Other application of EM

Marco Tulio Ribeiro

November 12th, 2013

Layout

EM for binomial

Seting

- Source: “What is the expectation maximization algorithm”, Do and Batzoglou

Seting

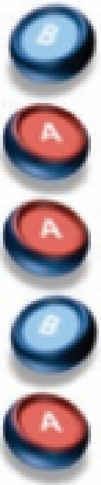
- Source: “What is the expectation maximization algorithm”, Do and Batzoglou
- Setting:
  - Two coins, A and B
  - Land on heads with $P = \theta_A, \theta_B$.

Seting

- Source: “What is the expectation maximization algorithm”, Do and Batzoglou
- Setting:
  - Two coins, A and B
  - Land on heads with $P = \theta_A, \theta_B$.
  - Choose one coin at random (50%), perform 10 tosses, record results

Seting

- Source: “What is the expectation maximization algorithm”, Do and Batzoglou
- Setting:
  - Two coins, A and B
  - Land on heads with $P = \theta_A, \theta_B$.
  - Choose one coin at random (50%), perform 10 tosses, record results
  - Do this five times.



H T T T H H T H T H

H H H H T H H H H H

H T H H H H H T H H

H T H T T T H H T T

T H H H T H H H T H

Formality

- $z = 1$ means coin A was chosen.
- Let $y_i \in Y$ be the number of heads in the sequence.
- For a single sequence of 10 tosses:

$$P(y_i, z|\theta) = \begin{cases} .5 * \binom{10}{y_i} \theta_A^{y_i} (1 - \theta_A)^{10 - y_i} & \text{if } z = 1 \\ .5 * \binom{10}{y_i} \theta_B^{y_i} (1 - \theta_B)^{10 - y_i} & \text{if } z = 0 \end{cases} \tag{1}$$

If we knew z...

- $$L(\theta|Y,z) = \prod_{i=1}^{5} \left(.5 * \binom{10}{y_i} \theta_A^{y_i} (1-\theta_A)^{10-y_i}\right)^{z_i} * \left(.5 * \binom{10}{y_i} \theta_B^{y_i} (1-\theta_B)^{10-y_i}\right)^{1-z_i}$$

If we knew z...

- Ignoring the binomial term:
- $$l(\theta|Y,z) = \sum_{i=1}^{5} z_i(log(.5) + y_i log(\theta_A) + (10 - y_i)log(1 - \theta_A)) + (1 - z_i)(log(.5) + y_i log(\theta_B) + (10 - y_i)log(1 - \theta_B))$$

MLE with we knew z.

- $\hat{\theta}_A =$

MLE with we knew z.

- $\hat{\theta}_A = \frac{\sum_{i=1}^{5} z_i y_i}{10 \sum_{i=1}^{5} z_i}$

MLE with we knew z.

- $\hat{\theta}_A = \frac{\sum_{i=1}^{5} z_i y_i}{10 \sum_{i=1}^{5} z_i}$
- $\hat{\theta}_B =$

MLE with we knew z.

- $\hat{\theta}_A = \frac{\sum_{i=1}^{5} z_i y_i}{10 \sum_{i=1}^{5} z_i}$
- $\hat{\theta}_B = \frac{\sum_{i=1}^{5} (1-z_i) y_i}{10 \sum_{i=1}^{5} (1-z_i)}$

Sure enough

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

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

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

We don't have z, though.

- Marginalizing Z out:

$$L(\theta|Y) = \prod_{i=1}^{5} \sum_{z \in (0,1)} P(y_i, z_i|\theta)$$

We don't have z, though.

- Marginalizing Z out:

$$L(\theta|Y) = \prod_{i=1}^{5} \sum_{z \in (0,1)} P(y_i, z_i|\theta)$$

- Which is equal to:

$$\prod_{i=1}^{5} .5 * (\binom{10}{y_i} \theta_A^{y_i} (1 - \theta_A)^{10-y_i}) +$$
$$.5 * (\binom{10}{y_i} \theta_B^{y_i} (1 - \theta_B)^{10-y_i})$$

Enter EM

- This function is usually hard to optimize

Enter EM

- This function is usually hard to optimize
- So we'll use EM.

Enter EM

- This function is usually hard to optimize
- So we'll use EM.
- Intuition: make z a random variable and take its expected value (given a current θ as truth.
- Then optimize over θ, and repeat.

Let's reason about z

-

$$P(z|Y,\theta) = \prod_{i=1}^{5} P(z_i|y_i,\theta) = \prod_{i=1}^{5} \frac{P(y_i, z_i|\theta)}{P(y_i|\theta)}$$

Let's reason about z

- $$P(z|Y,\theta) = \prod_{i=1}^{5} P(z_i|y_i,\theta) = \prod_{i=1}^{5} \frac{P(y_i, z_i|\theta)}{P(y_i|\theta)}$$
- Since we don't know z, let's leave it as a random variable and take its expected value

Let's reason about z

- $$P(z|Y,\theta) = \prod_{i=1}^{5} P(z_i|y_i,\theta) = \prod_{i=1}^{5} \frac{P(y_i, z_i|\theta)}{P(y_i|\theta)}$$
- Since we don't know z, let's leave it as a random variable and take its expected value
- We'll calculate the expected value of the log-likelihood leaving everything fixed but z. This is step one of EM.

