DD2434 Assignment 1A

Vianey Darsel

December 12, 2022

1 Dependency in a Directed Graphical Model

Question 1.1.1

The paths $Z_{d,n} \leftarrow \Theta_d \rightarrow Z_{d,n+1}$ and $W_{d,n} \leftarrow \beta_{1:K} \rightarrow Z_{d,n+1}$ are d-separated by $\{\Theta_d, \beta_{1:K}\}$. So, $W_{d,n}$ and $W_{d,n+1}$ are d-separated by $\{\Theta_d, \beta_{1:K}\}$. So, $W_{d,n} \perp W_{d,n+1} | \Theta_d, \beta_{1:K}$

Question 1.1.2

The path $\Theta_d \leftarrow \alpha \rightarrow \Theta_{d+1}$ is not d-separated by $\{Z_{d,1:N}\}$. So, Θ_d and Θ_{d+1} are not d-separated by $\{Z_{d,1:N}\}$. So, $\Theta_d \not\perp \Theta_{d+1} | Z_{d,1:N}$.

Question 1.1.3

The path $\Theta_k \leftarrow \alpha \rightarrow \Theta_j$ is d-separated by $\{\alpha; Z_{1:D,1:N}\}$. So, Θ_d and Θ_{d+1} are d-separated by $\{\alpha; Z_{1:D,1:N}\}$. So, $\Theta_d \perp \Theta_{d+1} | alpha; Z_{1:D,1:N}$.

Question 1.1.4

$W_{d,n} \not\perp W_{d,n+1} | \Lambda_d, \beta_{1:K}$

Question 1.1.5

$\Theta_d \not\perp \Theta_{d+1} | Z_{d,1:N}, Z_{d+1,1:N}$

Question 1.1.6

$\Lambda_d \not\perp \Lambda_{d+1} | \Phi, Z_{1:D,1:N}$

2 Likelihood of a Tree Graphical Mode

Question 1.2.7:

To use our code, you just need to run the file "run_file.py" in the 1A_2 folder. In this file, you can change the file to use by changing the value of the index of the list file. The function that we have implemented are object functions so they are in the file "Tree.py". The different methods added to the provided files are :

- Node.isLeaf() : return is the node is a leaf or not
- Node.isRoot() : return is the node is the root or not
- Node.Likelihood($\beta_{1:L}$) : return an array with the likelihood of the observation given the value of the parent node (if any) for all the possible values of the parent. Otherwise, it returns the global likelihood.
- Tree.Likelihood($\beta_{1:L}$) : return the likelihood of an observation (theta is included in the tree)

To implement the algorithm, we have used the following relation :

$$\mathbb{P}(\beta_{1:L}) = \sum_{k\in[K]} \mathbb{P}(\beta_{1:L}|X_{root} = i)\mathbb{P}(X_{root} = i) \tag{1}$$

We have reduced the size of the problem because $\mathbb{P}(\beta_{1:L}|X_{root} = i)$ is the same problem with one less depth. To start from the bottom. We can write :

$$\begin{aligned}
\mathbb{P}(\beta_{1:L}|X_{root} = i) &= \sum_{k_i\in[K]^{|\{X_i\in CHILDREN_{X_{root}}\}|}} \mathbb{P}\left(\bigcap_{X_i\in CHILDREN_{X_{root}}} (X_i = k_i), \beta_{1:L}|X_{root} = i\right) \\
&= \sum_{k_i\in[K]^{|\{X_i\in CHILDREN_{X_{root}}\}|}} \mathbb{P}\left(\bigcap_{X_i\in CHILDREN_{X_{root}}} \left(X_i = k_i, \beta_{u(X_i)}\right)|X_{root} = i\right) \\
&= \sum_{k_i\in[K]^{|\{X_i\in CHILDREN_{X_{root}}\}|}} \prod_{X_i\in CHILDREN_{X_{root}}} \mathbb{P}(X_i = k_i, \beta_{u(X_i)}|X_{root} = i) \\
&= \sum_{k_i\in[K]^{|\{X_i\in CHILDREN_{X_{root}}\}|}} \prod_{X_i\in CHILDREN_{X_{root}}} \mathbb{P}(\beta_{u(X_i)}|X_i = k_i)\mathbb{P}(X_i = k_i|X_{root} = i) \\
&= \sum_{k_i\in[K]^{|\{X_i\in CHILDREN_{X_{root}}\}|}} \prod_{X_i\in CHILDREN_{X_{root}}} \alpha_i(X_i = k_i)\mathbb{P}(X_i = k_i|X_{root} = i)
\end{aligned} \tag{2}$$

where :

