EM Algorithm

Zahoor Tanoli (PhD)

CUI Attock

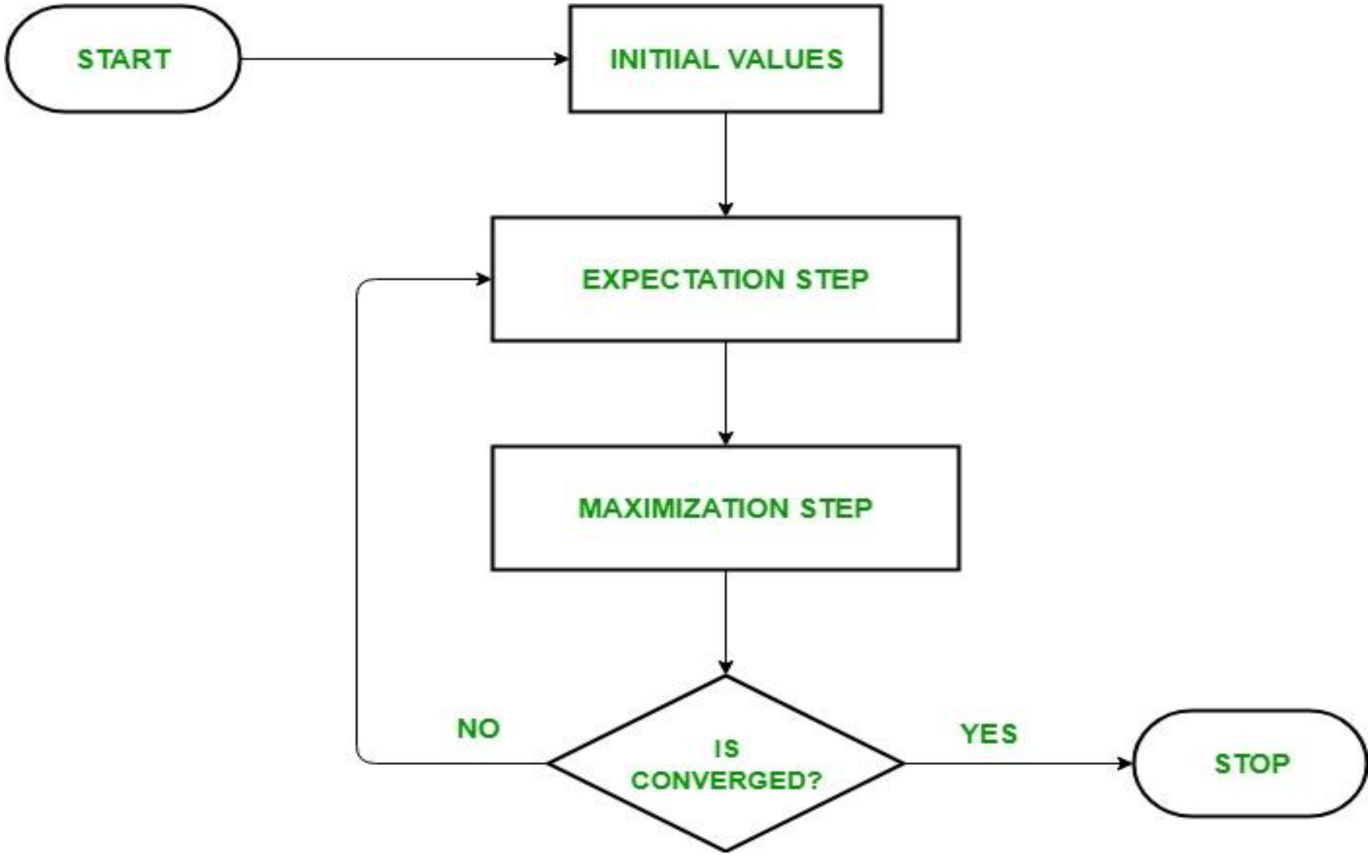
Background

- Expectation-Maximization algorithm can be used for the latent variables e.g. quality of life, business confidence, happiness and conservatism
- It was explained, proposed and given its name in a paper published in 1977 by Arthur Dempster, Nan Laird, and Donald Rubin
- Used to find the local maximum likelihood parameters of a statistical model in the cases where latent variables are involved and the data is missing or incomplete

