

COMPUTER SCIENCE TRIPOS Part IB – 2021 – Paper 6

7 Data Science (djw1005)

(a) Let x_t be the number of new COVID infections on date t . We anticipate approximately exponential growth or decay, $x_{t+1} \approx (1 + \lambda)x_t$, and we would like to estimate λ from a dataset (x_1, \dots, x_T) .

(i) Find the maximum likelihood estimator for λ for the model

$$X_{t+1} \sim \text{Poisson}((1 + \lambda)x_t)$$

[2 marks]

(ii) Find the maximum likelihood estimator for λ for the model

$$X_{t+1} \sim \text{Normal}((1 + \lambda)x_t, (\sigma x_t)^2)$$

[3 marks]

(iii) For the latter model, explain how to compute a 95% confidence interval for λ . Explain the resampling step carefully. [4 marks]

(b) We don't actually know the number of new infections x_t on date t : we only know the number of new positive test results, y_t . We anticipate $y_t \approx \beta_{\text{dow}(t)}x_t$, where $\text{dow}(t)$ gives the day of the week for date t . We would like to estimate not only λ but also $\beta_{\text{Mon}}, \dots, \beta_{\text{Sun}}$ from the dataset (y_1, \dots, y_T) .

(i) Propose a probability model for Y_{t+1} in terms of y_t . [5 marks]

(ii) Explain briefly how to estimate the parameters of your model. In your answer, you should consider whether or not the parameters are identifiable. [6 marks]