

EE5111 - Estimation Theory

(Spring 2019)

Mini Project 3

Released: March 14, 2020 Deadline: March 31, 2020

Study of Expectation Maximization (EM) algorithm

The objective of this exercise is to understand the Expectation Maximization (EM) algorithm. Assume that there are two biased coins, labelled as A and B . Coin A lands up as head with probability p and coin B lands up as head with probability q . An experiment is performed with n independent trials and in each trial, a coin is chosen at random (coin A with probability π and coin B with probability $1 - \pi$) and tossed m times (independently).

Let $z^{(i)} \in \{A, B\}$ be the label of the coin selected and $\mathbf{x}^{(i)} \in \{H, T\}^m$ be the observed sequence in the i^{th} trial of the experiment, $i \in 1, 2, \dots, n$. The labels of the coins $\{z^{(i)}, i \in 1, 2, \dots, n\}$ remain hidden and the observer could only record the faces $\{\mathbf{x}^{(i)}, i \in 1, 2, \dots, n\}$. Hence, the observations $\mathbf{x}^{(i)}$ can be considered as *i.i.d* samples from a *Mixture of two Binomial models*:

$$P_{\mathbf{X}}(\mathbf{x}) = \pi \times P_{\mathbf{X}}(\mathbf{x}|z^{(i)} = A) + (1 - \pi) \times P_{\mathbf{X}}(\mathbf{x}|z^{(i)} = B)$$

Our aim is to obtain maximum likelihood estimate for parameter vector $\boldsymbol{\theta}$ using Expectation Maximization (EM) algorithm. Generate the observations $(\mathbf{x}^{(i)}, z^{(i)})$ for following two cases

- (i) $\pi = 0.50$, $p = 0.35$ and $q = 0.60$.
- (ii) $\pi = 0.25$, $p = 0.35$ and $q = 0.60$.

Choose $m = 1, 10$ and $n = 10, 1000, 10000$. Run the EM algorithm for following experiments

Experiment: 1 Assume that we know π . Thus, the vector of parameters is given by $\boldsymbol{\theta} = [p \ q]^T$.

- (a) Use observations from case (i) and assume that $\pi = 0.50$ and initial estimates are $[0.45 \ 0.50]^T$
- (b) Use observations from case (ii) and assume that $\pi = 0.50$ and initial estimates are $[0.45 \ 0.50]^T$
- (c) Use observations from case (ii) and assume that $\pi = 0.25$ and initial estimates are $[0.45 \ 0.50]^T$

Experiment: 2 Assume that we do not know π . Thus, the vector of parameters is given by $\boldsymbol{\theta} = [\pi \ p \ q]^T$.

- (a) Use observations from case (ii) and initial estimates are $[0.50 \ 0.45 \ 0.50]^T$
- (b) Repeat above experiment with initial estimate $[0.50 \ 0.50 \ 0.45]^T$

Experiment: 3 Now use a Beta prior over π . Derive the update equations for MAP estimate. Use the following priors $Beta(1, 1)$, $Beta(1, 3)$, $Beta(2, 6)$, $Beta(3, 9)$.

Make the following inferences from the algorithm for the aforementioned choices of n , m and $\hat{\boldsymbol{\theta}}_0$:

- (1) Plot the *learning curve*¹ and show the convergence of the algorithm² for each experiment.
- (2) Observe how the final estimate of $\boldsymbol{\theta}$ from the algorithm (call it $\hat{\boldsymbol{\theta}}_{EM}$), and number of iterations needed for convergence, change when we increase m and n .
- (3) Report your observation about case (b) and case (c) of experiment 1, where in former case we are assuming a wrong value of π and in later case we are assuming the true value of π .
- (4) Report your observation about experiment 2 when we swap the prior for p and q .
- (5) Compare the result of experiment 3 with the result of experiment 2.

¹Plot of the estimate at iteration k , $\hat{\boldsymbol{\theta}}_k$ vs. iteration index, k .

²You shall consider that the algorithm has converged at iteration k when the update to any of the parameter is not more than $\epsilon = 10^{-6}$ (i.e., $\|\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\theta}}_{k-1}\|_\infty = \max(|\hat{\theta}_k - \hat{\theta}_{k-1}|) \leq \epsilon$).