Introduction to Statistical Learning

Omid Safarzadeh

February 2, 2022

Overview

1. Introduction
2. MLE
3. EM algorithm

***Acknowledgement**: This slide is prepared based on Do and Batzoglou, 2008

Introduction

Flip Coin Problem:

- $Pr(Head \quad A) = \theta_A$, $Pr(Head \quad B) = \theta_B$
- Our goal is to estimate $\theta = (\theta_A, \theta_B)$

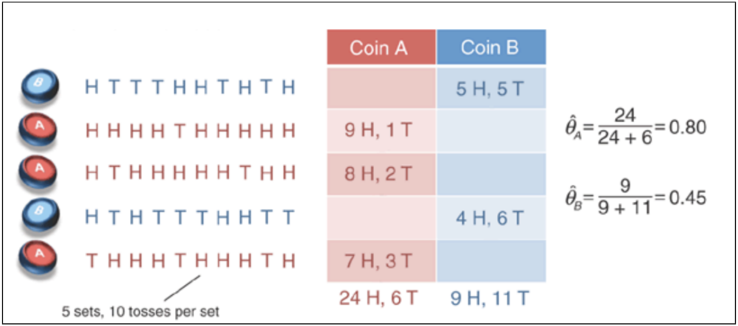
How to estimate θ

Let's design an experiment:

1. randomly choose one of the two coins (with equal probability), and perform ten independent coin tosses with the selected coin .
2. repeat above 5 times
3. $x = (x_1, x_2, ..., x_5)$. $x_i \in \{0, 1, ..., 10\}$ shows the number of heads in each set.
4. $z = (z_1, z_2, ..., z_5)$. $z_i \in \{A, B\}$ is the identity of the coin used during the i-th set of tosses
5. note that, we choose the coin at each set of experiments with 0.5 probability.
6. the complete data case

MLE of θ

If $\log P(x, z; \theta)$ is the logarithm of the joint probability (or log likelihood) of obtaining any particular vector of observed head counts x and coin types z, then the formulas solve for the parameters θ that maximize $logP(x, z; \theta)$ is:



Challenging Problem

we are given the recorded head counts x but not the identities z of the coins used for each set of tosses.

- z is hidden variables or latent factor.
- incomplete data case.
- EM-Algorithm!

Iterative EM

1. initial parameters, θ,
2. determine for each of the five sets whether coin A or coin B was more likely to have generated the observed flips using the current parameter estimates.

$$\hat{\theta_A}^{(0)} = 0.6$$
$$\Rightarrow P(x = 5|\theta_A = 0.6, z = A) =$$
$$(\hat{\theta_A}^{(0)})^5 * (1 - \hat{\theta_A}^{(0)})^5 = 0.000796262$$

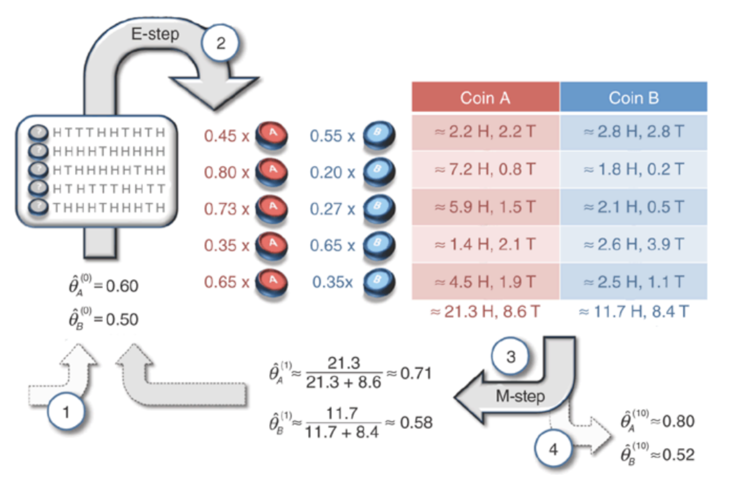
$$\hat{\theta_B}^{(0)} = 0.5$$
$$\Rightarrow P(x = 5|\theta_B = 0.6, z = B) =$$
$$(\hat{\theta_B}^{(0)})^5 * (1 - \hat{\theta_B}^{(0)})^5 = 0.000976563$$

Expectation Step

Now for each set estimate θ:

$$\hat{\theta_A} = P(z = A|Data) =$$
$$= \frac{P(x = 5|\theta_A = 0.6, z = A)}{P(x = 5|\theta_A = 0.6, z = A) + P(x = 5|\theta_B = 0.6, z = B)} = 0.449$$

$$\hat{\theta_B} = P(z = B|Data) =$$
$$= \frac{P(x = 5; \theta_A = 0.6, z = B)}{P(x = 5; \theta_A = 0.6, z = A) + P(x = 5; \theta_B = 0.6, z = B)} = 0.550$$


E-step
2
HTTTHHTHTH
HHHHTHHHHH
HTHHHHHTHH
HTHTTTHHTT
THHHTHHHTH
0.45 x
0.55 x
0.80 x
0.20 x
0.73 x
0.27 x
0.35 x
0.65 x
0.65 x
0.35x
Coin A
Coin B
≈ 2.2 H, 2.2 T
≈ 2.8 H, 2.8 T
≈ 7.2 H, 0.8 T
≈ 1.8 H, 0.2 T
≈ 5.9 H, 1.5 T
≈ 2.1 H, 0.5 T
≈ 1.4 H, 2.1 T
≈ 2.6 H, 3.9 T
≈ 4.5 H, 1.9 T
≈ 2.5 H, 1.1 T
≈ 21.3 H, 8.6 T
≈ 11.7 H, 8.4 T
$\hat{\theta}_A^{(0)} = 0.60$
$\hat{\theta}_B^{(0)} = 0.50$
1
$\hat{\theta}_A^{(1)} \approx \frac{21.3}{21.3 + 8.6} \approx 0.71$
$\hat{\theta}_B^{(1)} \approx \frac{11.7}{11.7 + 8.4} \approx 0.58$
3
M-step
4
$\hat{\theta}_A^{(10)} \approx 0.80$
$\hat{\theta}_B^{(10)} \approx 0.52$

Maximization Step

Now for each set find contribution of each set to coin A

$$0.45 * (5H, 5T) = (2.2H, 2.2T)$$

Now for each set find contribution of each set to coin B

$$0.45 * (5H, 5T) = (2.8H, 2.8T)$$

Sum these for each coin over the sets. then find new θ

Steps of EM Algorithm

1. Compute Posterior over latent variable z_j
2. Compute Expected likelihood under posterior of z_j
3. Maximize the expected likelihood
4. Repeat until convergence

References

Do, C. B., & Batzoglou, S. (2008). What is the expectation maximization algorithm? *Nature biotechnology*, *26*(8), 897–899.

The End