

MA3364 Tutorial Sheet 5, outline solutions<sup>1</sup>

6 April 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.

*Solution:* Well looking at it, the best estimator is

$$\begin{aligned}\hat{X}(a) &= 1 \\ \hat{X}(b) &= 2 \\ \hat{X}(c) &= 3\end{aligned}\tag{1}$$

and  $P_e = 1/2$ ;  $p(1|Y = a) = 1/2$  and so on. Now Fano's inequality tell us

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|Y)\tag{2}$$

So,  $H(P_e) = 1$  and hence the left hand side is

$$H(P_e) + P_e \log |\mathcal{X}| = 1 + \frac{1}{2} \log 3 \approx 1.79\tag{3}$$

and

$$H(X|Y = a) = \frac{1}{2} \log 2 + \frac{1}{2} \log 4 = \frac{3}{2}\tag{4}$$

and, since the entropy conditional on  $Y = b$  and  $Y = c$  will be the same

$$H(X|Y) = \frac{3}{2}\tag{5}$$

Furthermore,  $\hat{X}$  is determined by  $Y$  so  $H(X|Y) = 3/2$  as well.

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 an example of a pair of distinct distributions on the set  $\{0, 1\}$  where  $D(p||q) = D(q||p)$ .

*Solution:* So, by straightforward calculation

$$H(p) = \frac{1}{2} \log 2 + \frac{1}{2} \log 4 = \frac{3}{2}\tag{6}$$

and

$$H(q) = \log 3\tag{7}$$

Next,

$$D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log p(x)q(x)\tag{8}$$

so

$$D(p||q) = \frac{1}{2} \log \frac{3}{2} + \frac{1}{2} \log \frac{3}{4} = \log 3 - \frac{3}{2} = .08496\tag{9}$$

whereas

$$D(q||p) = \frac{1}{3} \log \frac{2}{3} + \frac{2}{3} \log \frac{4}{3} = \frac{5}{3} - \log 3 = .08170\tag{10}$$

so, although they have very similar values, the two KL divergences are different.

An example of a distribution with distinct  $p$  and  $q$  where the divergence is symmetric, try one distribution for two symbols with  $p(0) = p$  and  $p(1) = 1 - p$  and, conversely, the other with  $q(0) = 1 - p$  and  $q(1) = p$ , hence

$$D(p||q) = p \log \frac{p}{1-p} + (1-p) \log \frac{1-p}{p}\tag{11}$$

and

$$D(q||p) = (1-p) \log \frac{1-p}{p} + p \log \frac{p}{1-p}\tag{12}$$

which is the same.

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.

*Solution:* Well this is easy enough when you work out what the question is after; first  $D(p(x, y, z)||p(x)p(y)p(z)) = 0$  if and only if

$$p(x, y, z) = p(x)p(y)p(z)\tag{13}$$

in other words, when the distributions are independent. To expand it in entropies, just write out the definition and then expand the log

$$\begin{aligned}D(p(x, y, z)||p(x)p(y)p(z)) &= \sum_{x,y,z} p(x, y, z) \log \frac{p(x, y, z)}{p(x)p(y)p(z)} \\ &= \sum_{x,y,z} p(x, y, z) \log p(x, y, z) - \sum_{x,y,z} p(x, y, z) \log p(x) \\ &\quad - \sum_{x,y,z} p(x, y, z) \log p(y) - \sum_{x,y,z} p(x, y, z) \log p(z)\end{aligned}\tag{14}$$

<sup>1</sup>Conor Houghton, houghton@maths.tcd.ie, see also <http://www.maths.tcd.ie/~houghton/MA3364>

and now we just use the formula for the marginal densities, for example

$$\sum_{y,z} p(x, y, z) = p(x) \quad (15)$$

to give

$$\begin{aligned} D(p(x, y, z) \| p(x)p(y)p(z)) &= \sum_{x,y,z} p(x, y, z) \log p(x, y, z) - \sum_x p(x) \log p(x) \\ &\quad - \sum_y p(y) \log p(y) - \sum_z p(z) \log p(z) \\ &= H(X, Y, Z) - H(X) - H(Y) - H(Z) \end{aligned} \quad (16)$$

4. An alternative divergence is the  $\lambda$ -divergence,

$$D_\lambda(p \| q) = \lambda D_{\text{KL}}(p \| \lambda p + (1 - \lambda)q) + (1 - \lambda) D_{\text{KL}}(q \| \lambda p + (1 - \lambda)q) \quad (17)$$

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  $\lambda$  and  $(1 - \lambda)$  respectively. Explain this.

