Mini Project 3

Rohith D, Siddharth DP, Jay Shah, Aravint Annamalai, Sayan Mitra

April 22, 2021

1 Expressions for estimated values of various parameters

In this mini-project, we are estimating the values of the parameters of two Gaussian distributions and a Bernoulli random variable using Expectation Maximisation algorithm. The values of various parameters are as follows:

$$\hat{\mu_1} = \frac{\sum_{i=1}^{N}(1-r_i)x_i}{\sum_{i=1}^{N}(1-r_i)}$$

$$\hat{\sigma_1^2} = \frac{\sum_{i=1}^{N}(1-r_i)(x_i-\hat{\mu_1})^2}{\sum_{i=1}^{N}(1-r_i)}$$

$$\hat{\mu_2} = \frac{\sum_{i=1}^{N} r_i x_i}{\sum_{i=1}^{N} r_i}$$

$$\hat{\sigma_2^2} = \frac{\sum_{i=1}^{N} r_i(x_i-\hat{\mu_2})^2}{\sum_{i=1}^{N} r_i}$$

$$\hat{\pi} = \frac{1}{n}\sum r_i$$

2 Question1

2.1 Table of values

INIT	PARAM	TRUE VAL	N=10	N=1000	N=10000
[0.5,0.1,0.8,0.9,0.3]	μ_1	0.0	0.1384	-0.0423	0.00313
	μ_2	1.0	0.9058	0.9810	1.0039
	σ_1	0.8	0.4465	0.8358	0.7928
	σ_2	0.4	0.4306	0.3885	0.3914
	π	0.5	0.5	0.5	0.5
	no. of iter	-	47	29	32
[0.5,0.1,0.8,0.9,0.3]	μ_1	0.0	-0.9134	-0.4705	-0.3984
	μ_2	1.0	0.5193	0.5795	0.5984
	σ_1	0.8	0.0753	0.6522	0.6821
	σ_2	0.4	0.8558	0.6017	0.6443
	π	0.1	0.5	0.5	0.5
	no. of iter	-	14	71	32
[0.1,0.1,0.8,0.9,0.3]	μ_1	0.0	-0.3411	-0.0377	-0.0154
	μ_2	1.0	0.6185	1.0295	0.9692
	σ_1	0.8	0.7762	0.8016	0.8039
	σ_2	0.4	0.085	0.3235	0.4136
	π	0.1	0.1	0.1	0.1
	no. of iter	-	13	63	95

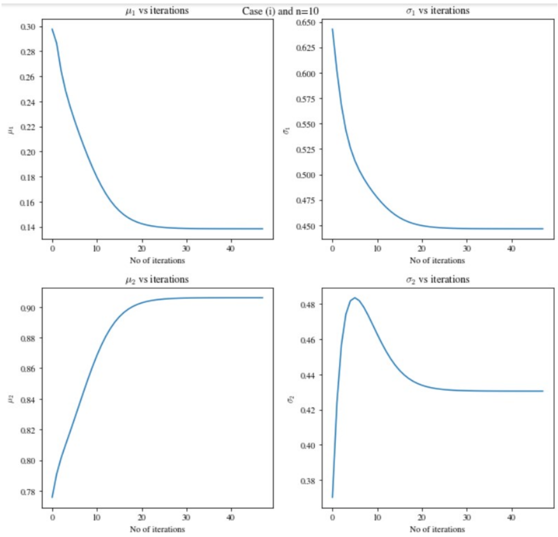
These are the true and the estimated values of various parameters which were obtained during a simulation.

2.2 Plots and inferences for Case (a)

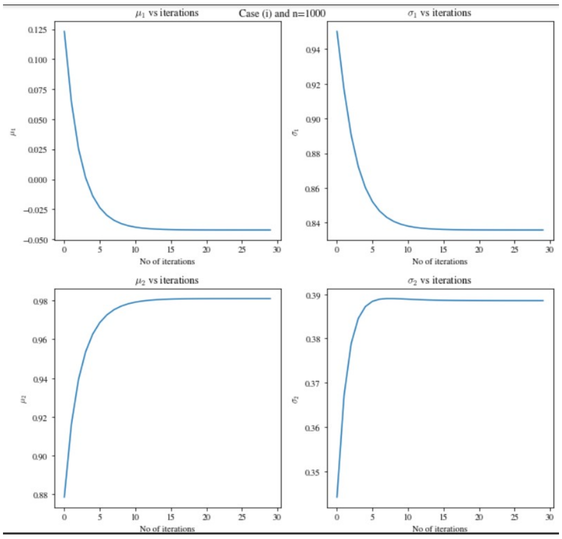
The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.0, \mu_2 = 1.0, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.5$$
$$\textbf{Initialized values} : \mu_1 = 0.1, \mu_2 = 0.8, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.5$$

