2010 Paper 7 Question 11

Information Theory and Coding

(*a*) Let X and Y be two discrete random variables whose respective sets of possible outcomes $\{x\}$ and $\{y\}$ are described by probability distributions $p(x)$ and $p(y)$, and by a joint probability distribution $p(x, y)$.

(*i*) Give an expression for the *mutual information* $I(X; Y)$ between X and Y, using only the probability distributions $p(x)$, $p(y)$, and $p(x, y)$. [2 marks]

(*ii*) In case X and Y are independent random variables, what becomes of their mutual information, and why? [1 mark]

(*iii*) Let the marginal entropy of random variable X be $H(X)$, and suppose that the two random variables X and Y are perfectly correlated with each other. In that case, prove that $I(X; Y) = H(X)$. [2 marks]

(*iv*) What is $I(X; X)$, the mutual information of a random variable with itself, in terms of $H(X)$? [1 mark]

(*b*) Prove that the information measure is additive: that the information gained from observing the combination of N independent events, whose probabilities are p_i for $i = 1....N$, is the *sum* of the information gained from observing each one of these events separately and in any order. [3 marks]

(*c*) An invertible transform generates projection coefficients by integrating the product of a signal onto each of a family of functions. In a reverse process, expansion coefficients can be used on those same functions to reproduce the signal. If the functions in question happen to form an orthonormal set, what is the consequence for the projection coefficients and the expansion coefficients? What penalty must be paid in the absence of orthogonality? Name one such penalised transform. [3 marks]

(*d*) In the Information Diagram (a plane whose axes are time and frequency), why does the Gabor–Heisenberg–Weyl *Uncertainty Principle* imply that information is *quantised* even in continuous signals – *i.e.,* that a signal's information content consists of a countable, limited number of independent quanta? [3 marks]

(*e*) Define the Kolmogorov algorithmic complexity K of a string of data. For a string of length N bits, how large might its Minimal Description Length be, and why? Comment on how, or whether, you can know that you have truly determined the Minimal Description Length for a set of data. Give a reasonable estimate of the Kolmogorov complexity K of a fractal, and explain why it is reasonable. [5 marks]