Exercise 1: EM Algorithm for Coin Toss Mixture Model

Problem Setup

We have two coins:

- Coin c_0: Heads probability q_0
- Coin c_1: Heads probability q_1

A player picks coin c_1 with probability θ and coin c_0 with probability $1-\theta$. For each trial i, we observe $x_i \in \{H, T\}$ and have an unobserved latent variable $z_i \in \{0, 1\}$ indicating which coin was chosen. Our goal is to estimate θ using the EM algorithm.

a) Complete-data likelihood $p(x, z|\theta)$

For a single observation (x, z):

$$p(x, z|\theta) = p(z|\theta) \cdot p(x|z) \tag{1}$$

where:

$$p(z = 1|\theta) = \theta, \quad p(z = 0|\theta) = 1 - \theta$$

If $z = 0$: $p(x = H|z = 0) = q_0$, $p(x = T|z = 0) = 1 - q_0$. If $z = 1$: $p(x = H|z = 1) = q_1$, $p(x = T|z = 1) = 1 - q_1$. Thus:

$$p(x, z|\theta) = \begin{cases} (1-\theta)q_0^{\mathbb{1}(x=H)}(1-q_0)^{\mathbb{1}(x=T)}, & z = 0 \\ \theta q_1^{\mathbb{1}(x=H)}(1-q_1)^{\mathbb{1}(x=T)}, & z = 1 \end{cases} \tag{2}$$

b) Observed-data likelihood $p(x|\theta)$

Marginalizing over z:

$$p(x|\theta) = (1-\theta)p(x|z = 0) + \theta p(x|z = 1) \tag{3}$$

So:

$$p(x|\theta) = (1-\theta)[q_0^{\mathbb{1}(x=H)}(1-q_0)^{\mathbb{1}(x=T)}] + \theta[q_1^{\mathbb{1}(x=H)}(1-q_1)^{\mathbb{1}(x=T)}] \tag{4}$$

This is a mixture model (Bernoulli mixture), not a Gaussian mixture.

c) Missing-data distribution $p(z|x,\theta)$

By Bayes' theorem:

$$p(z=1|x,\theta) = \frac{\theta p(x|z=1)}{p(x|\theta)} \tag{5}$$

So:

$$p(z=1|x,\theta) = \frac{\theta q_1^{\nVdash(x=H)}(1-q_1)^{\nVdash(x=T)}}{(1-\theta)q_0^{\nVdash(x=H)}(1-q_0)^{\nVdash(x=T)} + \theta q_1^{\nVdash(x=H)}(1-q_1)^{\nVdash(x=T)}} \tag{6}$$

Similarly $p(z=0|x,\theta) = 1-\gamma(x)$.

d) Q-function for EM

For N observations:

$$Q(\theta|\theta^{(t)}) = \mathbb{E}_{Z|X,\theta^{(t)}}[\log p(X,Z|\theta)] \tag{7}$$

Complete-data log-likelihood:

$$\log p(X,Z|\theta) = \sum_{i=1}^{N}[z_i \log\theta + (1-z_i)\log(1-\theta)] + \text{const} \tag{8}$$

So:

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{N}[\mathbb{E}[z_i]\log\theta + (1-\mathbb{E}[z_i])\log(1-\theta)] \tag{9}$$

Where $\mathbb{E}[z_i] = p(z_i=1|x_i,\theta^{(t)}) = \gamma_i$.

E-step

Compute:

$$\gamma_i = p(z_i=1|x_i,\theta^{(t)}) \tag{10}$$

e) M-step

Maximize:

$$Q(\theta) = \sum_{i=1}^{N}[\gamma_i\log\theta + (1-\gamma_i)\log(1-\theta)] \tag{11}$$

Derivative:

$$\frac{\partial Q}{\partial\theta} = \sum_{i=1}^{N}\frac{\gamma_i}{\theta} - \sum_{i=1}^{N}\frac{1-\gamma_i}{1-\theta} = 0 \tag{12}$$

Solve:

$$\theta^{(t+1)} = \frac{\sum_{i=1}^{N}\gamma_i}{N} \tag{13}$$

Final EM Algorithm

1. Initialize $\theta^{(0)}$.
2. **E-step**: Compute γ_i for each observation.
3. **M-step**: Update $\theta^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} \gamma_i$.
4. Repeat until convergence.