Week two task - TA notes

$I(X, Y) \geq 0$

If we have a joint distribution $p(x, y)$ for two random variables X and Y, we have two distributions on the space of (x, y) pairs, the original distribution $p(x, y)$ and the distribution given by marginalizing the distribution: $p(x)q(y)$.

So for convenience of notation, that is not calling too many things 'p', consider $d(r\|s)$ where $r(x, y) = p(x, y)$ and $s(x, y) = p(x)p(y)$. Now

$$d(r\|s) = \sum_{i,j} r(x_i, y_j) \log_2 \frac{r(x_i, y_j)}{s(x_i, y_j)} = \sum_{i,j} p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{1}$$

and hence $I(X, Y) = d(r\|s) \geq 0$ with equality if and only if $r(x, y) = s(x, y)$, that is, if the distributions are independent.

$H(X) \leq \log_2 n$

So if $q(x) = u(x)$ the uniform distribution with $u(x_i) = 1/n$ for all i we have

$$d(p|u) = \sum_i p(x_i) \log_2 p(x_i) - \sum_i p(x_i) \log_2 \frac{1}{n} \tag{2}$$

and the first term is $-H(X)$ and the second term is $-\log_2 n$ because the $\sum_i p(x_i) = 1$.

The coding question

	A	B	C	D
q	1/2	1/4	1/8	1/8
p	1/4	1/4	1/4	1/4
q-code	0	10	110	111
p-code	00	01	10	11

Check the relationship between the divergence and the difference in code lenghts, both using the code optimized to p and q.

So if $p(x)$ is the distribution, but we use the q-code, then

$$L = \frac{1}{4}(1 + 2 + 3 + 3) = \frac{9}{4} \tag{3}$$

whereas the code for $p(x)$ gives $L = 2$, so the penalty for using the wrong code is $1/4$ when

$$d(p\|q) = \frac{1}{4}\left(\log_2 \frac{1}{2} + \log_2 1 + 2\log_2 2\right) = \frac{1}{4}(-1 + 0 + 2) = \frac{1}{4} \tag{4}$$

Conversely is $q(x)$ is the distribution, using the p-code gives $L = 2$ whereas the efficient code gives $L = 7/4$ as before, giving, again, a gap of $1/4$.

$$d(q\|p) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 1 + \frac{1}{4}\log_2 \frac{1}{2} = \frac{1}{4} \tag{5}$$

The symmetry seen in this problem is not a general property.