CSU44062 – Computational Linguistics
Assignment 2

Brandon Dooley - #16327446

<u>Expectation Maximisation (EM)</u>

d	Z	tosses of chosen coin	
1	?	H	H
2	?	T	T

As detailed in the assignment we are concerned with the above scenario where a coin Z is tossed to choose between one of the other two coins A and B. The chosen coin is then tossed N (N=2) times and the outcome is recorded in regards to heads (H) or tails (T). In our given scenario the outcome of coin toss Z is hidden and is therefore the *hidden variable*.

Throughout this working we will use the following notation:

θ_a	$P(Z = a)$
θ_b	$P(Z = b)$
$\theta_{h\|a}$	$P(h \mid a)$ $i.e\ prob\ of\ a\ head\ on\ coin\ A$
$\theta_{t\|a}$	$P(t \mid a)$ $i.e\ prob\ of\ a\ tail\ on\ coin\ A$
$\theta_{h\|b}$	$P(h \mid b)$ $i.e\ prob\ of\ a\ head\ on\ coin\ B$
$\theta_{t\|b}$	$P(t \mid b)$ $i.e\ prob\ of\ a\ tail\ on\ coin\ B$
$\#(d, h)$	num of heads in trial d
$\#(d, t)$	num of tails in trial d
X^d	a given trial d
$\gamma_d(k)$	$P(Z = k \mid X^d)$

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

$\boldsymbol{\theta_a} = 0.5$
$\boldsymbol{\theta_b} = 0.5$
$\boldsymbol{\theta_{h|a}} = 0.75$
$\boldsymbol{\theta_{t|a}} = 0.25$
$\boldsymbol{\theta_{h|b}} = 0.5$
$\boldsymbol{\theta_{t|b}} = 0.5$

Iteration 1

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

$\boldsymbol{d} = \mathbf{1}: P(Z = A, HH) = 0.5 * 0.75 * 0.75 = 0.28125$
$\boldsymbol{d} = \mathbf{1}: P(Z = B, HH) = 0.5 * 0.5 * 0.5 = 0.125$
$\boldsymbol{d} = \mathbf{1}: \rightarrow sum\ = 0.40625$

$$\boldsymbol{d} = \mathbf{1}: \rightarrow \gamma_1(A)\ = \frac{P(Z = A, HH)}{P(Z = A, HH) + P(Z = B, HH)} = \frac{0.28125}{0.40625} = 0.692308$$

$$\boldsymbol{d} = \mathbf{1}: \rightarrow \gamma_1(B)\ = \frac{P(Z = B, HH)}{P(Z = A, HH) + P(Z = B, HH)} = \frac{0.125}{0.40625} = 0.307692$$

$\boldsymbol{d} = \mathbf{2}: P(Z = A, TT) = 0.5 * 0.25 * 0.25 = 0.03125$
$\boldsymbol{d} = \mathbf{2}: P(Z = B, TT) = 0.5 * 0.5 * 0.5 = 0.125$
$\boldsymbol{d} = \mathbf{2}: \rightarrow sum\ = 0.15625$

$$\boldsymbol{d} = \mathbf{2}: \rightarrow \gamma_2(A)\ = \frac{P(Z = A, TT)}{P(Z = A, TT) + P(Z = B, TT)} = \frac{0.03125}{0.15625} = 0.2$$

$$\boldsymbol{d} = \mathbf{2}: \rightarrow \gamma_2(B)\ = \frac{P(Z = B, TT)}{P(Z = A, TT) + P(Z = B, TT)} = \frac{0.125}{0.15625} = 0.8$$

$\boldsymbol{E}(\boldsymbol{A}) = \gamma_1(A)\ +\ \gamma_2(A) = 0.692308 + 0.2 = 0.892308$
$\boldsymbol{E}(\boldsymbol{B}) = \gamma_1(B)\ +\ \gamma_2(B) = 0.307692 + 0.8 = 1.10769$

$$\boldsymbol{E}(\boldsymbol{A}, \boldsymbol{H}) = \sum_d \gamma_d(A) * \#(d, h) = (0.692308 * 2) + (0.2 * 0) = 1.38462$$

$$\boldsymbol{E}(\boldsymbol{A}, \boldsymbol{T}) = \sum_d \gamma_d(A) * \#(d, t) = (0.692308 * 0) + (0.2 * 2) = 0.4$$

$$\boldsymbol{E}(\boldsymbol{B}, \boldsymbol{H}) = \sum_d \gamma_d(B) * \#(d, h) = (0.307692 * 2) + (0.8 * 0) = 0.615385$$

$$\boldsymbol{E}(\boldsymbol{B}, \boldsymbol{T}) = \sum_d \gamma_d(B) * \#(d, t) = (0.307692 * 0) + (0.8 * 2) = 1.6$$

Then from these 'expected' counts we re-estimate our parameters as follows:

$$\boldsymbol{est(\theta_a)} = \frac{E(A)}{E(A)+E(B)} = \frac{0.892308}{2} = 0.446154$$

$$\boldsymbol{est(\theta_b)} = \frac{E(B)}{E(A)+E(B)} = \frac{1.10769}{2} = 0.553846$$

$$\boldsymbol{est\left(\theta_{h|a}\right)} = \frac{E(A,H)}{\sum_X[E(A,X)]} = \frac{E(A,H)}{E(A,H) + E(A,T)} = \frac{1.38462}{1.38462 + 0.4} = 0.775862$$

$$\boldsymbol{est\left(\theta_{t|a}\right)} = \frac{E(A,T)}{\sum_X[E(A,X)]} = \frac{E(A,T)}{E(A,H) + E(A,T)} = \frac{0.4}{1.38462 + 0.4} = 0.224138$$

$$\boldsymbol{est\left(\theta_{h|b}\right)} = \frac{E(B,H)}{\sum_X[E(A,X)]} = \frac{E(B,H)}{E(B,H) + E(B,T)} = \frac{0.615385}{0.615385 + 1.6} = 0.277778$$

$$\boldsymbol{est\left(\theta_{t|b}\right)} = \frac{E(B,T)}{\sum_X[E(A,X)]} = \frac{E(B,T)}{E(B,H) + E(B,T)} = \frac{1.6}{0.615385 + 1.6} = 0.722222$$