- $\beta_{u(X_i)}$ are the values of the leaves of the tree generated by X_i.
- $\alpha_i(X_i = k_i) = \mathbb{P}(\beta_{u(X_i)}|X_i = k_i)$ and :
  $\alpha_i(X_i = x_i) = \sum_{\text{values of children of } X_i} \prod_{X_j\in textchildren} \alpha_j\mathbb{P}(X_j = x_j|X_i = k_i)$

Is is possible due to the structure of the tree. A β belongs to the tree generated by only one child. We have passed from depth d to depth $d-1$. Thus we can compute the likelihood by starting from the leaves to the root.

Question 1.2.8:

The results we have obtained with the given data can be read in Table 1

File	Small Tree	Medium Tree	Large Tree
Sample 0	0.0270	7.06e-18	2.04e-67
Sample 1	0.121	2.18e-18	3.12e-65
Sample 2	0.021	7.83e-19	4.50e-66
Sample 3	0.00763	1.99e-18	6.80e-68
Sample 4	0.00733	2.27e-17	2.06e-68

Table 1: Results of our algorithm with provided data

3 Simple Variational Inference

Question 1.3.9:

The code is implemented in the file VariationalInference_SupportFunction.py. We have chosen that all the variables converge to leave the loop. The update scheme is the following:

$$\begin{aligned}
a^* =& a + \frac{N+1}{2} \\
b^* =& b + \frac{1}{2}\mathbb{E}\left[\sum_{n=1}^{N}(X_n - \mu)^2 + \lambda(\mu - \mu')^2\right] \\
=& b + \frac{\sum_{n=1}^{N} X_n^2 + \lambda\mu'^2 - 2\mu^*\left(\sum_{n=1}^{N} X_n + \lambda\mu'\right) + (N+\lambda)\left(\frac{1}{\tau^*} + (\mu^*)^2\right)}{2} \\
\tau^* =& \mathbb{E}[\tau](N+\lambda) \\
=& \frac{a^*}{b^*}(N+\lambda) \\
\mu^* =& \frac{\lambda\mu' + \sum_{n=1}^{N} x_n}{\lambda + N}
\end{aligned} \tag{3}$$

Question 1.3.10:

$$p(\Theta|\mathcal{D}) = \frac{p(\mathcal{D},\Theta)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\Theta)p(\Theta)}{p(\mathcal{D})} \propto p(\mathcal{D}|\Theta)p(\Theta) \tag{4}$$

So we only need to compute $p(\mathcal{D}|\Theta)p(\Theta)$, where $\mathcal{D} = \mathbf{X}$ and $\Theta = \{\mu;\tau\}$. Since $\mu|\tau \hookrightarrow \mathcal{N}\left(\mu', \frac{1}{\lambda\tau}\right)$ and $\tau \hookrightarrow \mathcal{G}(a,b)$. So we expect to find $\mu|\tau \hookrightarrow \mathcal{N}\left(\mu^*, \frac{1}{\lambda^*\tau}\right)$ and $\tau \hookrightarrow \mathcal{G}(a^*,b^*)$.

$$\begin{aligned}
ln(p(\mu,\tau|\mathbb{X}) &\stackrel{\pm}{=} ln(p(\mathbb{X}|\mu,\tau) + ln(p(\mu,\tau)) \\
&= \sum_{n=1}^{N} ln(p(X_n|\mu,\tau) + ln(p(\mu|\tau)) + ln(p(\tau)) \\
&= \sum_{n=1}^{N}\left[\frac{1}{2}ln\left(\frac{\tau}{2\pi}\right) - \frac{\tau}{2}(X_n-\mu)^2\right] + \frac{1}{2}ln\left(\frac{\lambda\tau}{2\pi}\right) - \frac{\lambda\tau}{2}(\mu-\mu')^2 + ln\left(\frac{b^a}{\Gamma(a)}\right) + (a-1)ln(\tau) - b\tau \\
&\stackrel{\pm}{=} \left[\frac{N}{2}ln(\tau) - \sum_{n=1}^{N}\frac{\tau}{2}(X_n-\mu)^2\right] + \frac{1}{2}ln(\tau) - \frac{\lambda\tau}{2}(\mu^2 - 2\mu\mu' + \mu'^2) + (a-1)ln(\tau) - b\tau \\
&= ln(\tau)\left[\frac{N}{2} + a - 1\right] - \tau\left[\frac{\sum_{n=1}^{N} X_n^2}{2} + b + \frac{\lambda\mu'^2}{2}\right] - \frac{\sum_{n=1}^{N}\tau + \lambda\tau}{2}\mu^2 + \left[\tau\sum_{n=1}^{N} X_n + \lambda\tau\mu'\right]\mu \\
&= ln(\tau)\left[\frac{N}{2} + a - 1\right] - \tau\left[\frac{\sum_{n=1}^{N} X_n^2}{2} + b + \frac{\lambda\mu'^2}{2}\right] - \frac{N\tau + \lambda\tau}{2}\mu^2 + \left[\tau\sum_{n=1}^{N} X_n + \lambda\tau\mu'\right]\mu
\end{aligned} \tag{5}$$

