

Expectation Maximization Algorithm with Three-Coin Problem

Zengchang Qin (PhD)

3-Coin problem with Expectation Maximization

- 1) Given a coin with the probability p of being head.
The probability distribution is a Bernoulli distribution:

$$p(x) = p^x (1-p)^{1-x}$$

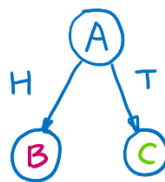
Given a sequence of results "HTTTHH...H"

p is estimated by:

$$p = \frac{\sum_{i=1}^N I(X_i = H)}{N}, \text{ where } I(\cdot) \text{ is an indicator function.}$$

- 2) There are three coins A, B and C with parameters π , p and q , respectively.

$$\begin{aligned} \textcircled{A} &\sim \pi \\ \textcircled{B} &\sim p \\ \textcircled{C} &\sim q \end{aligned}$$



If the tossing result of A is H, we toss coin B, otherwise we toss the coin C, the tossing results of B and C are recorded as a sequence.

E.g. HHTTHTHTTH

- 3) If we assume to know which coin generates the result.

For example:

HHTTHTHTTH
HHTTHTHTTH

The results of the coin B is in red and coin C is in green.

If we have the colored results, we can estimate parameters by:

$$\hat{\pi} = \frac{\sum_{i=1}^N (x_i = \text{Red})}{N} = \frac{\sum_{i=1}^N (x_i = \text{Red})}{\sum_{i=1}^N (x_i = \text{Red}) + \sum_{i=1}^N (x_i = \text{Green})} = \frac{6}{10}$$

$$\hat{p} = \frac{\sum_{i=1}^N (x_i = \text{Red}) \wedge (x_i = H)}{\sum_{i=1}^N (x_i = \text{Red})} = \frac{\sum_{i=1}^N (x_i = H)}{\sum_{i=1}^N (x_i = \text{Red})} = \frac{5}{6}$$

$$\hat{q} = \frac{\sum_{i=1}^N (x_i = \text{Green}) \wedge (x_i = H)}{\sum_{i=1}^N (x_i = \text{Green})} = \frac{\sum_{i=1}^N (x_i = H)}{\sum_{i=1}^N (x_i = \text{Green})} = \frac{1}{4}$$

4) For the above 3-coin problem, the parameter $\theta = (\pi, p, q)$

$$L(\theta) = \arg\max_{\theta} P(x|\theta)$$

$$P(x|\theta) = \sum_z P(x, z|\theta) \\ = \sum_z P(x|z, \theta) P(z|\theta)$$

where z is a latent (hidden) variable, $z = \{0, 1\}$

$z=1$ indicates the case of Coin **B** and $z=0$ is for Coin **C**. Therefore:

$$P(x|\theta) = \sum_z P(x|z, \theta) P(z|\theta) \\ = \underbrace{P(x|z=1, \theta)}_{\text{Bernoulli distribution}} \underbrace{P(z=1|\theta)}_{\text{Prior}} + \underbrace{P(x|z=0, \theta)}_{\text{Bernoulli distribution}} P(z=0|\theta)$$

$$P(x|z=1, \theta) = p^x (1-p)^{1-x}$$

$$P(x|z=0, \theta) = q^x (1-q)^{1-x}$$

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

$$P(x|\theta) = \underbrace{\pi p^x (1-p)^{1-x}}_{\text{Coin B}} + \underbrace{(1-\pi) q^x (1-q)^{1-x}}_{\text{Coin C}}$$

We use μ to represent the probability of x is tossing by B given current parameter $\theta = (\pi, p, q)$

$$\mu = \frac{\pi p^x (1-p)^{1-x}}{\pi p^x (1-p)^{1-x} + (1-\pi) q^x (1-q)^{1-x}} \quad x \in \{0, 1\}$$

$x=1$ indicates H and $x=0$ indicates T.

Given a sequence of tossing results $X = \{x_1, x_2, \dots, x_N\}$

$$\mu_i = \frac{\pi p^{x_i} (1-p)^{1-x_i}}{\pi p^{x_i} (1-p)^{1-x_i} + (1-\pi) q^{x_i} (1-q)^{1-x_i}}$$

5) Given a sequence of tossing results, we now have a probability estimation of being **B** or **C**.

H H T T H T H T H H



It is like H or T cannot be 100% of being red or green, it has a "probabilistic degree" of being red or green.

Based on Section (3) we can estimate θ using current μ .

$$\pi = \frac{\sum_{i=1}^N \mu_i}{N} \quad \text{percentage of being red of all } N \text{ tosses}$$

$$p = \frac{\sum_{i=1}^N \mu_i x_i}{\sum_{i=1}^N \mu_i} \quad \text{percentage of red H among red}$$

$$q = \frac{\sum_{i=1}^N (1-\mu_i) x_i}{\sum_{i=1}^N (1-\mu_i)} \quad \text{percentage of green H among green}$$

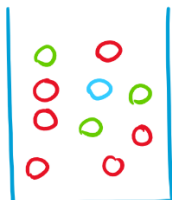
6) This can be Summarized as E-step and M-step.

E-step: calculating μ using θ

M-step: recalculate θ using μ from the last step.

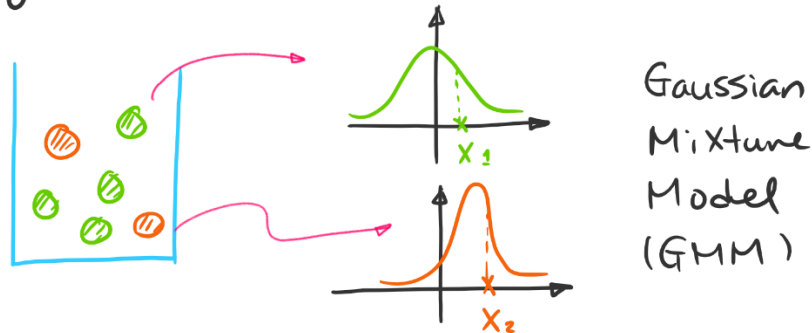
The problem can be considered as a mixture model.

To be precise, it is a **Bernoullian Mixture** model.

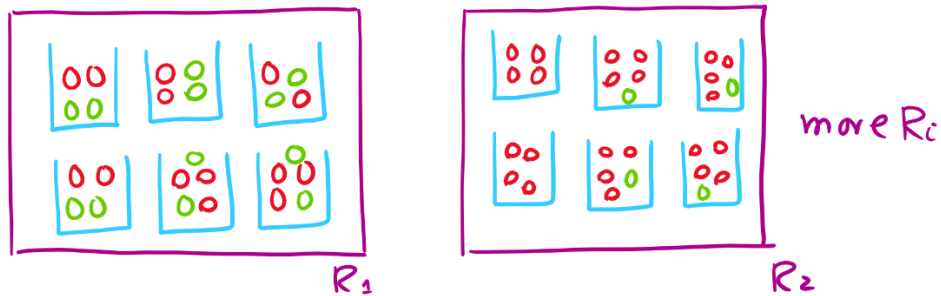


π can be considered as weights of the mixture of coin **B** and **C**.

Inside the box, it could be coins following Bernoulli distribution, could be dice, could be any event following some particular distribution, E.g.



7) The mixtures may grouped into different "rooms"



The mixture of boxes may following some distribution, pattern with some uncertainties.

$$(\tau_1(R_1), \tau_2(R_1)) \sim \text{Dir}(\alpha_1, \alpha_2)$$

$$(\tau_1(R_2), \tau_2(R_2)) \sim \text{Dir}(\beta_1, \beta_2)$$