New log likelihood

- $$E[l(\theta|Y,Z)|Y=y,\theta_0] = \sum_{i=1}^{5} E[z_i|Y_i=y_i,\theta_0](log(.5)+y_i log(\theta_A)+(10-y_i)log(1-\theta_A))+ E[(1-z_i)|Y_i=y_i,\theta_0](log(.5)+y_i log(\theta_B)+(10-y_i)log(1-\theta_B))$$
- Note that I substituted z_i for $E[z_i|Y_i=y_i,\theta_0]$.

New log likelihood

- $$E[l(\theta|Y,Z)|Y=y,\theta_0] =$$
$$\sum_{i=1}^{5} E[z_i|Y_i=y_i,\theta_0](log(.5)+y_i log(\theta_A)+(10-y_i)log(1-\theta_A))+$$
$$E[(1-z_i)|Y_i=y_i,\theta_0](log(.5)+y_i log(\theta_B)+(10-y_i)log(1-\theta_B))$$
- Note that I substituted z_i for $E[z_i|Y_i=y_i,\theta_0]$.
- Only z is random, so I was able to push the expectation inside.

What's $E[z_i|Y_i = y_i, \theta_0]$?

- z_i is a binary random variable, so

$$E[z_i|Y_i = y_i, \theta_0] = P(z_i = 1|Y_i = y_i, \theta_0) =$$
$$\frac{P(z_i = 1, Y_i = y_i|\theta_0)}{P(z_i = 1, Y_i = y_i|\theta_0) + P(z_i = 0, Y_i = y_i|\theta_0)} =$$
$$\frac{.5 * \theta_{0A}^{y_i}(1 - \theta_{0A})^{10-y_i}}{.5 * \theta_{0A}^{y_i}(1 - \theta_{0A})^{10-y_i} + .5 * \theta_{0B}^{y_i}(1 - \theta_{0B})^{10-y_i}}$$

- I ommited the binomial terms.

Step one: Expectation

- Since we know $E[z_i|Y_i = y_i, \theta_0]$ we can calculate $E[l(\theta|Y = y, \theta_0)]$ for arbitrary θ values.
- Let's denote:

$$Q(\theta|\theta_0, Y) = E[l(\theta|Y = y, \theta_0)]$$

Step two: maximization

- In step two, we find the value of θ that maximizes $Q(\theta|\theta_0, Y)$
- Let $E[z_i|Y_i = y_i, \theta_0] = E(z_i)$. MLE becomes:
  - $\hat{\theta}_A =$

Step two: maximization

- In step two, we find the value of θ that maximizes $Q(\theta|\theta_0, Y)$
- Let $E[z_i|Y_i = y_i, \theta_0] = E(z_i)$. MLE becomes:
  - $\hat{\theta}_A = \frac{\sum_{i=1}^{5} E(z_i) y_i}{10 \sum_{i=1}^{5} E(z_i)}$

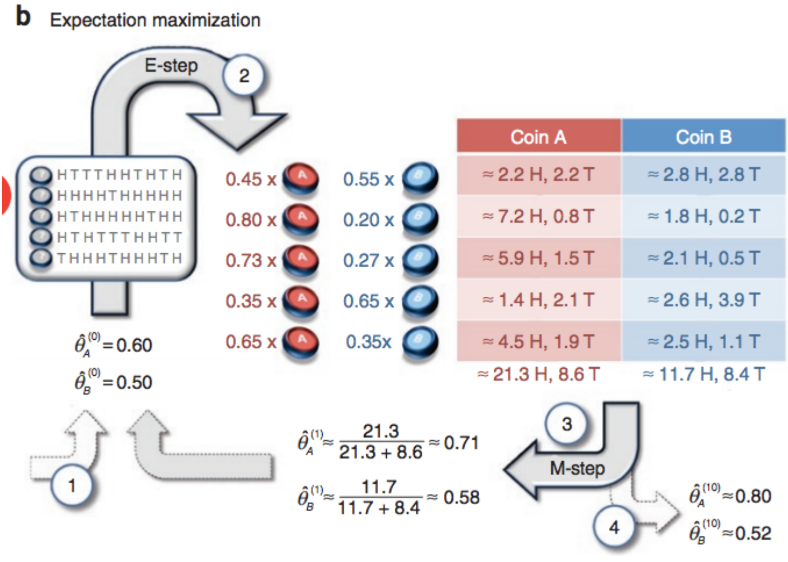
Step two: maximization

- In step two, we find the value of θ that maximizes $Q(\theta|\theta_0, Y)$
- Let $E[z_i|Y_i = y_i, \theta_0] = E(z_i)$. MLE becomes:
  - $\hat{\theta}_A = \frac{\sum_{i=1}^{5} E(z_i) y_i}{10 \sum_{i=1}^{5} E(z_i)}$
  - $\hat{\theta}_B =$

