8 Data Science (djw1005)

Let $(E_0, E_1, \dots)$ be a Markov chain generated by

$$E_{t+1} = \lambda E_t + \text{Normal}\big(0, (1-\lambda^2)\sigma^2\big)$$

where $0 \le \lambda < 1$.

(*a*) What is meant by "stationary distribution"? Show that $\text{Normal}(0, \sigma^2)$ is a stationary distribution for this Markov chain. [3 marks]

(*b*) What is meant by "memorylessness"? Give an expression for the log likelihood of a sequence of values $(e_1, \dots, e_n)$, given the value for E_0. [4 marks]

Klaus Hasselmann, who won the 2021 Nobel Prize, studied climate models in which short-term random fluctuations can have longer-term effects. Suppose we're given a dataset $(y_0, y_1, \dots, y_n)$ of temperatures at timepoints $t = 0, 1, \dots, n$, and we use a Hasselmann-style model,

$$Y_t = \alpha + \beta \sin(2\pi\omega t) + \gamma t + E_t$$

where $(E_0, E_1, \dots)$ is a Markov chain as described above, and α, β, and γ are unknown parameters to be estimated.

(*c*) Give an expression for the log likelihood of $(y_1, \dots, y_n)$, given the value for Y_0. [*Hint:* First find the distribution of Y_{t+1} given Y_t.] [6 marks]

(*d*) What is meant by a "linear model"? Write out a linear model that can be used to estimate the unknown parameters α, β, and γ (treating all other parameters as known). Identify the feature vectors. [7 marks]