UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: Probability and Random Processes

Problem Set 13
Spring 2018

Issued: Monday April 30, 2018 **Due:** N/A

---

1. **Higher-Order Markov Chains**

   Let k be a fixed positive integer. A stochastic process $(X_n)_{n\in\mathbb{N}}$ taking values in a discrete state space $\mathcal{X}$ is called a **kth order (time homogeneous) Markov chain** if for all $n \in \mathbb{N}$ and all feasible sequences $x_0, x_1, \dots, x_{n+k} \in \mathcal{X}$,

$$\begin{aligned}\mathbb{P}(X_{n+k} = x_{n+k} \mid X_0 = x_0, X_1 = x_1, \dots, X_{n+k-1} = x_{n+k-1})\\ = \mathbb{P}(X_{n+k} = x_{n+k} \mid X_n = x_n, \dots, X_{n+k-1} = x_{n+k-1})\\ = P_k(x_{n+k} \mid x_n, \dots, x_{n+k-1}).\end{aligned}$$

   In other words, the transition to the next state depends only on the previous k states. For example, if X_n represents the position of a particle moving with constant velocity at time n, then the system is a second-order Markov chain because the previous two position measurements are needed to infer the particle's velocity.

   Show that we can "embed" $(X_n)_{n\in\mathbb{N}}$ into a *first-order* Markov chain $(Z_n)_{n\in\mathbb{N}}$ with an augmented state space, in the sense that X_n can be recovered from Z_n. This allows us to apply algorithms such as the Viterbi algorithm to systems with higher orders of dependence.

2. **Most Likely Sequence of States**

   In this problem, we give an example of an HMM and a sequence of observations which demonstrates that the most likely sequence of hidden states (i.e., the output of the Viterbi algorithm) is *not* the same as computing the most likely state at each time. Your task is to verify that the following example works:

   Consider a HMM with two states $\{0, 1\}$ and the hidden state is observed through a BSC with error probability $1/3$. The hidden state transitions according to $P(0, 0) = P(1, 1) = 3/4$. Assume that the initial state is equally likely to be 0 or 1. We see the observation 0 at time 0 and 1 at time 1.

3. **EM for Censored Exponential Data**

   A common application of the EM algorithm is for **censored data** in statistics. Let n be a fixed positive integer denoting the sample size; let $X_1, \dots, X_n$ be i.i.d. Exponential(λ) random variables; let $c_1, \dots, c_n$ be known positive constants, and suppose that we observe $Y_i := \mathbb{1}\{X_i > c_i\}$ for each $i = 1, \dots, n$. In other words, we do not get to observe the actual values of $X_1, \dots, X_n$. We only get to observe whether the ith data point is greater than the level c_i. We would like to find the MLE for the rate λ. If we knew the values of $X_1, \dots, X_n$, then we would use $\hat{\lambda} := n/(\sum_{i=1}^n X_i)$.

Applying the EM algorithm to the following problem, we will alternate between the following steps. First, initialize a guess $\hat{\lambda}^{(0)}$. Then, for $t = 0, 1, 2, \dots$:

- **E step**: Compute $\bar{X}^{(t)} := \mathbb{E}_{\hat{\lambda}^{(t)}}[n^{-1} \sum_{i=1}^n X_i \mid Y_1, \dots, Y_n]$, where the notation $\mathbb{E}_{\hat{\lambda}^{(t)}}$ means you should calculate the expectation as if

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exponential}(\hat{\lambda}^{(t)}).$$

- **M step**: The next estimate of the parameter is $\hat{\lambda}^{(t+1)} := 1/\bar{X}^{(t)}$.

(Do not worry about why the E and M steps look the way they do.)

(a) Verify that the MLE estimate of λ given $X_1, \dots, X_n$ is $\hat{\lambda} = n/(\sum_{i=1}^n X_i)$.

(b) Explicitly write out what the E step looks like.

(c) Write out the joint PMF for the observations $Y_1, \dots, Y_n$. Is it possible to find the MLE for λ given $Y_1, \dots, Y_n$ directly?

4. **EM for a Simple HMM**

Consider an HMM $(X_i)_{i \in \mathbb{N}}$ on the state space $\{0, 1\}$, where

$$P(0, 1) = P(1, 0) = \theta \in (0, 1)$$

is an unknown parameter. The hidden state is observed through a BSC with known error probability $\epsilon \in (0, 1)$; let the observations be denoted $(Y_i)_{i \in \mathbb{N}}$. For a fixed positive integer n, suppose that we observe $Y_0, Y_1, \dots, Y_n$. The initial hidden state is equally likely to be 0 or 1.

(a) What is the MLE for θ given $X_1, \dots, X_n$? (Use the notation

$$T = \sum_{i=1}^n \mathbb{1}\{X_i \neq X_{i-1}\}$$

for the number of times that the hidden state switches between 0 and 1.)

(b) We will now derive an EM algorithm to estimate θ given $Y_1, \dots, Y_n$. Initialize a guess $\hat{\theta}^{(0)}$. For $t = 0, 1, 2, \dots$:

- **E step**: Compute $\bar{X}^{(t)} := \mathbb{E}_{\hat{\theta}^{(t)}}[n^{-1} T \mid Y_0, Y_1, \dots, Y_n]$.
- **M step**: In this case, the next parameter estimate is $\hat{\theta}^{(t+1)} := \bar{X}^{(t)}$.

Explicitly write out what the E step is.