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:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>$p(x)$</th>
<th>$q(x)$</th>
<th>$C_1(x)$</th>
<th>$C_2(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$1/2$</td>
<td>$1/2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$1/4$</td>
<td>$1/8$</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>110</td>
<td>101</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$1/16$</td>
<td>$1/8$</td>
<td>1110</td>
<td>110</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>$1/16$</td>
<td>$1/8$</td>
<td>1111</td>
<td>111</td>
</tr>
</tbody>
</table>

(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]