If there is no data ($N = 0$), we only have $ln(p(\Theta))$. So,

$$ln(p(\Theta)) = ln(\tau)\,[a-1] - \tau\left(b + \frac{\lambda\mu'^2}{2}\right) - \frac{\lambda\tau}{2}\mu^2 + \lambda\tau\mu'\mu \tag{6}$$

Thus, we can identify the new parameter.

$$\mu|\tau \hookrightarrow \mathcal{N}\left(\mu^*, \frac{1}{\lambda^*\tau}\right)$$

$$\tau \hookrightarrow \mathcal{G}(a^*, b^*)$$

$$
\begin{aligned}
\text{with } \lambda^* &= N + \lambda \\
\mu^* &= \frac{\sum_{n=1}^{N} X_n + \lambda\mu'}{N+\lambda} \\
a^* &= \frac{N}{2} + a \\
b^* &= \frac{\sum_{n=1}^{N} X_n^2}{2} + b + \frac{\lambda\mu'^2}{2} - \frac{\lambda^*(\mu^*)^2}{2}
\end{aligned}
$$

Question 1.3.11:

The file runfile.py provides examples of model comparisons for various parameters. We have computed the results from the Variational Inference algorithm and from the Exact Posterior for 3 different orders of magnitude:

- $\lambda >> N$
- $\lambda = N$
- $\lambda << N$

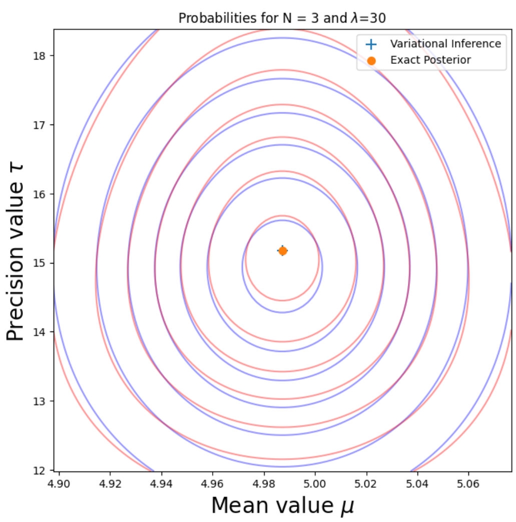

Figure 1: Comparison for $\lambda >> N$

Figure 1 and 2 show that the Variational results are very close from the Posterior. The figures are also similar.

However, Figure 3 presents an even better similarity. The iso-lines are almost overlaid. With an increase of the number of observations, the loss of the Variational techniques is reduced.

Figure 4 shows the difference between the prior distribution and the distribution of the posterior one.

