Question: How to prove $I(X;Y) \geq 0$

Proof:

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

According to Theorem 2.6.3 (Information inequality),

$$\begin{aligned} I(X;Y) &= D\big(p(x,y) \| p(x)p(y)\big) \\ &\geq 0 \end{aligned}$$

when $p(x,y) = p(x)p(y)$, $I(X;Y) = 0$.

So if we can prove $D(p\|q) \geq 0$, we can prove $I(X;Y) \geq 0$.

Proof of $D(p\|q) \geq 0$:

$$D(p\|q) = \sum_{x \in \chi} p(x) \log \frac{p(x)}{q(x)}$$

Since $\ln x \leq x - 1$ when x>0 (can be observed in Figure 1, also could be proved by monotonicity principle), we have $\log\left(\frac{q(x)}{p(x)}\right) \leq \frac{q(x)}{p(x)} - 1$

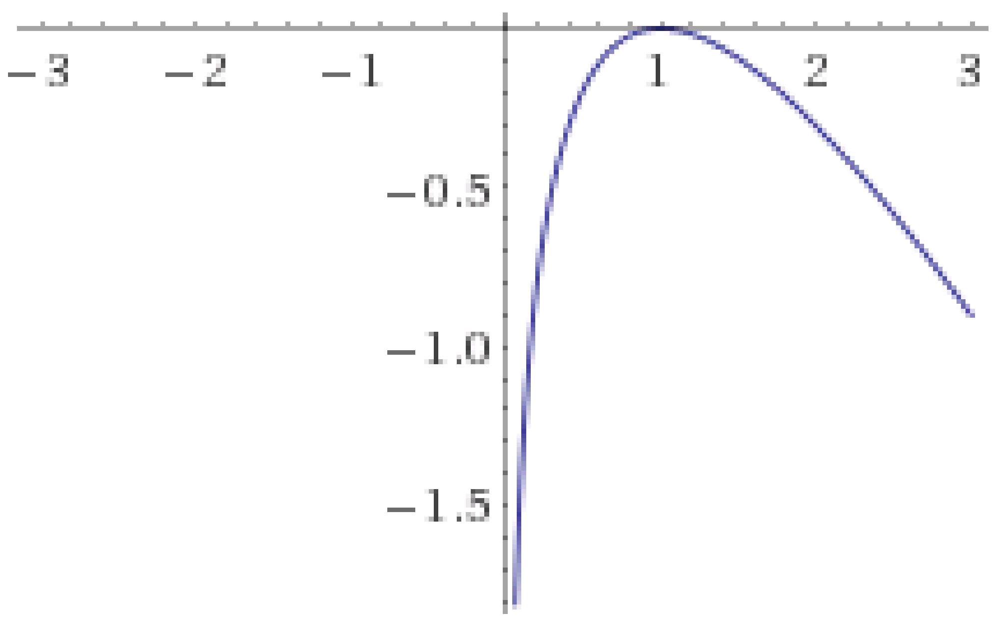

Figure 1 the plot of lnx-(x-1)

$$
\begin{aligned}
-D(p\|q) &= \sum_{x\in\chi} p(x)\log\frac{q(x)}{p(x)} \\
&\le \sum_{x\in\chi} p(x)\left(\frac{q(x)}{p(x)}-1\right) \\
&= \sum_{x\in\chi} p(x)\left(\frac{q(x)}{p(x)}-1\right) \\
&= \sum_{x\in\chi} p(x)\left(\frac{q(x)}{p(x)}-1\right) \\
&= \sum_{x\in\chi} q(x)-\sum_{x\in\chi} p(x) \\
&= 0
\end{aligned}
$$

Therefore,

$$D(p\|q)\ge 0$$

With equality if p(x)=q(x)

The proof of $D(p\|q)\ge 0$ in the book (P28, Elements of Information Theory) is as follows:

Theorem 2.6.3 (*Information inequality*) *Let* $p(x), q(x), x \in \mathcal{X}$*, be two probability mass functions. Then*

$$D(p||q) \geq 0 \tag{2.82}$$

with equality if and only if $p(x) = q(x)$ *for all* x.

Proof: Let $A = \{x : p(x) > 0\}$ be the support set of $p(x)$. Then

$$-D(p||q) = -\sum_{x \in A} p(x) \log \frac{p(x)}{q(x)} \tag{2.83}$$

$$= \sum_{x \in A} p(x) \log \frac{q(x)}{p(x)} \tag{2.84}$$

$$\leq \log \sum_{x \in A} p(x) \frac{q(x)}{p(x)} \tag{2.85}$$

$$= \log \sum_{x \in A} q(x) \tag{2.86}$$

$$\leq \log \sum_{x \in \mathcal{X}} q(x) \tag{2.87}$$

$$= \log 1 \tag{2.88}$$

$$= 0, \tag{2.89}$$

We can also notice that in the book (P29, Elements of Information Theory), it utilize $I(X;Y) \geq 0$ to prove $H(X|Y) \leq H(X)$