Step two: maximization

- In step two, we find the value of θ that maximizes $Q(\theta|\theta_0, Y)$
- Let $E[z_i|Y_i = y_i, \theta_0] = E(z_i)$. MLE becomes:
  - $\hat{\theta}_A = \frac{\sum_{i=1}^{5} E(z_i) y_i}{10 \sum_{i=1}^{5} E(z_i)}$
  - $\hat{\theta}_B = \frac{\sum_{i=1}^{5} (1-E(z_i)) y_i}{10 \sum_{i=1}^{5} (1-E(z_i))}$

Going back to the example



Doing one iteration

- They begin with $\theta_A = .6, \theta_B = .5$

Doing one iteration

- They begin with $\theta_A = .6, \theta_B = .5$
- $Y = (5, 9, 8, 4, 7)$

Doing one iteration

- They begin with $\theta_A = .6, \theta_B = .5$
- $Y = (5, 9, 8, 4, 7)$
- This yieds $E[z|Y, \theta_0] = (0.44914893, 0.80498552, 0.73346716, 0.35215613, 0.64721512)$

Doing one iteration

- They begin with $\theta_A = .6, \theta_B = .5$
- $Y = (5, 9, 8, 4, 7)$
- This yieds $E[z|Y, \theta_0] = (0.44914893, 0.80498552, 0.73346716, 0.35215613, 0.64721512)$
- What does this mean?

Python code

```
def main():
  Y = np.array([5, 9, 8, 4, 7])
  theta_hat = np.array([.6, .5])
  previous = theta_hat.copy()
  pi = np.array([.5, .5])
  while True:
    # E-step
    pzi1 = pi[0] * theta_hat[0] ** Y * (1 - theta_hat[0]) ** (10 - Y)
    pzi0 = pi[1] * theta_hat[1] ** Y * (1 - theta_hat[1]) ** (10 - Y)
    ezk = pzi1 / (pzi0 + pzi1)
    # M - step
    theta_hat[0] = sum(ezk * Y) / (10 * sum(ezk))
    theta_hat[1] = sum((1 - ezk) * Y) / (10 * sum((1 - ezk)))
    # print ezk
    # print theta_hat
    if (theta_hat == previous).all():
      break
    previous = theta_hat.copy()
  print theta_hat
```

- Output: $\hat{\theta}_A = 0.79678907, \hat{\theta}_B = 0.51958312$

What happens if π is not $(.5, .5)$?

- $$E[l(\theta|Y,Z)|Y=y,\theta_0] = \sum_{i=1}^{5} E[z_i|Y_i=y_i,\theta_0](log(\pi_0)+y_i log(\theta_A)+(10-y_i)log(1-\theta_A))+ E[(1-z_i)|Y_i=y_i,\theta_0](log(1-\pi_0)+y_i log(\theta_B)+(10-y_i)log(1-\theta_B))$$

What happens if π is not $(.5, .5)$?

- $$E[l(\theta|Y,Z)|Y=y,\theta_0] =$$
$$\sum_{i=1}^{5} E[z_i|Y_i=y_i,\theta_0](log(\pi_0)+y_i log(\theta_A)+(10-y_i)log(1-\theta_A))+$$
$$E[(1-z_i)|Y_i=y_i,\theta_0](log(1-\pi_0)+y_i log(\theta_B)+(10-y_i)log(1-\theta_B))$$

- $$E[z_i|Y_i=y_i,\theta_0] = P(z_i=1|Y_i=y_i,\theta_0) =$$
$$\frac{\pi_0 * \theta_{0A}^{y_i}(1-\theta_{0A})^{10-y_i}}{\pi_0 * \theta_{0A}^{y_i}(1-\theta_{0A})^{10-y_i} + (1-\pi_0) * \theta_{0B}^{y_i}(1-\theta_{0B})^{10-y_i}}$$

Reasoning about π

- What's the MLE for π_0 when we know the labels z?

Reasoning about π

- What's the MLE for π_0 when we know the labels z?
- $\hat{\pi}_0 = \frac{\sum_{i=1}^{5} z_i}{5}$

Reasoning about π

- What's the MLE for π_0 when we know the labels z?
- $\hat{\pi}_0 = \frac{\sum_{i=1}^{5} z_i}{5}$
- And when we don't know the labels?

Reasoning about π

- What's the MLE for π_0 when we know the labels z?
- $\hat{\pi}_0 = \frac{\sum_{i=1}^{5} z_i}{5}$
- And when we don't know the labels?
- $\hat{\pi}_0 = \frac{\sum_{i=1}^{5} E(z_i)}{5}$

Reasoning about π

- What's the MLE for π_0 when we know the labels z?
- $\hat{\pi}_0 = \frac{\sum_{i=1}^{5} z_i}{5}$
- And when we don't know the labels?
- $\hat{\pi}_0 = \frac{\sum_{i=1}^{5} E(z_i)}{5}$