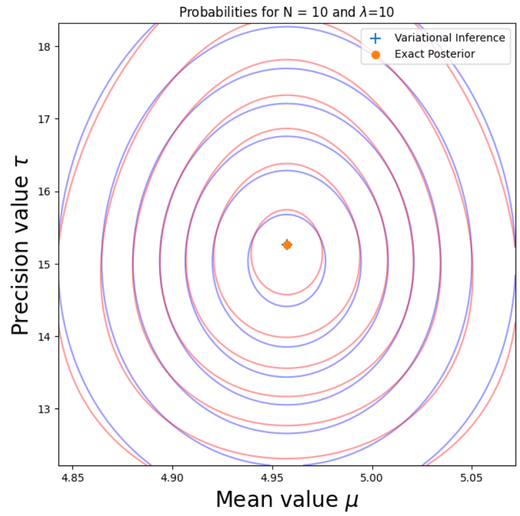

Figure 2: Comparison for $\lambda = N$

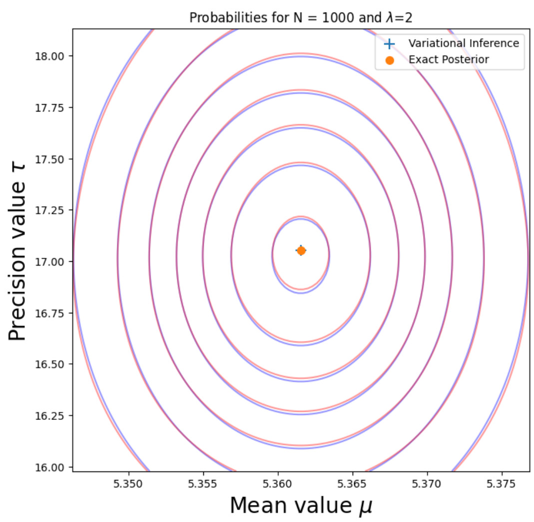

Figure 3: Comparison for $\lambda << N$

4 Super Epicentra - Expectation-Maximization

Question 1.4.12:

In this model, we have :

- $Z^n \hookrightarrow Cat(\pi)$
- $X_1^n|Z^n \hookrightarrow \mathcal{N}(\mu_{1,Z^n}, \frac{1}{\tau_{1,Z^n}})$
- $X_2^n|Z^n \hookrightarrow \mathcal{N}(\mu_{2,Z^n}, \frac{1}{\tau_{2,Z^n}})$
- $S^n|Z^n \hookrightarrow \mathcal{P}(\lambda_{Z^n})$

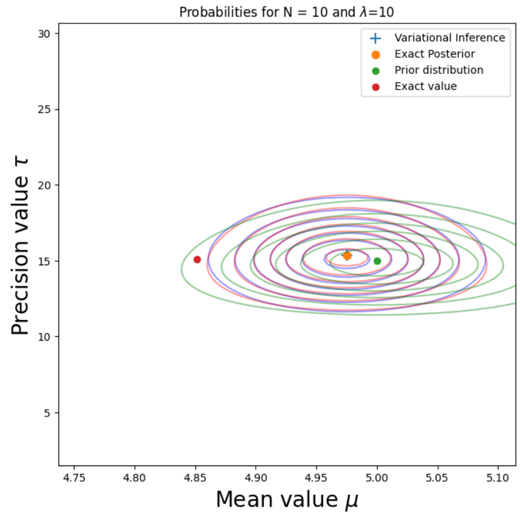

Figure 4: Comparison posterior and prior

Thus, we can compute $p(X_1^n, X_2^n, S^n, Z^n|\mu, \tau, \pi)$ and $\mathbb{P}(Z^n|X_1^n, X_2^n, S^n, \mu, \tau, \pi)$. On the one hand,

$$
\begin{aligned}
p(X_1, X_2, S, Z|\mu, \tau, \lambda, \pi) &= \prod_{n=1}^{N} p(X_1^n, X_2^n, S^n, Z^n|\mu, \tau, \lambda, \pi) \\
&= \prod_{n=1}^{N} p(X_1^n, X_2^n, S^n|Z^n, \mu, \tau, \lambda) p(Z^n|\pi) \\
&= \prod_{n=1}^{N} p(X_1^n|Z^n, \mu_1, \tau_1) p(X_2^n|Z^n, \mu_2, \tau_2) \mathbb{P}(S^n|Z^n, \lambda) p(Z^n|\pi) \\
&= \prod_{n=1}^{N} \mathcal{N}(\mu_{Z^n,1}, \tau_{Z^n,1}) \mathcal{N}(\mu_{Z^n,2}, \tau_{Z^n,2}) \mathcal{P}(S^n|\lambda_{Z^n}) \pi_{Z^n}
\end{aligned}
\tag{7}
$$

We also have

$$
\mathbb{P}(Z|X_1, X_2, S, \mu, \tau, \lambda, \pi) = \prod_{n=1}^{N} \mathbb{P}(Z^n|X_1^n, X_2^n, S^n, \mu, \tau, \lambda, \pi)
$$

$$
\mathbb{P}(Z^n = c|X_1^n, X_2^n, S^n, \mu, \tau, \lambda, \pi) = \frac{p(X_1^n, X_2^n, S^n|Z^n = c, \mu, \lambda, \tau)\mathbb{P}(Z^n = c|\pi)}{\sum_{k=1}^{K} p(X_1^n, X_2^n, S^n, Z^n = k|\mu, \tau, \lambda, \pi)}
$$

