Nonnegativity of Relative Entropy

http://www.isi.ee.ethz.ch/teaching/courses/it1/ Handout of September 28, 2016

In this handout we prove the following theorem in a way slightly different from that in the book.

Theorem 1. *Let $p(\cdot)$ and $q(\cdot)$ be two probability mass functions of a chance variable X taking value in a finite set $\mathcal{X}$. The relative entropy $D(p\|q)$ satisfies*

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

with equality if, and only if,

$$p(x) = q(x), \qquad \forall\, x \in \mathcal{X}. \tag{2}$$

Proof. Note that the validity of (1) does not depend on whether $D(p\|q)$ is measured in bits or nats or other units, therefore we only need to prove (1) in one unit, and results for other units follow immediately. Throughout this proof, we shall measure information in nats, namely, we shall assume all logarithms to be natural. For natural logarithms we have

$$\log \alpha \leq \alpha - 1, \qquad \forall\, \alpha \geq 0, \tag{3}$$

with equality if, and only if, $\alpha = 1$. Using this, we prove (1) by showing that $-D(p\|q) \leq 0$:

$$\begin{aligned}
-D(p\|q) &= \sum_{\substack{x\in\mathcal{X}:\\ p(x)>0}} p(x) \log \frac{q(x)}{p(x)} \\
&\stackrel{(a)}{\leq} \sum_{\substack{x\in\mathcal{X}:\\ p(x)>0}} p(x) \left(\frac{q(x)}{p(x)} - 1 \right) \\
&= \sum_{\substack{x\in\mathcal{X}:\\ p(x)>0}} q(x) - 1 \\
&\stackrel{(b)}{\leq} 1 - 1 \\
&= 0.
\end{aligned}$$

Note that for probability mass functions $p(\cdot)$ and $q(\cdot)$ satisfying (2), inequalities (a) and (b) are both satisfied with equality and hence $D(p\|q) = 0$. This shows that Condition (2) is sufficient for achieving equality in (1). It thus remains to show that (2) is also necessary for equality in (1). In order to have $D(p\|q) = 0$, we must have equality in (a) and in (b). By the necessary condition for equality in (3) we conclude that (a) holds with equality only if

$$p(x) = q(x), \qquad \forall\, x \in \mathcal{X} \colon p(x) > 0. \tag{4}$$

Note that since $p(\cdot)$ is a probability mass function it must sum to one and hence (4) implies equality in (b), because (4) implies

$$\sum_{\substack{x \in \mathcal{X}: \\ p(x)>0}} q(x) = \sum_{\substack{x \in \mathcal{X}: \\ p(x)>0}} p(x) = \sum_{x \in \mathcal{X}} p(x) = 1. \tag{5}$$

where the second equality follows because adding zero-terms to a sum does not change the sum, and where the last equality follows because $p(\cdot)$ is a probability mass function.

Since also $q(\cdot)$ is a probability mass function it must sum to 1 and thus (4)—via (5)—also implies

$$q(x) = 0, \quad \forall\, x \in \mathcal{X} \colon p(x) = 0. \tag{6}$$

Combining (4) and (6) we obtain that $p(x) = q(x)$, $\forall\, x \in \mathcal{X}$, is a necessary condition for equality in (1). $\square$