**COMPUTER SCIENCE TRIPOS  Part IB – mock – Paper 6**
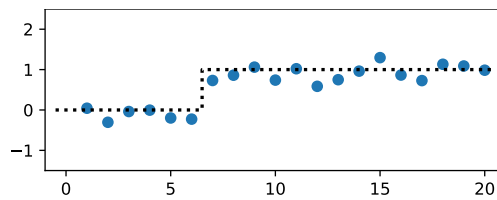
## 1  Foundations of Data Science (DJW)

(a) A 0/1 signal is being transmitted. The transmitted signal at timeslot $i \in \{1, \ldots, n\}$ is $x_i \in \{0, 1\}$, and we have been told that this signal starts at 0 and then flips to 1, i.e. there is a parameter $\theta \in \{1, \ldots, n-1\}$ such that $x_i = 1_{i>\theta}$. The value of this parameter is unknown. The channel is noisy, and the received signal in timeslot $i$ is

$$Y_i \sim x_i + \text{Normal}(0, \varepsilon^2)$$

where $\varepsilon$ is known.



(i) Given received signals $(y_1, \ldots, y_n)$, find an expression for the log likelihood, $\log \Pr(y_1, \ldots, y_n\,;\,\theta)$. Explain your working. [5 marks]

(ii) Give pseudocode for finding the maximum likelihood estimator $\hat{\theta}$. [5 marks]

(b) I have been monitoring average annual river levels for many years, and I have collected a dataset $(z_0, \ldots, z_n)$ where $z_i$ is the level in year $i$ since I started monitoring. I believe that for the first few years the level each year was roughly what it was the previous year, plus or minus some random variation; but that some year a drought started, and since then the level has decreased on average each year. I would like to estimate when the drought started. I do not know the other parameters.

(i) Propose a probability model for my dataset. [5 marks]

(ii) Explain how to fit your model. [5 marks]