From Equation 7, we can deduce that :

$$
\begin{aligned}
\mathbb{P}(Z^n = c|X_1^n, X_2^n, S^n, \mu, \tau, \lambda, \pi) &= \frac{p(X_1^n, X_2^n, S^n|Z^n = c, \mu, \tau, \lambda)\mathbb{P}(Z^n = c|\pi)}{\sum_{k=1}^{K} p(X_1^n, X_2^n, S^n|Z^n = k, \mu, \tau)\mathbb{P}(Z^n = k|\pi)} \\
&= \frac{\mathcal{N}(X_1^n|\mu_{c,1}, \tau_{c,1})\mathcal{N}(X_2^n|\mu_{c,2}, \tau_{c,2})\mathcal{P}(S^n|\lambda_c)\pi_c}{\sum_{k=1}^{K} \mathcal{N}(X_1^n|\mu_{k,1}, \tau_{k,1})\mathcal{N}(X_2^n|\mu_{k,2}, \tau_{k,2})\mathcal{P}(S^n|\lambda_k)\pi_k} \\
A_{n,c,\Theta^{old}} &= \frac{\mathcal{N}(X_1^n|\mu_{c,1}, \tau_{c,1})\mathcal{N}(X_2^n|\mu_{c,2}, \tau_{c,2})\mathcal{P}(S^n|\lambda_c)\pi_c}{\sum_{k=1}^{K} \mathcal{N}(X_1^n|\mu_{k,1}, \tau_{k,1})\mathcal{N}(X_2^n|\mu_{k,2}, \tau_{k,2})\mathcal{P}(S^n|\lambda_k)\pi_k}
\end{aligned}
\tag{8}
$$

From Equation 7 and 8, we can compute $\mathbb{E}_{\mathbb{P}(Z|X_1,X_2,S,\mu^{old},\tau^{old},\lambda^{old},\pi^{old})}\left[ln(p(X_1,X_2,S,Z|\mu,\tau,\lambda,\pi))\right]$.

$$
\begin{aligned}
E &= \mathbb{E}_{\mathbb{P}(Z|X_1,X_2,S,\mu^{old},\tau^{old},\lambda^{old},\pi^{old})}\left[ln(p(X_1,X_2,S,Z|\mu,\tau,\lambda,\pi))\right] \\
&= \mathbb{E}_{\mathbb{P}(Z|X_1,X_2,S,\mu^{old},\tau^{old},\lambda^{old},\pi^{old})}\left[ln\left(\prod_{n=1}^{N}\mathcal{N}(\mu_{Z^n,1},\tau_{Z^n,1})\mathcal{N}(\mu_{Z^n,2},\tau_{Z^n,2})\mathcal{P}(S^n|\lambda_{Z^n})\pi_{Z^n}\right)\right] \\
&= \mathbb{E}_{\mathbb{P}(Z|X_1,X_2,S,\mu^{old},\tau^{old},\lambda^{old},\pi^{old})}\left[\sum_{n=1}^{N} ln\left(\mathcal{N}(\mu_{Z^n,1},\tau_{Z^n,1})\mathcal{N}(\mu_{Z^n,2},\tau_{Z^n,2})\mathcal{P}(S^n|\lambda_{Z^n})\pi_{Z^n}\right)\right] \\
&= \sum_{n=1}^{N}\mathbb{E}_{\mathbb{P}(Z|X_1,X_2,S,\mu^{old},\tau^{old},\lambda^{old},\pi^{old})} ln\left(\mathcal{N}(\mu_{Z^n,1},\tau_{Z^n,1})\mathcal{N}(\mu_{Z^n,2},\tau_{Z^n,2})\mathcal{P}(S^n|\lambda_{Z^n})\pi_{Z^n}\right) \\
&= \sum_{n=1}^{N}\mathbb{E}_{\mathbb{P}(Z^n|X_1^n,X_2^n,S^n,\mu^{old},\tau^{old},\lambda^{old},\pi^{old})} ln\left(\mathcal{N}(\mu_{Z^n,1},\tau_{Z^n,1})\mathcal{N}(\mu_{Z^n,2},\tau_{Z^n,2})\mathcal{P}(S^n|\lambda_{Z^n})\pi_{Z^n}\right) \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K}\mathbb{P}(Z^n=k|X_1^n,X_2^n,S^n,\mu^{old},\tau^{old},\lambda^{old},\pi^{old}) ln\left(\mathcal{N}(\mu_{k,1},\tau_{k,1})\mathcal{N}(\mu_{k,2},\tau_{k,2})\mathcal{P}(S^n|\lambda_k)\pi_k\right) \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K} A_{n,k,\Theta^{old}}\left[ln(\mathcal{N}(\mu_{k,1},\tau_{k,1})) + ln(\mathcal{N}(\mu_{k,2},\tau_{k,2})) + ln(\mathcal{P}(S^n|\lambda_k)) + ln(\pi_k)\right] \\
&\overset{\pm}{=} \sum_{n=1}^{N}\sum_{k=1}^{K} A_{n,k,\Theta^{old}}\left[\frac{1}{2}ln(\tau_{k,1}) - \frac{\tau_{k,1}}{2}(X_1^n-\mu_{k,1})^2 + \frac{1}{2}ln(\tau_{k,2}) - \frac{\tau_{k,2}}{2}(k_2-\mu_{k,2})^2\right. \\
&\left. -\lambda_k + S_n ln(\lambda_k) + ln(\pi_k)\right] \\
&= \sum_{k=1}^{K}\left[\frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}ln(\tau_{k,1}) - \tau_{k,1}\frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}(X_1^n-\mu_{k,1})^2 + \frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}ln(\tau_{k,2})\right. \\
&\left. -\tau_{k,2}\frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}(X_2^n-\mu_{k,2})^2 - \sum_{n=1}^{N}A_{n,k,\Theta^{old}}\lambda_k + \sum_{n=1}^{N}A_{n,k,\Theta^{old}}S_n ln(\lambda_k) + \sum_{n=1}^{N}A_{n,k,\Theta^{old}}ln(\pi_k)\right]
\end{aligned}
\tag{9}
$$

