MA3364 Tutorial Sheet 5[1]

12 March 2010

1. (C&T 2.32) Fano. X and Y are two independent variables with $\mathcal{X} = \{1, 2, 3\}$ and $\mathcal{Y} = \{a, b, c\}$. $p(1, a) = p(2, b) = p(3, c) = 1/6$, all the other probabilities are $1/12$. Let $\hat{X}(Y)$ be an estimator for X based on Y and let $P_e = Pr[\hat{X} \neq X]$.

   (a) Find the minimum probability of error estimator $\hat{X}(Y)$ and the associated P_e.

   (b) Evaluate Fano's inequality for this problem and compare.

2. (C&T 2.35 and 2.36). Consider two distributions over the set $\{a, b, c\}$: $p(a) = 1/2$ and $p(b) = p(c) = 1/4$ and $q(a) = q(b) = q(c) = 1/3$. Find $H(p)$, $H(q)$, $D(p\|q)$ and $D(q\|p)$ and verify that in this case $D(p\|q) \neq D(q\|p)$. Conversely, give and example of a pair of distinct distributions on the set $\{0, 1\}$ where $D(p\|q) = D(q\|p)$.

3. (C&T 2.37) Let X,Y and Z be three random variable with a joint probability distribution $p(x, y, z)$. The relative entropy between the joint distribution and the product of the marginals is $D(p(x, y, z)\|p(x)p(y)p(z))$; expand this in terms of entropies. When is it zero.

4. An alternative divergence is the λ-divergence,

$$D_\lambda(p\|q) = \lambda D_{\text{KL}}(p\|\lambda p + (1-\lambda)q) + (1-\lambda)D_{\text{KL}}(q\|\lambda p + (1-\lambda)q) \qquad (1)$$

   Unlike the KL divergence, this is symmetric in p and q. According to Wikipedia this can be interpreted as the expected information gain about X from discovering which probability distribution X is drawn from, p or q, if they currently have probabilities λ and $(1-\lambda)$ respectively. Explain this.

5. For $\lambda = 1/2$, the λ-divergence is known as Jensen-Shannon divergence. Show this satisfies

   (a) $D_{\text{JS}}(p_1, p_2) \geq 0$ with equality if and only if $p_1 = p_2$.

   (b) $D_{\text{JS}}(p_1, p_2) = D_{\text{JS}}(p_2, p_1)$.

   However, it does not satisfy the triangular inequality and is therefore not a metric. Give an example of distributions p_1, p_2 and p_3 such that

$$D_{\text{JS}}(p_1, p_2) + D_{\text{JS}}(p_2, p_3) < D_{\text{JS}}(p_1, p_3) \qquad (2)$$

[1] Conor Houghton, houghton@maths.tcd.ie, see also http://www.maths.tcd.ie/~houghton/MA3364