Pencil-and-paper execution of EM on coin toss example
deadline: Mon 1st Nov 4pm

October 20, 2021

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, and then chosen coin is then tossed N times. The whole procedure is repeated D times. 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 data set to left below, with just 2 trials, and chosen coin tossed just twice (it was considered in the slides), and also suppose the notations to the right (also same as in slides)

d	Z	$\boldsymbol{X}$: tosses of chosen coin	
1	?	H	H
2	?	T	T

θ_a	$P(Z=a)$
θ_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

Below under 'ITERATION 1' there is a detailed working through[1] of a first EM iteration, commencing from particular start values for θ_a, $\theta_{h|a}$, $\theta_{h|b}$.

For the assignment

1. *you have to give a similar working through of the second iteration.*
2. *using quantities already determined in the first part, work out the entire data set probability (a) at the initial parameters (b) at the parameters reached after the first iteration, bearing in mind that the data set probability is:*

$$\prod_d [\sum_Z P(Z, \boldsymbol{X}^d)]$$

Note the *Coin EM Lab* program can be used to show what the parameter values and data probability ought to come out to be.

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

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

[1] The slides contain something similar, though with different start values

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:

$\theta_a=0.4$, $\theta_b=0.6$,

$\theta_{h|a}=0.45$, $\theta_{t|a}=0.55$

$\theta_{h|b}=0.6$, $\theta_{t|b}=0.4$

ITERATION 1

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

$d=1: p(Z=A,HH)=0.4\times0.45\times0.45=0.081$
$d=1: p(Z=B,HH)=0.6\times0.6\times0.6=0.216$

$d = 1 :\to sum = 0.297$
$d = 1 :\to \gamma_1(A) = 0.272727$
$d = 1 :\to \gamma_1(B) = 0.727273$
$d = 2 : p(Z = A, TT) = 0.4 \times 0.55 \times 0.55 = 0.121$
$d = 2 : p(Z = B, TT) = 0.6 \times 0.4 \times 0.4 = 0.096$
$d = 2 :\to sum = 0.217$
$d = 2 :\to \gamma_2(A) = 0.557604$
$d = 2 :\to \gamma_2(B) = 0.442396$

Armed with these γ 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

$E(A) = \gamma_1(a) + \gamma_2(a) = 0.272727 + 0.557604 = 0.830331$
$E(B) = \gamma_1(b) + \gamma_2(b) = 0.727273 + 0.442396 = 1.16967$

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

$E(A, H) = \sum_d \gamma_d(a)\#(d, h) = 0.272727 \times 2 + 0.557604 \times 0 = 0.545455$
$E(A, T) = \sum_d \gamma_d(a)\#(d, t) = 0.272727 \times 0 + 0.557604 \times 2 = 1.11521$
$E(B, H) = \sum_d \gamma_d(b)\#(d, h) = 0.727273 \times 2 + 0.442396 \times 0 = 1.45455$
$E(B, T) = \sum_d \gamma_d(b)\#(d, t) = 0.727273 \times 0 + 0.442396 \times 2 = 0.884793$

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

$est(\theta_a) = E(A)/2 = 0.830331/2 = 0.415165$
$est(\theta_b) = E(B)/2 = 1.16967/2 = 0.584835$
$est(\theta_{h|a}) = E(A, H)/\sum_X[E(A, X)] = 0.545455/1.66066 = 0.328456$
$est(\theta_{t|a}) = E(A, T)/\sum_X[E(A, X)] = 1.11521/1.66066 = 0.671544$
$est(\theta_{h|b}) = E(B, H)/\sum_X[E(B, X)] = 1.45455/2.33934 = 0.621777$
$est(\theta_{t|b}) = E(B, T)/\sum_X[E(B, X)] = 0.884793/2.33934 = 0.378223$

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