We need to add a constraint on π_k since $\sum_{k=1}^{K}\pi_k = 1$. So we finally have:

$$
\begin{aligned}
E = \sum_{k=1}^{K}&\left[\frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}ln(\tau_{k,1}) - \tau_{k,1}\frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}(X_1^n-\mu_{k,1})^2 + \frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}ln(\tau_{k,2})\right. \\
&-\tau_{k,2}\frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{2}(X_2^n-\mu_{k,2})^2 - \sum_{n=1}^{N}A_{n,k,\Theta^{old}}\lambda_k + \sum_{n=1}^{N}A_{n,k,\Theta^{old}}S_n ln(\lambda_k) \\
&\left.+\sum_{n=1}^{N}A_{n,k,\Theta^{old}}ln(\pi_k)\right] + \beta\left(\sum_{k=1}^{K}\pi_k = 1\right)
\end{aligned}
\tag{10}
$$

Now we have to maximize E (Equation 10) with respect to each parameter.

$$
\begin{aligned}
\frac{\partial E}{\partial \pi_k} &= 0 \\
&= \frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{\pi_k} + \beta \\
\Leftrightarrow \pi_k &= -\frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{\beta}
\end{aligned}
$$

So $\pi_k \propto \sum_{n=1}^{N}A_{n,k,\Theta^{old}}$; so :

$$
\pi_k = \frac{\sum_{n=1}^{N}A_{n,k,\Theta^{old}}}{\sum_{c=1}^{K}\sum_{n=1}^{N}A_{n,c,\Theta^{old}}}
\tag{11}
$$

$$\begin{aligned}
\frac{\partial E}{\partial \mu_{k,1}} &= 0 \\
&= \tau_{k,1} \sum_{n=1}^{N} A_{n,k,\Theta^{old}} \left(X_1^n - \mu_{k,1} \right) \\
\Leftrightarrow \sum_{n=1}^{N} A_{n,k,\Theta^{old}} \mu_{k,1} &= \sum_{n=1}^{N} A_{n,k,\Theta^{old}} X_1^n \\
\Leftrightarrow \qquad\qquad \mu_{k,1} &= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} X_1^n}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial E}{\partial \tau_{k,1}} &= 0 \\
&= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}}{2\tau_{k,1}} - \sum_{n=1}^{N} \frac{A_{n,k,\Theta^{old}}}{2} \left(X_1^n - \mu_{k,1} \right)^2 \\
\Leftrightarrow \tau_{k,1} \left[\sum_{n=1}^{N} A_{n,k,\Theta^{old}} \left(X_1^n - \mu_{k,1} \right)^2 \right] &= \sum_{n=1}^{N} A_{n,k,\Theta^{old}} \\
\Leftrightarrow \qquad\qquad \tau_{k,1} &= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} \left(X_1^n - \mu_{k,1} \right)^2} \\
\Leftrightarrow \qquad\qquad \tau_{k,1} &= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} \left(X_1^n - \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} X_1^n}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}} \right)^2}
\end{aligned}$$