Steps of Algorithms

- Given a set of incomplete data, consider a set of starting parameters.
- Expectation step (E – step): Using the observed available data of the dataset, estimate (guess) the values of the missing data.
- Maximization step (M – step): Complete data generated after the expectation (E) step is used in order to update the parameters.
- Repeat point 2 and 3 until convergence

Flow chart for EM algorithm



Usage of EM algorithm

- Used to fill the missing data in a sample
- Used as the basis of unsupervised learning of clusters
- It can be used for the purpose of estimating the parameters of Hidden Markov Model (HMM)
- Used for discovering the values of latent variables

EM Algorithm
The Case for Two Conditions with Success or Failure

Imagine there are two coins A and B

One is more likely to get Heads, the other more likely to get Tails.

You pick one at random and toss it. Which one was it?

Well, let's do this five times:

- pick a coin randomly
- toss it 10 times
- record the number of heads and tails

Then, get the average number of heads for each coin.

[1] Cuong B. Do and Serafim Batzoglou (2008)

EM Algorithm
A Tutorial

a Maximum likelihood

		Coin A	Coin B
B	H T T T H H T H T H		5 H, 5 T
A	H H H H T H H H H H	9 H, 1 T	
A	H T H H H H H T H H	8 H, 2 T	
B	H T H T T T H H T T		4 H, 6 T
A	T H H H T H H H T H	7 H, 3 T	
		24 H, 6 T	9 H, 11 T

5 sets, 10 tosses per set

$$\hat{\theta}_A = \frac{24}{24+6} = 0.80$$

$$\hat{\theta}_B = \frac{9}{9+11} = 0.45$$

EM Algorithm
A Tutorial

That was easy: Coin A yields heads 80% of the time, Coin B 45% of the time.

What if we are given ONLY the results of our coin tosses

Can we guess the percentage of heads that each coin yields?

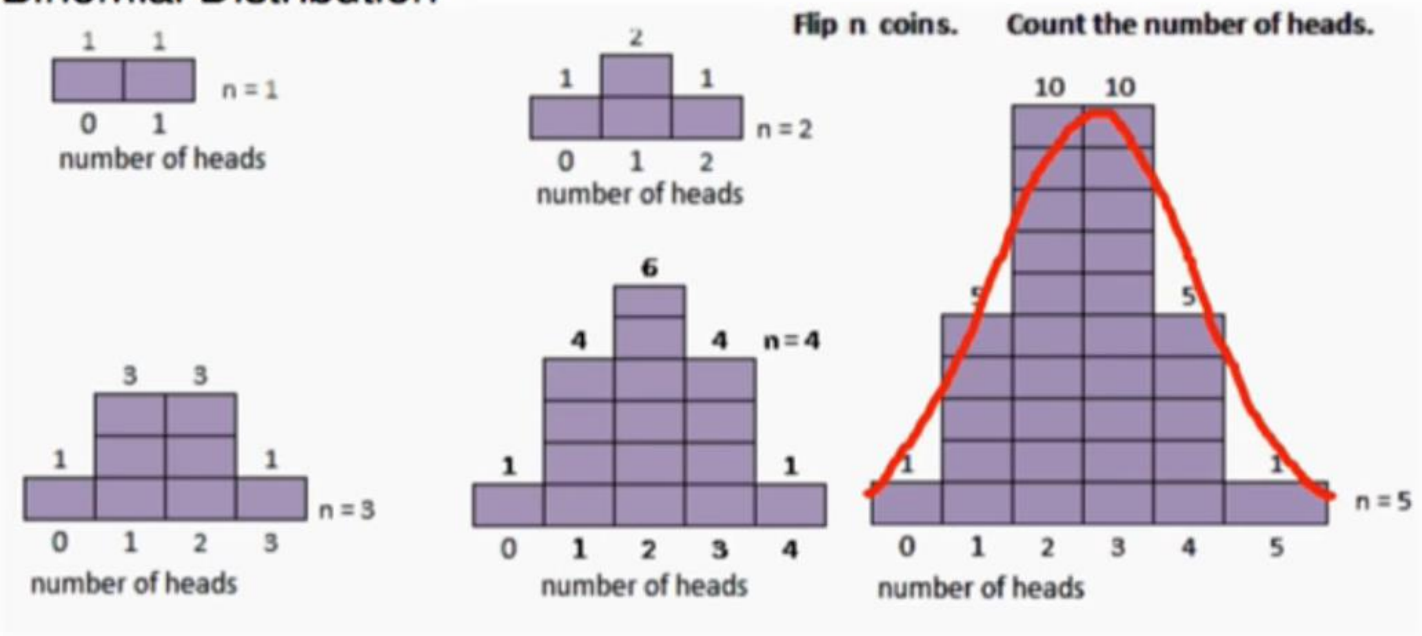
Can we guess which coin was picked for each set of 10 coin tosses?

