Engineering Tripos Part IIA THIRD YEAR

Paper 3F8: Inference

Example Sheet 3: Sequence Modelling and Monte Carlo Methods

Straightforward questions are marked †
*Tripos standard (but not necessarily Tripos length) questions are marked **

Markov Models

1. Markov Models: fitting bi-gram models

   A data scientist observes part of a long sequence that contains $K = 3$ characters: $ABAAABBABCCCBC$. She would like to use a bi-gram model to fit the data with parameters $p(y_1 = k|\theta) = \pi_k^0$ and $p(y_t = k|y_{t-1} = l, \theta) = T_{k,l}$.

   (a) Write down the log-likelihood for the model and optimise it with respect to π^0 and T to find the maximum likelihood parameter estimates.

   (b) Is the maximum-likelihood estimate sensible? How might you improve the estimate?

2. Markov Models: Gaussian AR(1) models

   A data scientist observes a sequence of scalar variables $y_{1:T} = \{y_t\}_{t=1}^T$ generated from a Gaussian AR(1) process $y_t = \lambda y_{t-1} + \epsilon_t$ where $\epsilon_t \sim \mathcal{N}(\mu, \sigma^2)$.

   She knows that the invariant distribution of the process has the following properties

$$\lim_{t\to\infty} \mathbb{E}(y_t) = \mu_\infty, \quad \lim_{t\to\infty} \mathbb{E}(y_t^2) = \sigma_\infty^2 + \mu_\infty^2, \quad \lim_{t\to\infty} \mathbb{E}(y_t y_{t-1}) = \alpha_\infty.$$

   (a) Derive the parameters of the Gaussian AR(1) process $\{\lambda, \mu, \sigma^2\}$ in terms of the properties of the invariant distribution $\{\mu_\infty, \sigma_\infty^2, \alpha_\infty\}$.

   (b) The data scientist reinterprets the original Markov model in terms of a new random variable z_t such that $y_t = z_t + \mu_\infty$. State the form of the distribution $p(z_{1:T})$ required for this model to be equivalent to the original one.

3. Discrete Valued Hidden Markov Models*

   (a) Provide the probabilistic equations that define a Hidden Markov Model (HMM) for observed data that takes discrete values. Indicate what aspects of the model the following terms refer to: *initial state probabilities*, *transition matrix* and *emission matrix*.

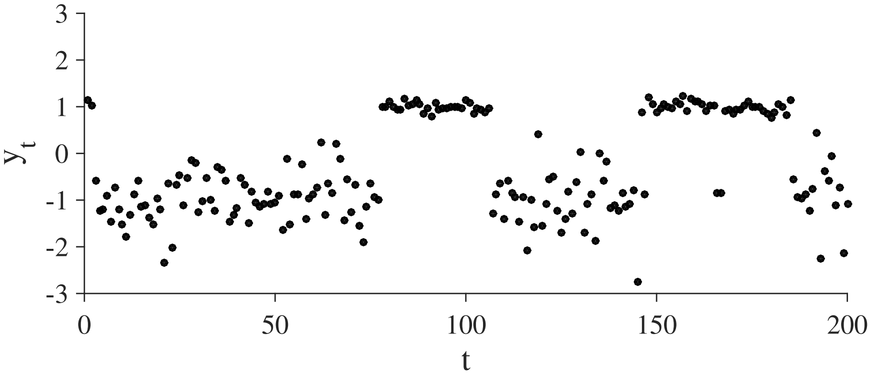
   (b) Consider a dataset consisting of the following string of 160 symbols from the alphabet $\{A, B, C\}$:

   $AABBBACABBBACAAAAAAAAABBBACAAAAABACAAAAAA$
   $BBBBACAAAAAAAAAAAABACABACAABBACAAABBBBACA$
   $AABACAAAABACAABACAAABBACAAAABBBBACABBACAA$
   $AAAABACABACAAABACAABBBACAAAABACABBACA$

   Carefully analyse the string. Describe an HMM model for the string. Your description should include the number of states in the HMM, the transition matrix including the values of the elements of the matrix, the emission matrix including the values of its elements, and the initial state probabilities. Explain your reasoning.

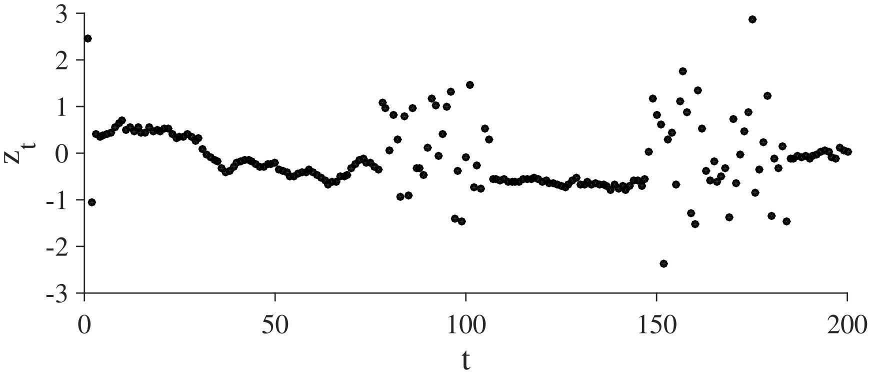
4. Probabilistic Modelling using HMMs for continuous valued observations

(a) A machine learner observes the time-series, y_t, shown below:



Suggest a suitable Hidden Markov Model (HMM) for this sequence and state the model's probabilistic equations. Indicate plausible numerical values for the parameters where possible.

