# Pencil-and-paper execution of EM on coin toss example deadline: 4pm Mon Oct 14

October 2, 2019

This assignment concerns the scenario considered in the slides, where a coin $Z$ is tossed to choose between one of two other coins $A$ and $B$. The chosen coin is then tossed $N$ times. The whole procedure is repeated $D$ times.From the complete data set it would be easy to determine the probabilities of $Z$ indicating $A$, and for each of $A$ and $B$, their probabilities of turning up heads. The *hidden variable* variant is where as data you just know on each trial what the $N$ coin tosses yielded with the chosen coin, but you don't know which coin was being tossed: the outcome on $Z$ is *hidden*.

Suppose the following data set, where there are just 2 trials, and the chosen coin is tossed just 2 times (this was considered in the slides):

| $d$ | Z | tosses of chosen coin | |
|---|---|---|---|
| 1 | ? | H | H |
| 2 | ? | T | T |

Below there is a detailed working through of a first iteration of the EM procedure for estimating parameters applied to this. **For the assignment you have to give a similar working through of the second iteration.**

Suppose the following notation

| | |
|---|---|
| $\theta_a$ | $P(Z = a)$ |
| $\theta_b$ | $P(Z = b)$ |
| $\theta_{h\|a}$ | $P(h\|a)$ ie. the head prob of coin $A$ |
| $\theta_{t\|a}$ | $P(t\|a)$ ie. the tail prob of coin $A$ |
| $\theta_{h\|b}$ | $P(h\|b)$ ie. the head prob of coin $B$ |
| $\theta_{t\|b}$ | $P(t\|b)$ ie. the tail prob of coin $B$ |
| $\#(d, h)$ | num of heads in trial $d$ |
| $\#(d, t)$ | num of tails in trial $d$ |

If $\boldsymbol{X}^d$ is a particular trial – ie. outcomes of the $N$ tosses of a chosen coin – the probability of the version where the chosen coin was $A$ is given by

$$
\begin{aligned}
P(Z = a, \boldsymbol{X}^d) &= P(Z = a) \times P(h|a)^{\#(d,h)} \times P(t|a)^{\#(d,t)} \\
&= \theta_a \times \theta_{h|a}^{\#(d,h)} \times \theta_{t|a}^{\#(d,t)}
\end{aligned}
$$

and likewise the probability of the version where the chosen coin was $B$ is given by

$$
\begin{aligned}
P(Z = b, \boldsymbol{X}^d) &= P(Z = b) \times P(h|b)^{\#(d,h)} \times P(t|b)^{\#(d,t)} \\
&= \theta_b \times \theta_{h|b}^{\#(d,h)} \times \theta_{t|b}^{\#(d,t)}
\end{aligned}
$$

From these joint probability formula can work out the *conditional probabilities* for the hidden variable:

$$\begin{aligned} P(Z=a|\boldsymbol{X}^d) &= \frac{P(Z=a,\boldsymbol{X}^d)}{\sum_c P(Z=c,\boldsymbol{X}^d)} \\ P(Z=b|\boldsymbol{X}^d) &= \frac{P(Z=b,\boldsymbol{X}^d)}{\sum_c P(Z=c,\boldsymbol{X}^d)} \end{aligned}$$

In the slides we used the notation $\gamma_d(Z)$ for this, where $d$ is index of the data item.

On the particular data set at hand the joint probability formulae are particularly simple

$$\begin{aligned} P(Z=a,\boldsymbol{X}^1) &= \theta_a \times \theta_{h|a}^2 \\ P(Z=b,\boldsymbol{X}^1) &= \theta_b \times \theta_{h|b}^2 \\ P(Z=a,\boldsymbol{X}^2) &= \theta_a \times \theta_{t|a}^2 \\ P(Z=b,\boldsymbol{X}^2) &= \theta_b \times \theta_{t|b}^2 \end{aligned}$$

and thus the conditional probalities are:

$$\begin{aligned} \gamma_1(a) &= \frac{\theta_a \times \theta_{h|a}^2}{\theta_a \times \theta_{h|a}^2 + \theta_b \times \theta_{h|b}^2} \\ \gamma_1(b) &= \frac{\theta_b \times \theta_{h|b}^2}{\theta_a \times \theta_{h|a}^2 + \theta_b \times \theta_{h|b}^2} \\ \gamma_2(a) &= \frac{\theta_a \times \theta_{t|a}^2}{\theta_a \times \theta_{t|a}^2 + \theta_b \times \theta_{t|b}^2} \\ \gamma_2(b) &= \frac{\theta_b \times \theta_{t|b}^2}{\theta_a \times \theta_{t|a}^2 + \theta_b \times \theta_{t|b}^2} \end{aligned}$$

