Information Theory I

Fall Semester 2017

Prof. Dr. A. Lapidoth

Signal and Information Processing Laboratory

Institut für Signal- und Informationsverarbeitung

Nonnegativity of Relative Entropy

http://www.isi.ee.ethz.ch/teaching/courses/it1.html

Theorem 1. *Let $p(\cdot)$ and $q(\cdot)$ be probability mass functions on 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. Since the claim is true for $D(p\|q) = \infty$, we assume from now on that $p(x) > 0$ implies $q(x) > 0$ for all $x \in \mathcal{X}$, which is equivalent to $D(p\|q) < \infty$. The claims do not depend on whether $D(p\|q)$ is measured in bits or nats, so we use the natural logarithm. Using that

$$\ln \alpha \leq \alpha - 1 \qquad \forall\, \alpha > 0 \tag{3}$$

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

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

Equality holds if and only if both (a) and (b) hold with equality. The equality condition for (3) implies that (a) holds with equality if and only if

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

Since $q(\cdot)$ is a probability mass function,

$$1 = \sum_{x \in \mathcal{X}} q(x) = \sum_{x \in \mathcal{X}\,:\,p(x)>0} q(x) + \sum_{x \in \mathcal{X}\,:\,p(x)=0} q(x),$$

so (b) holds with equality if and only if

$$q(x) = 0 \qquad \forall\, x \in \mathcal{X} : p(x) = 0,$$

which is equivalent to

$$p(x) = q(x) \qquad \forall\, x \in \mathcal{X} : p(x) = 0. \tag{5}$$

Combining (4) and (5), we see that (1) holds with equality if and only if (2) is satisfied. ∎