(b) The machine learner is provided with a second set of observations z_t that were measured simultaneously with y_t, shown below:



Extend the HMM you proposed for part (a) so that it can jointly model the first and second set of observations.

5. Inference in HMMs with Discrete Hidden States[†]

   A Hidden Markov Model contains a discrete hidden state variable x_t that takes one of two values and a discrete observed state y_t that also takes one of two values. The hidden state has a transition probability,

$$\left[\begin{array}{cc} P(x_t = 1|x_{t-1} = 1) & P(x_t = 1|x_{t-1} = 2) \\ P(x_t = 2|x_{t-1} = 1) & P(x_t = 2|x_{t-1} = 2) \end{array}\right] = \left[\begin{array}{cc} T_{11} & T_{12} \\ T_{21} & T_{22} \end{array}\right] = \left[\begin{array}{cc} 2/3 & 1/3 \\ 1/3 & 2/3 \end{array}\right].$$

   The filtering distribution at time $t-1$ is

$$P(x_{t-1}|y_{1:t-1}) = \left[\begin{array}{c} P(x_{t-1} = 1|y_{1:t-1}) \\ P(x_{t-1} = 2|y_{1:t-1}) \end{array}\right] = \left[\begin{array}{c} 1/4 \\ 3/4 \end{array}\right].$$

   (a) Compute the predictive distribution for the next hidden state variable, $P(x_t|y_{1:t-1})$.

   (b) Explain how your solution to part (a) can be used to compute the filtering distribution $P(x_t|y_{1:t})$. What additional piece of information would you require to carry out this computation?

6. Forecasting in Linear Gaussian State Space Models[*]

   A simple linear Gaussian state space model with scalar hidden state variables x_t has been used to model scalar observations y_t,

$$p(x_t|x_{t-1}, \lambda, \sigma^2) = \mathcal{N}(x_t; \lambda x_{t-1}, \sigma^2), \quad p(y_t|x_t, \sigma_y^2) = \mathcal{N}(y_t; x_t, \sigma_y^2).$$

   The Kalman filter recursions have been used to process T observations, $y_{1:T}$, in order to return the posterior distribution over the Tth latent state, $p(x_T|y_{1:T}) = \mathcal{N}(x_T; \mu_T, \sigma_T^2)$.

   (a) Explain how to transform the posterior distribution over the Tth latent state into a forecast for the observations one time step into the future, i.e. express $p(y_{T+1}|y_{1:T})$ in terms of μ_T and σ_T^2.

   (b) Now provide a forecast for the observations τ time steps into the future by expressing $p(y_{T+\tau}|y_{1:T})$ in terms of μ_T and σ_T^2.

   (c) What happens to $p(y_{T+\tau}|y_{1:T})$ as $\tau \to \infty$?

Selected solutions and hints

1. a) $\log p(y_{1:T}|\theta) = \log \pi_1^0 + 2\log(T_{11}T_{12}T_{32}T_{33}) + 3\log T_{21} + \log(T_{22}T_{23})$

$$T = \begin{bmatrix} 2/5 & 2/5 & 0 \\ 3/5 & 1/5 & 1/3 \\ 0 & 2/5 & 2/3 \end{bmatrix}, \quad \pi^0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

2. a) $\lambda = (\alpha_\infty - \mu_\infty^2)/\sigma_\infty^2$, $\mu = \mu_\infty(1 + (\mu_\infty^2 - \alpha_\infty)/\sigma_\infty^2)$, $\sigma^2 = \sigma_\infty^2 \left(1 - \left(\frac{\alpha_\infty - \mu_\infty^2}{\sigma_\infty^2}\right)^2\right)$

3. b) pay close attention to repeated patterns and remember that some parts of a HMM can be deterministic

4. b) consider whether the low variance z_t regions are correlated through time and whether a standard HMM could model this

5. a) $p(x_t|y_{1:t-1}) = [5,7]^\top/12$

6. b) $p(y_{T+\tau}|y_{1:T}) = \mathcal{N}(y_{T+\tau}; \lambda^\tau \mu_T, \sigma_y^2 + \lambda^{2\tau}\sigma_T^2 + \sigma^2 \sum_{t'=0}^{\tau-1} \lambda^{2t'})$

Richard E. Turner