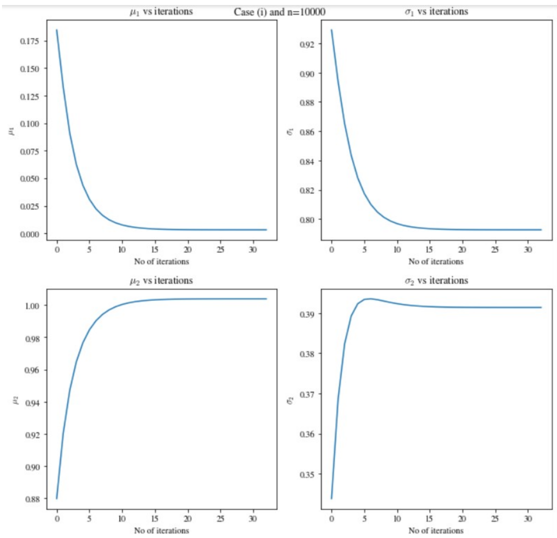
The learning curves are plotted below:



(a) N=10



(b) N=1000



(c) N=10000

Figure 1: Learning curves for Case-(a)

This is a really nice case where the values of various parameters are initialized close to the true values. Hence, the algorithm faces no issues whatsoever to converge to the true value of the parameter, given sufficient number of samples and number of iterations.

As the number of samples increases, the algorithm converges more towards the true value, and the difference between the true and the predicted value decreases gradually, which is what is expected. Even the learning becomes smoother as the number of samples increase.

2.3 Plots and inferences for Case (b)

The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.0, \mu_2 = 1.0, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.1$$
$$\textbf{Initialized values} : \mu_1 = 0.1, \mu_2 = 0.8, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.5$$

The learning curves are plotted in the next page.

This is a very peculiar case. In Experiment-1, the value of π is fixed and it is not allowed to be updated. But here, the true value of π is 0.1, but the initialized value of π is 0.5, which is not allowed to change as well. The value of π affects the whole distribution by a massive extent.

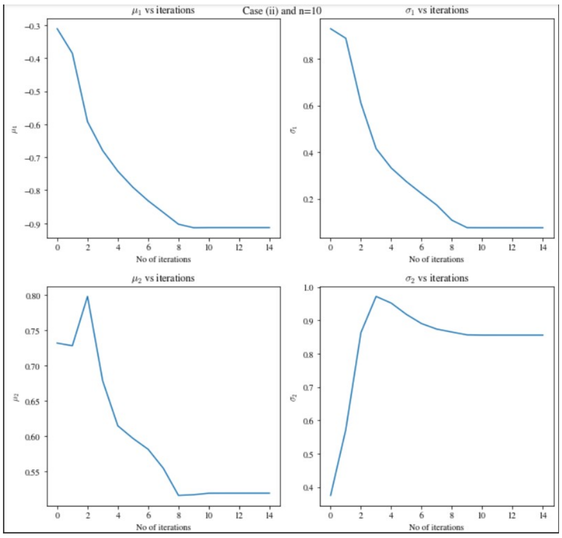
The EM algorithm tries to fit the distribution to the new value of π, but fails, even when the number of samples generated are high. The values of various parameters which are estimated, isn't even close to the actual value, even when a large number of samples are taken. This can be clearly observed in the table above.

2.4 Plots and inferences for Case (c)

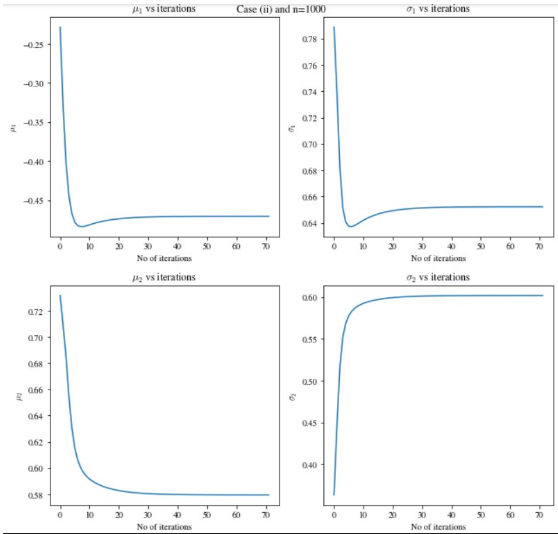
The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.0, \mu_2 = 1.0, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.1$$
$$\textbf{Initialized values} : \mu_1 = 0.1, \mu_2 = 0.8, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.1$$