To carry out an EM estimation of the parameters given the data we need some initial setting of the parameters. We will suppose this is:

$\theta_a = \frac{1}{2}$, $\theta_b = \frac{1}{2}$,
$\theta_{h|a} = \frac{3}{4}$, $\theta_{t|a} = \frac{1}{4}$
$\theta_{h|b} = \frac{1}{2}$, $\theta_{t|b} = \frac{1}{2}$

**ITERATION 1**

For each piece of data have to first compute the conditional probabilities of the hidden variable given the data:

$d = 1 : p(Z = A, HH) = 0.5 \times 0.75 \times 0.75 = 0.28125$
$d = 1 : p(Z = B, HH) = 0.5 \times 0.5 \times 0.5 = 0.125$
$d = 1 :\rightarrow sum = 0.40625$
$d = 1 :\rightarrow \gamma_1(A) = 0.692308$
$d = 1 :\rightarrow \gamma_1(B) = 0.307692$
$d = 2 : p(Z = A, TT) = 0.5 \times 0.25 \times 0.25 = 0.03125$
$d = 2 : p(Z = B, TT) = 0.5 \times 0.5 \times 0.5 = 0.125$
$d = 2 :\rightarrow sum = 0.15625$

$d = 2 :\rightarrow \gamma_2(A) = 0.2$
$d = 2 :\rightarrow \gamma_2(B) = 0.8$

Armed with these $\gamma$ values we now treat each data item $\boldsymbol{X}^d$ as if it splits into two versions, one filling out $Z$ as $a$, and with 'count' $\gamma_d(a)$, and one filling out $Z$ as $b$, and with 'count' $\gamma_d(b)$.

We then go through this virtual corpus accumulating counts of certain kinds of event. For events of hidden variable being $Z = a$ and $Z = b$ we get

$E(A) = \gamma_1(a) + \gamma_2(a) = 0.692308 + 0.2 = 0.892308$
$E(B) = \gamma_1(b) + \gamma_2(b) = 0.307692 + 0.8 = 1.10769$

Then we need to go through the $Z = a$ cases and count types of coin toss, and likewise for $Z = b$ cases

$E(A, H) = \sum_d \gamma_d(a)\#(d, h) = (0.692308 \times 2 + 0.2 \times 0) = 1.38462$
$E(A, T) = \sum_d \gamma_d(a)\#(d, t) = (0.692308 \times 0 + 0.2 \times 2) = 0.4$
$E(B, H) = \sum_d \gamma_d(b)\#(d, h) = (0.307692 \times 2 + 0.8 \times 0) = 0.615385$
$E(B, T) = \sum_d \gamma_d(b)\#(d, t) = (0.307692 \times 0 + 0.8 \times 2) = 1.6$

Then from these 'expected' counts we re-estimate parameters

$est(\theta_a) = E(A)/2 = 0.892308/2 = 0.446154$
$est(\theta_b) = E(B)/2 = 1.10769/2 = 0.553846$
$est(\theta_{h|a}) = E(A, H)/\sum_X[E(A, X)] = 1.38462/(1.38462 + 0.4) = 1.38462/1.78462 = 0.775862$
$est(\theta_{t|a}) = E(A, T)/\sum_X[E(A, X)] = 0.4/(1.38462 + 0.4) = 0.4/1.78462 = 0.224138$
$est(\theta_{h|b}) = E(B, H)/\sum_X[E(B, X)] = 0.615385/(0.615385{+}1.6) = 0.615385/2.21538 = 0.277778$
$est(\theta_{t|b}) = E(B, T)/\sum_X[E(B, X)] = 1.6/(0.615385 + 1.6) = 1.6/2.21538 = 0.722222$

Note the denominator 2 in the re-estimation formula for $\theta_a$. We could have written the denominator as $E(A) + E(B)$, but this is $\sum_d \gamma_d(a) + \sum_d \gamma_d(b) = \sum_d[\gamma_d(a) + \gamma_d(b)] = \sum_d[1] = 2$