f(x)$ is monotonically increasing over $(1, +\infty)$

Hence, for $x > 0$, $lnx \geq 1 - \frac{1}{x}$.

4. *Conditioning reduces entropy.* Show that $H(Y|X) \leq H(Y)$ with equality iff X and Y are independent.

Proof

Claim:$D(p||q) \geq 0$, with equality iff $p(x) = q(x)$ for all x.

$$\begin{aligned}
D(p||q) &= \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{p(y)} \\
&= -\sum_{x\in\mathcal{X}} p(x)\log\frac{p(y)}{p(x)} \\
&\geq -\log\sum_{x\in\mathcal{X}} p(x)\frac{q(x)}{p(x)} \\
&= -\log\sum_{x\in\mathcal{X}} q(x) \\
&= -\log 1 \\
&= 0
\end{aligned}$$

Also, $H(Y) - H(Y|X) = I(X;Y) = D(p(x,y)||p(x)p(y)) \geq 0$ and equality holds iff $p(x,y) = p(x)p(y)$, in which $p(x)$ and $p(y)$ are independent.

5. Show that $I(X;Y|Z) \geq 0$ with equality iff $X \to Z \to Y$.

Proof

Shown in Question 4 that

$$I(X;Y) = D(p(x,y)||p(x)p(y)) \geq 0$$

we can expand the conclusion to conditional mutual information given Z since that

$$I(X;Y|Z) = \boldsymbol{E}_{p(x,y,z)}\log\frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)} \geq 0$$

and also equality holds iff X and Y are independent conditioning to Z, which is also a necessary condition of Markov chain $X \to Z \to Y$.

6. *Data processing.* Let $X_1 \to X_2 \to X_3 \to ... \to X_n$ form a Markov chain, i.e.,

$$p(x_1, x_2, ..., x_n) = p(x_1)p(x_2|x_1)...p(x_n|x_{n-1})$$

Reduce $I(X_1; X_2, ..., X_n)$ to its simplest form.

Answer

From the chain rule for mutual information we have

$$\begin{aligned} I(X_1; X_2, ..., X_n) &= I(X_1; X_2) + I(X_1; X_3|X_2) + ... + I(X_1; X_n|X_2, ..., X_{n-2}) \\ &= I(X_1; X_2) \end{aligned}$$

7. Let X and Y be two random variables and let Z be independent of (X, Y). Show that $I(X; Y) \geq I(X; g(Y, Z))$ for any function g.

Proof

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && \textbf{by definition} \\ &= H(X) - H(X|(Y,Z)) && \textbf{independence between X and Z} \\ &\geq H(X) - H(X|g(Y,Z)) && H(X|(Y,Z)) \leq H(X|g(Y,Z)) \\ & && \textbf{with equality iff } (Y,Z) \textbf{ is a function of } g(Y,Z) \\ &= I(X; g(Y,Z)) \end{aligned}$$

8. *Bottleneck.* Suppose that a (non-stationary) Markov chain starts in one of n states, necks down to $k < n$ states, and then fans back to $m > k$ states. Thus $X_1 \to X_2 \to X_3$, that is, $p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$, for all $x_1 \in \{1, 2, ..., n\}$, $x_2 \in \{1, 2, ..., k\}$,

$x_3 \in \{1, 2, ..., m\}$.

(a) Show that the dependence of X_1 and X_3 is limited by the bottleneck by proving that $I(X_1; X_3) \leq \log k$

(b) Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.

Proof

(a) By the data processing inequality

$$\begin{aligned}
I(X_1; X_3) &\leq I(X_1; X_2) \\
&= H(X_2) - H(X_2|X_1) \\
&\leq H(X_2) \\
&\leq \log k
\end{aligned}$$

(b) If $k = 1$, then

$$\begin{aligned}
I(X_1; X_3) &\leq \log 1 \\
&= 0
\end{aligned}$$

in which $I(X_1; X_3) \geq 0$, thus $I(X_1; X_3) = 0 \rightarrow X_1$ and X_3 are independent.