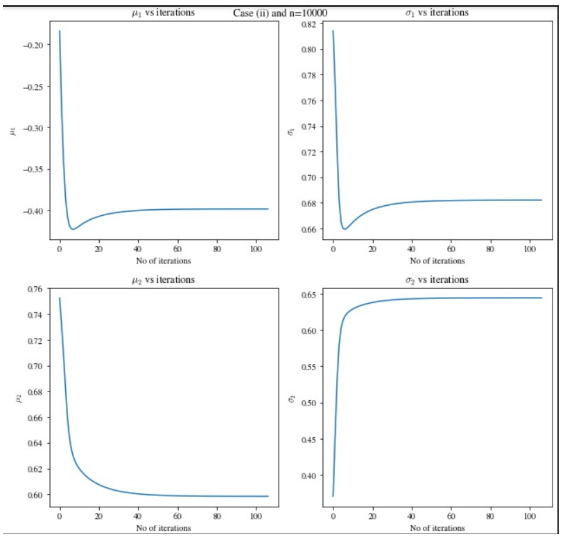
The problem due to mismatched values of π is absent here. This is now similar to Case (a) and similar convergence is observed. The learning curves for this are plotted in Page-6.



(a) N=10

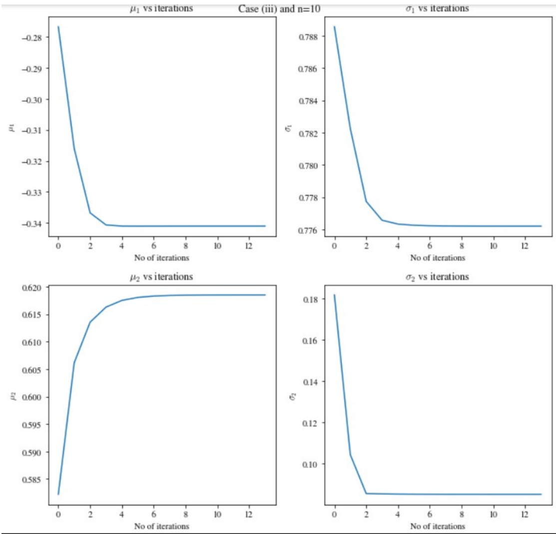


(b) N=1000

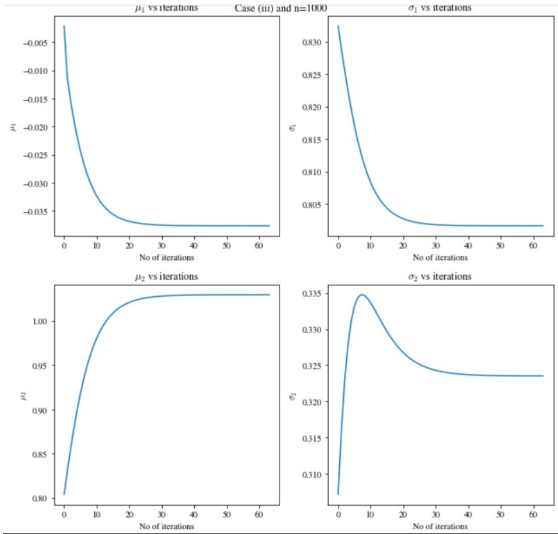


(c) N=10000

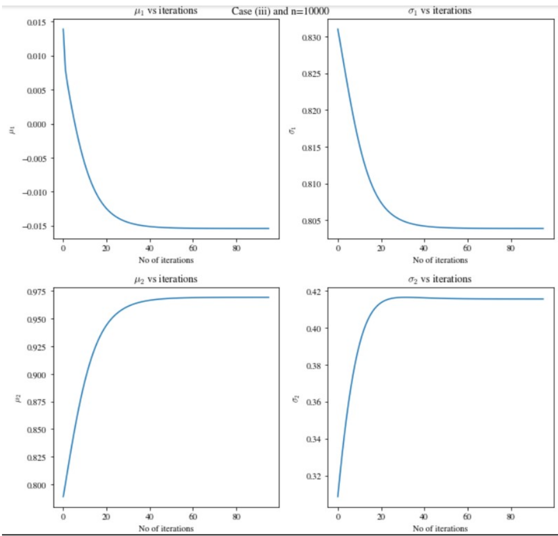
Figure 2: Learning curves for Case-(b)



(a) N=10



(b) N=1000



(c) N=10000

Figure 3: Learning curves for Case-(c)