With the symmetry of the roles we have:

$$\begin{aligned}
\mu_{k,1} &= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} X_1^n}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}} \\
\tau_{k,1} &= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} \left(X_1^n - \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} X_1^n}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}} \right)^2} \\
\mu_{k,2} &= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} X_2^n}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}} \\
\tau_{k,2} &= \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} \left(X_2^n - \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} X_2^n}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}} \right)^2}
\end{aligned} \tag{12}$$

$$\begin{aligned}
\frac{\partial E}{\partial \lambda_k} &= 0 \\
&= -\sum_{n=1}^{N} A_{n,k,\Theta^{old}} + \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} S_n}{\lambda_k} \\
\Leftrightarrow \lambda_k \sum_{n=1}^{N} A_{n,k,\Theta^{old}} &= \sum_{n=1}^{N} A_{n,k,\Theta^{old}} S_n
\end{aligned}$$

So :

$$\lambda_k = \frac{\sum_{n=1}^{N} A_{n,k,\Theta^{old}} S_n}{\sum_{n=1}^{N} A_{n,k,\Theta^{old}}} \tag{13}$$

From Equation 11,12 and 13, we can compute the maximization step of the EM algorithm.

Question 1.4.13:

The algorithm is implemented in the functions of the file EMfunctions.

Question 1.4.14:

the algorithm is very sensitive to the initial values and to extreme points. Thus, it happens that a "cluster" k contains only one value. The observations of data in 3 dimensions from Figure 5 seems to show that 3 is a good number for the number of different "clusters". The EM algorithm (initialized

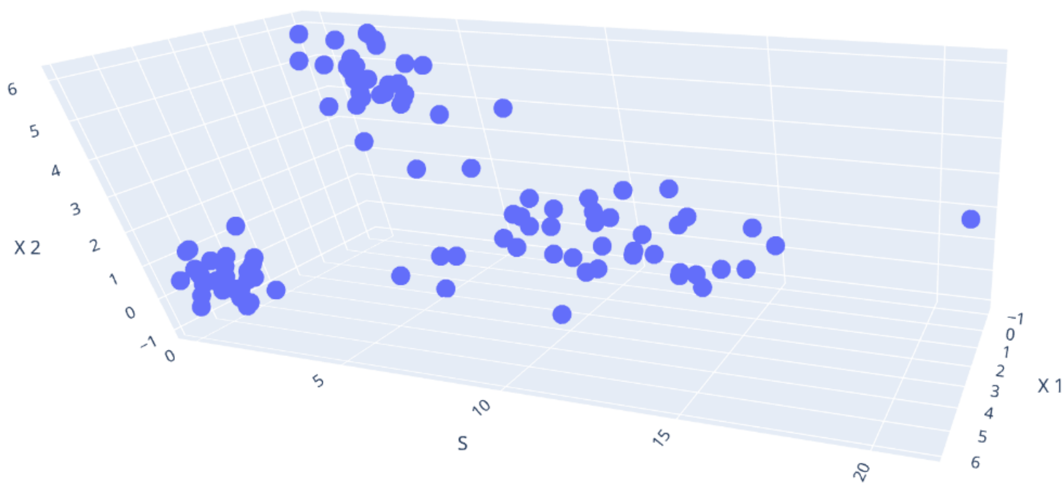

Figure 5: Original distribution of data

with maximal distance values) gives the expected output (Figure 6). However, with a larger number of clusters, it detects individual behaviours, as Figure 7. The BIC value could be a good tools to get the optimal number of distributions k. But, our calculations have detected that the probability of the individual points are extremely high (due to the precision of the normal distributions which is very high), so the tool seems to look for more distributions than it seems to really exist.