One way to think about this is:

1. Assign random averages to both coins
2. For each of the 5 rounds of 10 coin tosses
    - Check the percentage of heads
    - Find the probability of it coming from each coin
    - Compute the expected number of heads: using that probability as a weight, multiply it by the number of heads
    - Record those numbers
    - Re-Compute new means for coin A and B.
3. With these new means go back to step 2.

EM Algorithm
How do Coin Tosses Behave?[2]

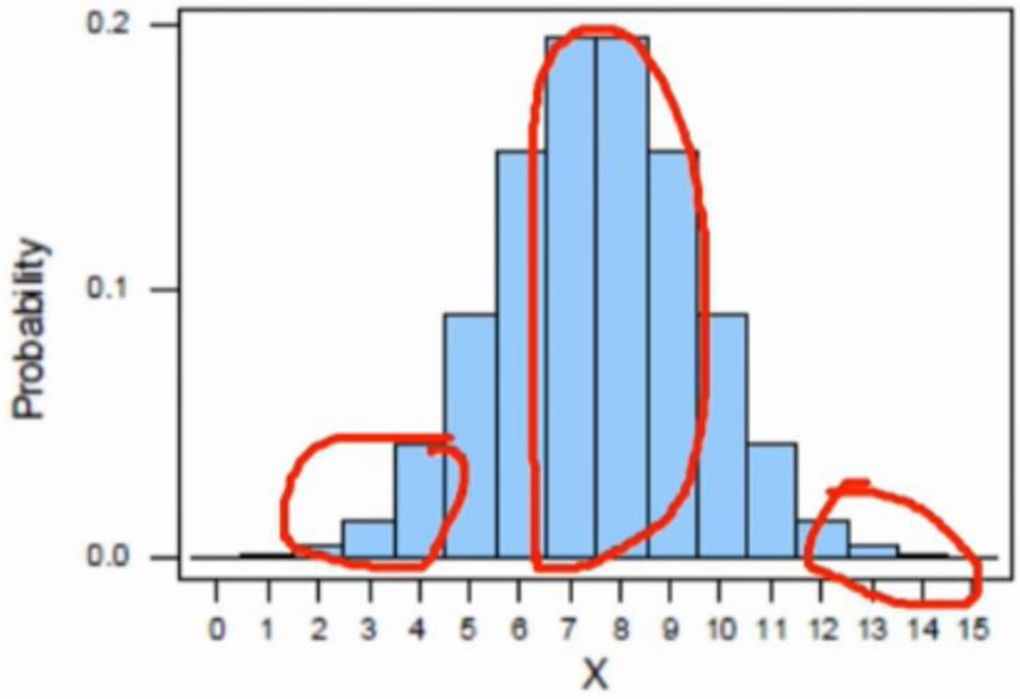
Binomial Distribution



[2] Math Spoken Here. Binomial Distribution

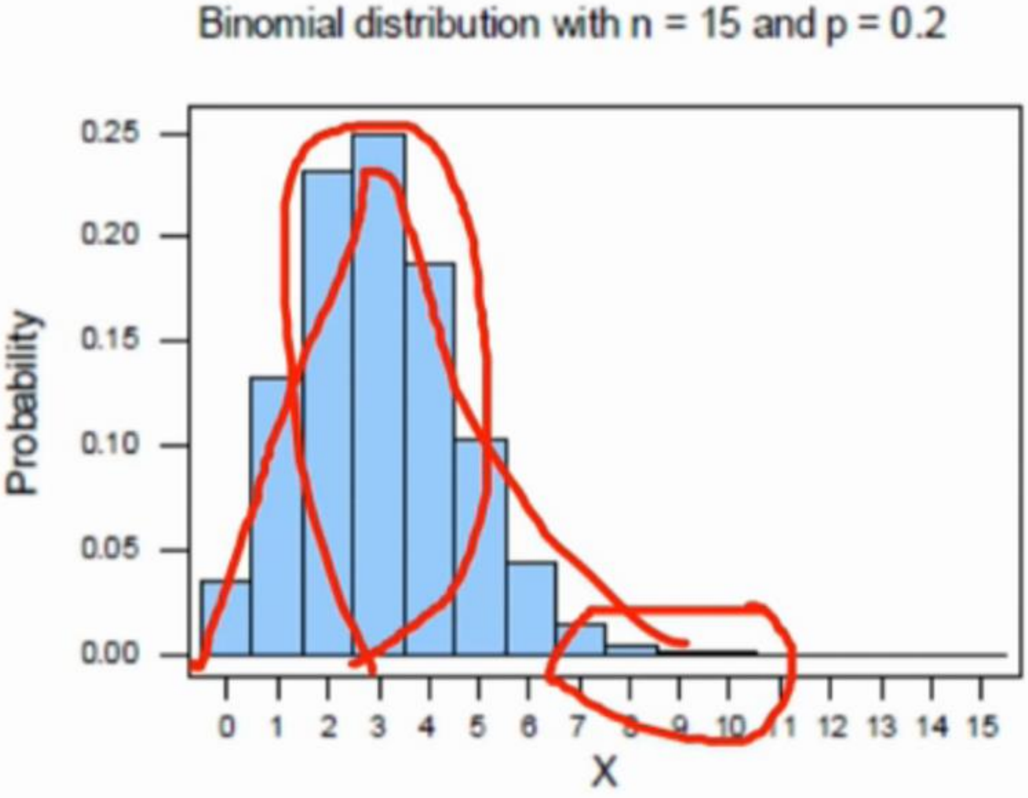
EM Algorithm
How do Coin Tosses Behave?[3]

Binomial distribution with n = 15 and p = 0.5



[3]PennState Eberly College of Science Statistics Online

EM Algorithm
How do Coin Tosses Behave?[4]



[4]PennState Eberly College of Science Statistics Online

EM Algorithm
A Tutorial

So, the five rounds of 10 coin tosses with $\theta_A = 0.6$; $\theta_B = 0.5$[5] yield:

