

COMPUTER SCIENCE TRIPOS Part IB – 2023 – Paper 6

6 Data Science (djw1005)

Here are two binary sequences:

$$x = 111010100100001011110011$$

$$y = 00000001111111111110000$$

- (a) Consider a model in which each element of x is an independent Bernoulli random variable, $\text{Bin}(1, p)$. Estimate p , and give a formula for the log likelihood of the x sequence. Repeat for y . Explain why the log likelihoods are the same for the two sequences. [4 marks]
- (b) The model from part (a) might seem a poor choice for y . Explain how to conduct a hypothesis test to determine whether this is so. In your answer you should define the test statistic you have invented for the test, and you should explain your choice of one-sided or two-sided testing. [8 marks]
- (c) Consider an alternative model, in which y is generated from a two-state Markov chain, with the first item of y drawn from the chain's stationary distribution. Give a formula for the log likelihood of y , and explain how to fit the model. [8 marks]

[*Note:* You do not need to find numerical answers. You may be interested to know that for the independent model the log likelihoods are both -16.55 , and for the Markov model the log likelihood of x is -16.47 and that for y is -9.00 .]