6 Information Theory and Coding (JGD)

(*a*) Two random variables X and Y are correlated. The marginal probabilities $p(X)$ and $p(Y)$ are known, as is their joint probability $p(X,Y)$. Give an expression for the conditional probability $p(X|Y)$ using the known quantities. Then, using $p(X)$, $p(Y)$, and $p(X|Y)$, give an expression for the information gained, in bits, from observing Y after X was already observed. [2 marks]

(*b*) Let the random variable X be five possible symbols $\{\alpha, \beta, \gamma, \delta, \epsilon\}$. Consider two probability distributions $p(x)$ and $q(x)$ over these symbols, and two possible coding schemes $C_1(x)$ and $C_2(x)$ for this random variable:

Symbol	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
α	1/2	1/2	0	0
β	1/4	1/8	10	100
γ	1/8	1/8	110	101
δ	1/16	1/8	1110	110
ϵ	1/16	1/8	1111	111

(*i*) Calculate $H(p)$, $H(q)$, and relative entropies (Kullback-Leibler distances) $D(p||q)$ and $D(q||p)$. [4 marks]

(*ii*) Show that the average codeword length of C_1 under p is equal to $H(p)$, and thus C_1 is optimal for p. Show that C_2 is optimal for q. [2 marks]

(*iii*) Now assume that we use code C_2 when the distribution is p. What is the average length of the codewords? By how much does it exceed the entropy $H(p)$? Relate your answer to $D(p||q)$. [2 marks]

(*iv*) If we use code C_1 when the distribution is q, by how much does the average codeword length exceed $H(q)$? Relate your answer to $D(q||p)$. [2 marks]

(*c*) Compare and contrast the compression strategies deployed in the JPEG and JPEG-2000 protocols. Include these topics: the underlying transforms used; their computational efficiency and ease of implementation; artefacts introduced in lossy mode; typical compression factors; and their relative performance when used to achieve severe compression rates. [5 marks]

(*d*) Discuss the following concepts in Kolmogorov's theory of pattern complexity: how writing a program that generates a pattern is a way of compressing it, and executing such a program decompresses it; fractals; patterns that are their own shortest possible description; and Kolmogorov incompressibility. [3 marks]