*Solution:* So the idea here is to work through the Wikipedia description. Imagine  $X$  is a random variable over a set  $\mathcal{X}$  with two steps, step  $A$  has two possibilities ‘ $P$ ’ with probability  $\lambda$  and ‘ $Q$ ’ with probability  $1 - \lambda$ . Step  $B$  involves the actual selection of the symbol from  $\mathcal{X}$ , this is selected according to  $p$  if the result from  $A$  is  $P$  and according to  $q$  if the result from  $A$  is  $Q$ . Thus, the probability for  $x$  is

$$r(x) = \lambda p(x) + (1 - \lambda)q(x) \quad (18)$$

and so the entropy is

$$\begin{aligned} H(X) &= - \sum_x r(x) \log r(x) \\ &= -\lambda \sum_x p(x) \log r(x) - (1 - \lambda) \sum_x q(x) \log r(x) \end{aligned} \quad (19)$$

On the otherhand; if the result of step  $A$  is known to be  $P$  the probability distribution for selecting the element of  $\mathcal{X}$  is  $p$ , so

$$H(X|A = P) = - \sum_x p(x) \log p(x) \quad (20)$$

and, if the result is  $Q$

$$H(X|A = Q) = - \sum_x q(x) \log q(x) \quad (21)$$

Hence

$$\begin{aligned} H(X|A) &= \lambda H(H(X|A = P)) + (1 - \lambda) H(H(X|A = Q)) \\ &= -\lambda \sum_x p(x) \log p(x) - (1 - \lambda) \sum_x q(x) \log q(x) \end{aligned} \quad (22)$$

Hence the information gain from knowing the result of  $A$  is

$$\begin{aligned} I(X; A) &= H(X) - H(X|A) \\ &= -\lambda \sum_x p(x) \log \frac{r(x)}{p(x)} - (1 - \lambda) \sum_x q(x) \log \frac{r(x)}{q(x)} \\ &= D_\lambda(p \| q) \end{aligned} \quad (23)$$

5. For  $\lambda = 1/2$ , the  $\lambda$ -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

$$D_{\text{JS}}(p_1, p_2) + D_{\text{JS}}(p_2, p_3) < D_{\text{JS}}(p_1, p_3) \quad (24)$$

*Solution:* So, substituting  $\lambda = 1/2$  gives

$$D_{\text{JS}}(p \| q) = \frac{1}{2} D_{\text{KL}}(p \| r) + \frac{1}{2} D_{\text{KL}}(q \| r) \quad (25)$$

where  $r = (p + q)/2$ . From this the first two properties are straight forward, its non-negativeness follows from the fact that it is the sum of two KL divergences, equality can only occur is the two divergences on the right hand side are both zero, so  $p = r$  and  $q = r$  and hence  $p = q$ ; the symmetry under exchange of  $p$  and  $q$  is also obvious from the formula.

To find a counter example to the triangular inequality, obviously we want  $p_1$  and  $p_3$  to be as different as possible, one guess would be to try, in the obvious notation,  $p_1 = (1, 0)$ ,  $p_3 = (0, 1)$  and  $p_2 = (1/2, 1/2)$ , so for  $p_1$  and  $p_2$ ;  $r_{12} = (p_1 + p_2)/2 = (3/4, 1/4)$  so

$$D(p_1 \| r_{12}) = \log 4/3 = 2 - \log 3 \quad (26)$$

whereas

$$D(p_2 \| r_{12}) = \frac{1}{2} \log 2/3 + \frac{1}{2} \log 2 = 1 - \frac{1}{2} \log 3 \quad (27)$$

and so

$$D_{\text{JS}}(p_1 \| p_2) = \frac{3}{2} - \frac{3}{4} \log 3 \quad (28)$$

and, it is clear  $D_{\text{JS}}(p_2 \| p_3)$  will be the same. That leaves  $p_1$  to  $p_3$ ;  $r_{13} = (p_1 + p_3)/2 = (1/2, 1/2)$ , so

$$D(p_1 \| r_{13}) = D(p_3 \| r_{13}) = \log 2 = 1 \quad (29)$$

and hence  $D_{\text{JS}}(p_1 \| p_3) = 1$ . Now, we have

$$D_{\text{JS}}(p_1 \| p_2) + D_{\text{JS}}(p_2 \| p_3) = 3 - \frac{3}{2} \log 3 < D_{\text{JS}}(p_1 \| p_3) = 1 \quad (30)$$

since

$$3 - \frac{3}{2} \log 3 \approx .623 \quad (31)$$