Iteration 2

Using our new estimations for our parameters we repeat again for a second iteration with the following values:

$\boldsymbol{\theta_a} = 0.446154$
$\boldsymbol{\theta_b} = 0.553846$
$\boldsymbol{\theta_{h|a}} = 0.775862$
$\boldsymbol{\theta_{t|a}} = 0.224138$
$\boldsymbol{\theta_{h|b}} = 0.277778$
$\boldsymbol{\theta_{t|b}} = 0.722222$

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

$\boldsymbol{d} = \mathbf{1}: P(Z = A, HH) = 0.446154 * 0.775862 * 0.775862 = 0.268568$
$\boldsymbol{d} = \mathbf{1}: P(Z = B, HH) = 0.553846 * 0.277778 * 0.277778 = 0.042735$
$\boldsymbol{d} = \mathbf{1}: \rightarrow sum \ = 0.311303$

$$\boldsymbol{d} = \mathbf{1}: \rightarrow \gamma_1(A) \ = \frac{P(Z = A, HH)}{P(Z = A, HH) + P(Z = B, HH)} = \frac{0.268568}{0.311303} = 0.862722$$

$$\boldsymbol{d} = \mathbf{1}: \rightarrow \gamma_1(B) \ = \frac{P(Z = B, HH)}{P(Z = A, HH) + P(Z = B, HH)} = \frac{0.042735}{0.311303} = 0.137278$$

$\boldsymbol{d} = \mathbf{2}: P(Z = A, TT) = 0.446154 * 0.224138 * 0.224138 = 0.022414$
$\boldsymbol{d} = \mathbf{2}: P(Z = B, TT) = 0.553846 * 0.722222 * 0.722222 = 0.288889$
$\boldsymbol{d} = \mathbf{2}: \rightarrow sum \ = 0.311303$

$$\boldsymbol{d} = \mathbf{2}: \rightarrow \gamma_2(A) \ = \frac{P(Z = A, TT)}{P(Z = A, TT) + P(Z = B, TT)} = \frac{0.022414}{0.311303} = 0.072001$$

$$\boldsymbol{d} = \mathbf{2}: \rightarrow \gamma_2(B) \ = \frac{P(Z = B, TT)}{P(Z = A, TT) + P(Z = B, TT)} = \frac{0.288889}{0.311303} = 0.927999$$

$\boldsymbol{E(A)} = \ \gamma_1(A) \ + \ \gamma_2(A) = 0.862722 + 0.072001 = 0.934723$
$\boldsymbol{E(B)} = \ \gamma_1(B) \ + \ \gamma_2(B) = 0.137278 + 0.927999 = 1.065277$

$$\boldsymbol{E(A,H)} = \sum_d \gamma_d(A) * \#(d,h) = (0.862722 * 2) + (0.072001 * 0) = 1.725444$$

$$\boldsymbol{E(A,T)} = \sum_d \gamma_d(A) * \#(d,t) = (0.862722 * 0) + (0.072001 * 2) = 0.144002$$

$$\boldsymbol{E(B,H)} = \sum_d \gamma_d(B) * \#(d,h) = (0.137278 * 2) + (0.927999 * 0) = 0.274556$$

$$\boldsymbol{E(B,T)} = \sum_d \gamma_d(B) * \#(d,t) = (0.137278 * 0) + (0.927999 * 2) = 1.855998$$

Then from these 'expected' counts we re-estimate our parameters as follows:

$$\boldsymbol{est(\theta_a)} = \frac{E(A)}{E(A) + E(B)} = \frac{0.934723}{2} = 0.4673615$$

$$\boldsymbol{est(\theta_b)} = \frac{E(B)}{E(A) + E(B)} = \frac{1.065277}{2} = 0.5326385$$

$$\boldsymbol{est\left(\theta_{h|a}\right)} = \frac{E(A,H)}{\sum_X [E(A,X)]} = \frac{E(A,H)}{E(A,H) + E(A,T)} = \frac{1.725444}{1.869446} = 0.922971$$

$$\boldsymbol{est\left(\theta_{t|a}\right)} = \frac{E(A,T)}{\sum_X [E(A,X)]} = \frac{E(A,T)}{E(A,H) + E(A,T)} = \frac{0.144002}{1.869446} = 0.077029$$

$$\boldsymbol{est\left(\theta_{h|b}\right)} = \frac{E(B,H)}{\sum_X [E(A,X)]} = \frac{E(B,H)}{E(B,H) + E(B,T)} = \frac{0.274556}{2.130554} = 0.128866$$

$$\boldsymbol{est\left(\theta_{t|b}\right)} = \frac{E(B,T)}{\sum_X [E(A,X)]} = \frac{E(B,T)}{E(B,H) + E(B,T)} = \frac{1.855998}{2.130554} = 0.871134$$

Thus at the end of the second iteration our new estimates for the parameters are:

$\boldsymbol{\theta_a} = 0.4673615$
$\boldsymbol{\theta_b} = 0.5326385$
$\boldsymbol{\theta_{h|a}} = 0.922971$
$\boldsymbol{\theta_{t|a}} = 0.077029$
$\boldsymbol{\theta_{h|b}} = 0.128866$
$\boldsymbol{\theta_{t|b}} = 0.871134$