1	H	T	T	T	H	H	T	H	T	H
2	H	H	H	H	H	T	H	H	H	H
3	H	T	H	H	H	H	H	T	H	H
4	H	T	H	T	T	T	H	H	T	T
5	T	H	H	H	T	H	H	H	T	H

Let's take the first round: $\frac{5}{10}$ heads and $\frac{5}{10}$ tails.

compute the likelihood that it was coin "A" and coin "B" using the binomial distribution with mean probability θ on n trials with k successes. $p(k) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$

[5] θ_i is the average number of heads for coin i. Initially it is randomly assigned

EM Algorithm
A Tutorial: E-Step

So, We have:

$\theta_A = 0.6$; $\theta_B = 0.5$

1	H	T	T	T	H	H	T	H	T	H
2	H	H	H	H	H	T	H	H	H	H
3	H	T	H	H	H	H	H	T	H	H
4	H	T	H	T	T	T	H	H	T	T
5	T	H	H	H	T	H	H	H	T	H

Let's take the first round: $\frac{5}{10}$ heads and $\frac{5}{10}$ tails.

likelihood of "A"$= p_A(h)^h(1 - p_A(h))^{10-h} = 0.0007962624$
likelihood of "B"$= p_B(h)^h(1 - p_B(h))^{10-h} = 0.0009765625$
Normalizing I get probabilities: 0.45 and 0.55