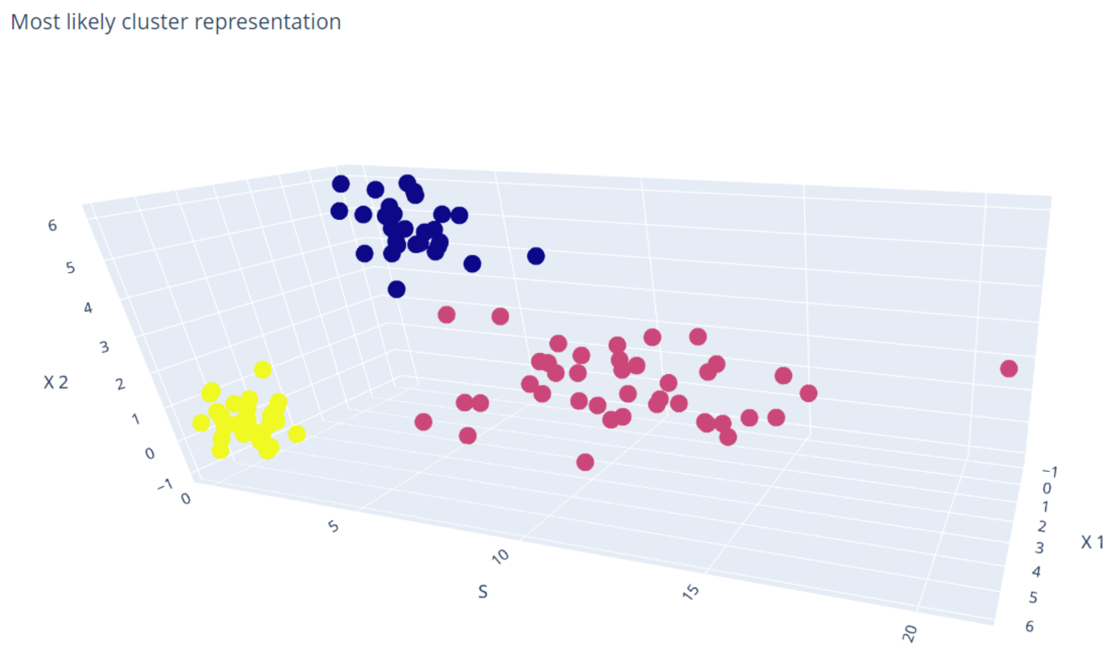

Figure 6: Distribution of the clusters for $k = 3$

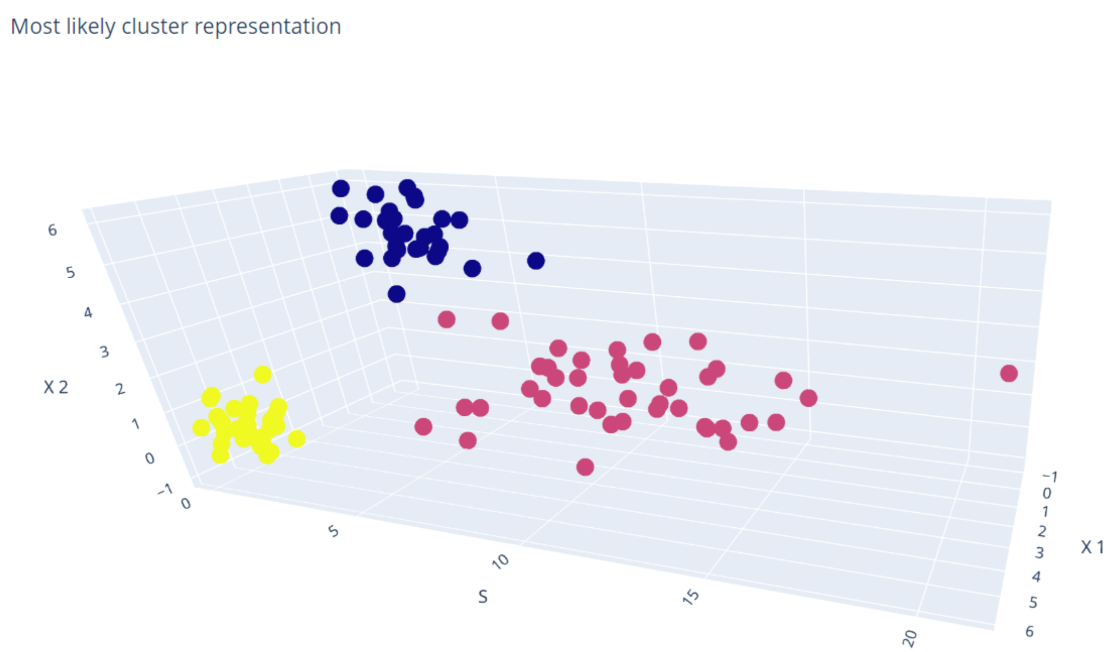

Figure 7: Distribution of the clusters for $k = 5$