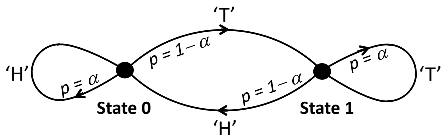
COMPUTER SCIENCE TRIPOS Part II – 2021 – Paper 8

9 Information Theory (jgd1000)

(*a*) A "smart" coin (one with memory) is tossed, whose frequencies of coming up heads ('H') or tails ('T') are equal; but with probability α the outcomes repeat the previous one ($0 < \alpha < 1$).



(*i*) Suppose you know $\alpha = 0.75$, and you observe that a particular outcome is the opposite of the previous one. How much information, in bits, is associated with this improbable observation? [1 mark]

(*ii*) Treating α as a free parameter, provide an expression for the entropy $H(\alpha)$ of this two-state Markov process. What is the maximum possible value of $H(\alpha)$, and how is that compatible with your answer in (*i*)? [3 marks]

(*b*) Consider two discrete probability distributions $p(x)$ and $q(x)$ over the same set of four values $\{x\}$ of a random variable:

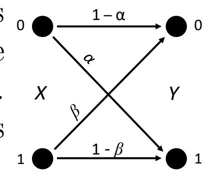
$p(x)$	1/8	1/8	1/4	1/2
$q(x)$	1/4	1/4	1/4	1/4

(*i*) Calculate the cross-entropy $H(p, q)$ between $p(x)$ and $q(x)$. [2 marks]

(*ii*) Calculate their Kullback-Leibler distance $D_{\mathrm{KL}}(p\|q)$. [2 marks]

(*iii*) Comment on the use of metrics $H(p, q)$ and $D_{\mathrm{KL}}(p\|q)$ in machine learning and for calculating the efficiency of codes. [2 marks]

(*c*) Consider an asymmetric binary channel whose input source is the alphabet $X = \{0, 1\}$ with probabilities $(0.5, 0.5)$ and whose outputs are $Y = \{0, 1\}$, but with asymmetric error probabilities. Thus an input 0 is flipped with probability α, but an input 1 is flipped with probability β.



(*i*) Give its channel matrix $p(y_k|x_j)$ and the output probabilities. [3 marks]

(*ii*) Show that the capacity C of this asymmetric binary channel is minimised, $C = 0$, for any combination (α, β) in which $\alpha + \beta = 1$. [2 marks]

(*d*) In the Information Diagram developed by Dennis Gabor, explain the concept of an "atom" and what is irreducible about it. Draw several atoms in this plane representing different trade-offs, labelling the axes of the plane, and explain what all the atoms have in common despite their differences. Write a parameterised expression $f(t)$ defining atoms as functions of time, and explain what makes atoms an optimal basis for representing information in signals. [5 marks]