EM Algorithm
A Tutorial. M-Step

So, We have:

$\theta_A = 0.6;\ \theta_B = 0.5$

1	H	T	T	T	H	H	T	H	T	H
2	H	H	H	H	H	T	H	H	H	H
3	H	T	H	H	H	H	H	T	H	H
4	H	T	H	T	T	T	H	H	T	T
5	T	H	H	H	T	H	H	H	T	H

Recap: $P(Coin = A) = 0.45;\ P(Coin = B) = 0.55$

Estimating likely number of heads and tails from:

- "A": $H = 0.45 \times 5\ heads = 2.2\ heads;\ \ T = 0.45 \times 5\ tails = 2.2\ tails$
- "B": $H = 0.55 \times 5\ heads = 2.8\ heads;\ \ T = 0.55 \times 5\ tails = 2.8\ tails$

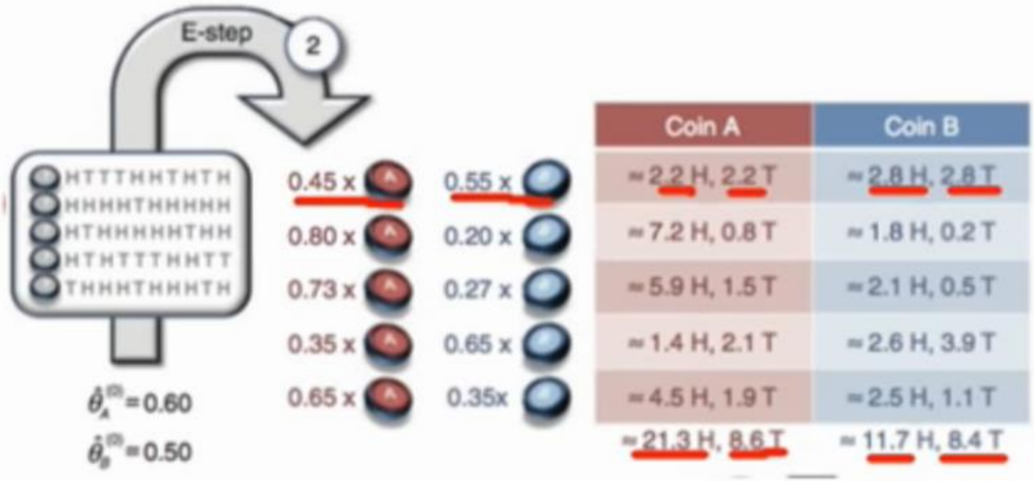
Do the same for all five runs

EM Algorithm
A Tutorial

So, We have:

$\theta_A = 0.6;\ \theta_B = 0.5$

b Expectation maximization



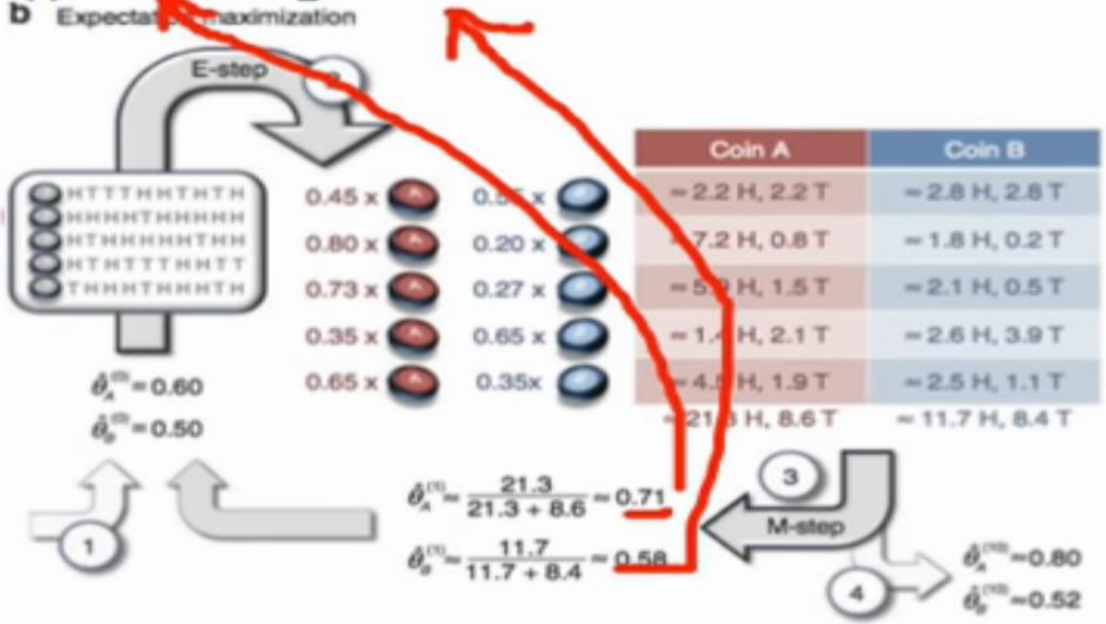
Compute the new probabilities for each coin ($\frac{H}{H+T}$)

