Positivity of the KL Divergence[1]

26 February 2008

I misunderstood how the proof of the positivity of the KL divergence (Theorem 2.6.3 in C&L ed3) works, so I am doing it again here.

Theorem

Given two probability distributions p and q on a set $\mathcal{X}$ then

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p(x) = q(x)$ for all $x \in \mathcal{X}$.

Proof: By definition

$$D(p\|q) = \sum_{x\in\mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \tag{2}$$

if $p(x) = 0$ x does not contribute to the sum, so

$$D(p\|q) = \sum_{x\in\mathcal{A}} p(x) \log \frac{p(x)}{q(x)} = -\sum_{x\in\mathcal{A}} p(x) \log \frac{q(x)}{p(x)} \tag{3}$$

Now let Y be the random variable $q(X)/p(X)$ where X is distributed according to $p(x)$, so there is probability $p(x)$ $X = x$, in which case $Y = p(x)/q(x)$. Of course more than one x might yield the same $p(x)/q(x)$, we don't know the map is $1 - 1$ and

$$p_Y(y) = \sum_{x:q(x)/p(x)=y} p(x) \tag{4}$$

This is the maybe obvious point I missed, we are interested in the log of a new random variable. Now

$$D(p\|q) = E(-\log Y) \tag{5}$$

and we can apply the Jensen inequality, so

$$D(p\|q) = E(-\log Y) \geq -\log EY = -\log \sum_{x\in\mathcal{A}} p(x) \frac{p(x)}{q(x)} \tag{6}$$

[1]Conor Houghton, `houghton@maths.tcd.ie` please send me any corrections.

with equality if and only if $Y = EY$ with probability one, so $p(x) = Cq(x)$ for all x, C must be one since p and q are both probability distributions. Finally, since log is increasing and $q(x)$ is positive for all $x \in \mathcal{X}$

$$D(p\|q) \geq -\log \sum_{x\in\mathcal{A}} p(x)\frac{p(x)}{q(x)} \geq -\log \sum_{x\in\mathcal{X}} q(x) = -\log 1 = 0 \tag{7}$$

and this proves the theorem.