## 7 Foundations of Data Science (djw1005)

Consider the probability model

$$\begin{array}{ccccccccc} X_0 & \to & X_1 & \to & X_2 & \to & X_3 & \to & \cdots \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & Y_1 & & Y_2 & & Y_3 & & \end{array}$$

where $(X_0, X_1, \dots)$ is a Markov chain on state space $\{0,1\}$ with transition probabilities $P_{01} = p$, $P_{10} = q$; and where each $Y_i$ is normally distributed with mean $X_i$ and variance $\sigma^2$.

We are given a sequence of observations $(y_1, y_2, \dots, y_n)$, and we wish to make an inference about the unobserved values $(X_1, X_2, \dots, X_n)$. We will take $0 < p < 1$, $0 < q < 1$, and $\sigma > 0$ to be known, and we will assume that $X_0$ is sampled from the Markov chain's stationary distribution.

(*a*) Write out the transition matrix for the Markov chain $(X_0, X_1, \dots)$. Calculate its stationary distribution. [4 marks]

(*b*) Writing $\vec{X}$ for $(X_0, X_1, \dots, X_n)$, and writing $\vec{Y}$ for $(Y_1, \dots, Y_n)$, and similarly $\vec{x}$ and $\vec{y}$, find expressions for

$$\mathbb{P}(\vec{X} = \vec{x}) \quad \text{and for} \quad \mathbb{P}(\vec{Y} = \vec{y} \mid \vec{X} = \vec{x}).$$

[4 marks]

(*c*) Give pseudocode for a function `rx(n)` that generates a random $\vec{X}$. Give pseudocode to generate a weighted sample from the posterior distribution of $\vec{X}$ conditional on the observed data $\vec{Y} = \vec{y}$. [8 marks]

(*d*) Let $Z = n^{-1} \sum_{i=1}^{n} X_i$. Give pseudocode to find a 95% confidence interval for $Z$, conditional on the observed data $\vec{Y} = \vec{y}$. [4 marks]