COMPUTER SCIENCE TRIPOS Part II – 2021 – Paper 9

9 Information Theory (jgd1000)

(*a*) A long-term and self-replicating data storage system based on DNA sequences is being developed. Advantages include huge information density ($\sim 10^{19}$ bits/cm^3) and extreme persistence: dinosaur DNA can still be extracted from fossils. The letters A,C,G,T each occur with equal probability, independently, without sequence constraints. Consider sequences consisting of 100 such letters.

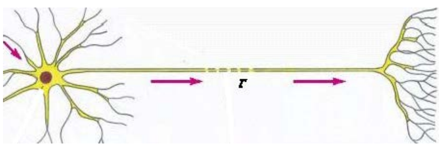
(*i*) How many sequences are possible, and with what probabilities? [2 marks]

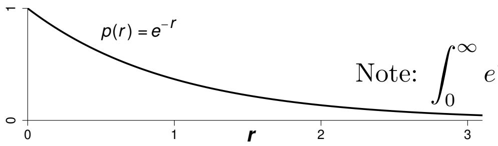
(*ii*) Random variable X selects such a sequence. Calculate $H(X)$, the entropy of X, starting from Shannon's definition. [2 marks]

(*iii*) Sequence replication may be corrupted such that the last two letters are reproduced randomly in the post-replication sequences, denoted Y. What is the conditional entropy $H(X|Y)$, and what is the mutual information $I(X;Y)$ for this error-prone replication process? [4 marks]

(*b*) Financial markets generate daily asset valuations like the time-series $f(t)$ in the left panel, reflecting the dynamics of greed and fear. But underlying such fluctuating indices there may exist meaningful trends, such as a business cycle (right panel). Write an auto-correlation integral that can extract the coherent quasi-periodic signal on the right from noisy valuations $f(t)$, and explain how computing the Fourier transform $F(\omega)$ of $f(t)$ makes it efficient. [5 marks]

(*c*) Brain tissue contains about 10^5 neurones per mm^3, and each neurone has a single output axon whose length r (in dimensionless units) before terminating at synapses to other neurones has probability density distribution $p(r) = e^{-r}$.





Note: $\displaystyle\int_0^\infty e^{-r} dr = 1$

(*i*) Define differential entropy h for continuous random variables in terms of general probability density distribution $p(x)$, and then calculate the value of h in bits for this axonal length distribution $p(r) = e^{-r}$. [5 marks]

(*ii*) If the axon's branching terminals make altogether about 1,000 synapses (connections) with different neurones within the axonal tree's 1 mm^3 volume, uniformly distributed, roughly how many bits of entropy describe the uncertainty of whether a neurone gets such a connection? [2 marks]