That gives you the new maximized parameter θ for each coin

EM Algorithm
A Tutorial

So, We have:

$\theta_A = 0.6; \ \theta_B = 0.5$

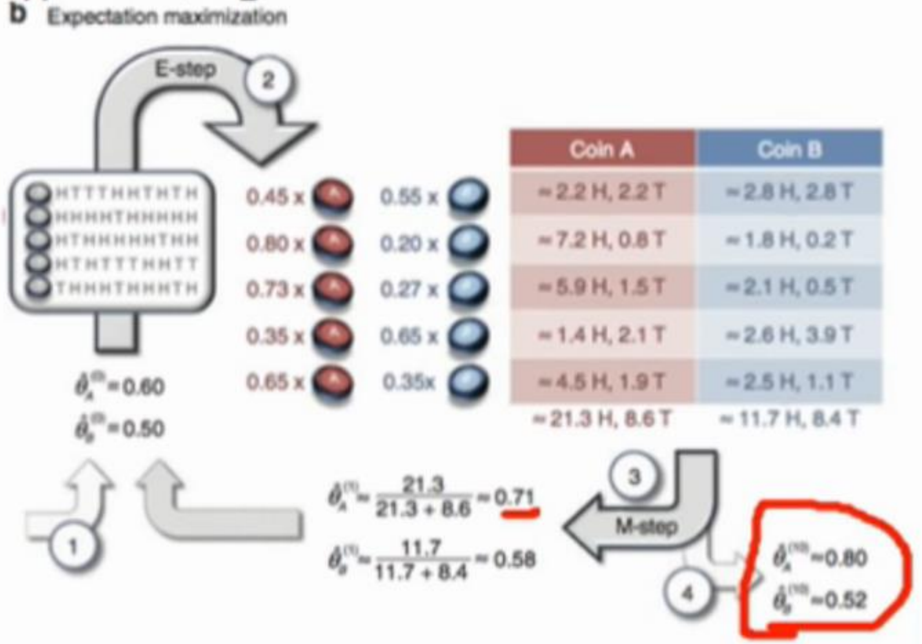


Repeat E-Step and M-Step until convergence

EM Algorithm
A Tutorial

So, We have:

$\theta_A = 0.6; \ \theta_B = 0.5$



Repeat E-Step and M-Step until convergence

Application Areas

1. Credit Risk Modeling: EM algorithm is used in credit risk modeling to estimate the parameters of statistical models, such as the default probability or hazard rate. It helps in modeling the creditworthiness of borrowers and predicting default events.

2. Credit Scoring and Fraud Detection: EM algorithm is applied in credit scoring models to estimate the parameters that determine the creditworthiness of individuals or businesses. It is also used in fraud detection models to estimate the parameters that distinguish fraudulent transactions from legitimate ones.

3. Text Mining and Topic Modeling: EM algorithm is employed in text mining tasks, such as topic modeling using Latent Dirichlet Allocation (LDA). EM helps estimate the latent variables (topics) and the associated probabilities for each word in the documents.