3 Question2

3.1 Table of values

INIT	PARAM	TRUE VAL	N=10	N=1000	N=10000
[0.45,0.1,0.8,0.9,0.3]	π	0.5	0.29	0.46	0.51
	μ_1	0.0	0.25	0.04	-0.06
	μ_2	1.0	1.18	1.01	1.009
	σ_1	0.8	0.35	0.8	0.785
	σ_2	0.4	0.112	0.39	0.396
	no. of iter	-	37	173	257
[0.45,0.8,0.1,0.3,0.9]	π	0.5	0.53	0.34	0.516
	μ_1	0.0	0.915	0.96	1.023
	μ_2	1.0	0.028	-0.211	0.035
	σ_1	0.8	0.125	0.41	0.389
	σ_2	0.4	0.33	0.683	0.813
	no. of iter	-	31	447	235
[0.45,0.1,0.8,0.9,0.3]	π	0.1	0.8	0.136	0.097
	μ_1	0.0	1.79	-0.006	-0.004
	μ_2	1.0	-0.26	0.847	0.99
	σ_1	0.8	0.136	0.787	0.818
	σ_2	0.4	0.516	0.420	0.4205
	no. of iter	-	52	599	1016

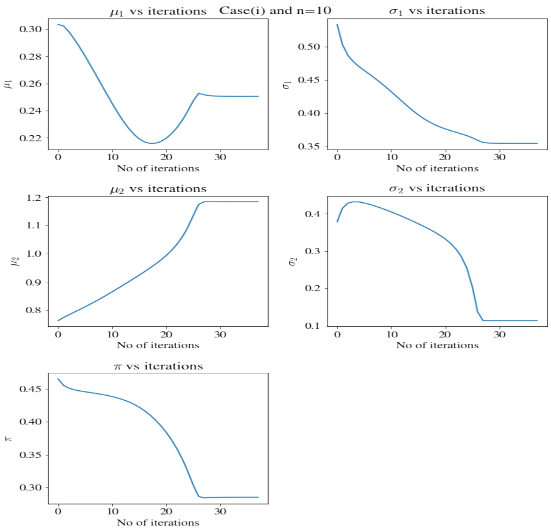
These are the true and the estimated values of various parameters which were obtained during a simulation.

3.2 Plots and inferences for Case (a)

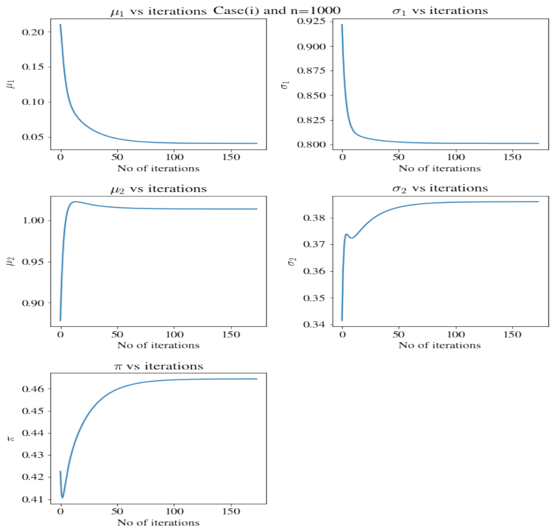
The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.0, \mu_2 = 1.0, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.5$$
$$\textbf{Initialized values} : \mu_1 = 0.1, \mu_2 = 0.8, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.45$$

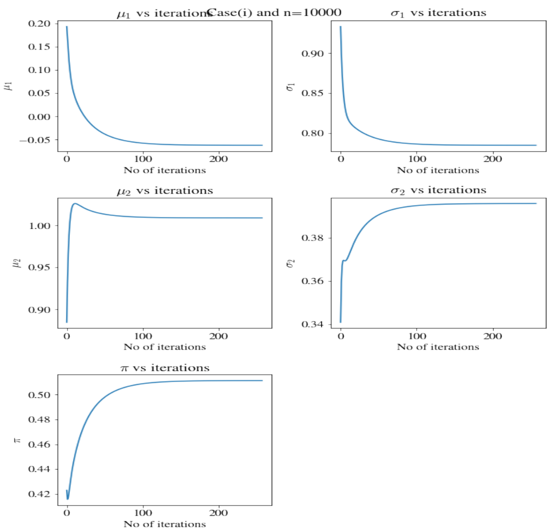
The learning curves are plotted below:



(a) N=10



(b) N=1000



(c) N=10000

Figure 4: Learning curves for Case-(a)

Here, the initialized values are quite close to the true value, and hence the EM algorithm faces no issues whatsoever to converge closer to the actual value. As the number of samples, N, increases, the value of the parameters converge closer to the true value, but it takes more iterations to converge.

3.3 Plots and inferences for Case (b)

The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.0, \mu_2 = 1.0, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.5$$
$$\textbf{Initialized values} : \mu_1 = 0.8, \mu_2 = 0.1, \sigma_1 = 0.3, \sigma_2 = 0.9, \pi = 0.45$$

The learning curves are plotted in the next page.

Here, we can see that the priors μ_1 and μ_2, σ_1 and σ_2 are swapped. As a result, the initialized value of the parameters are quite far away from the actual value. As we see in the table, μ_1 seems to converge to 1.0, which is the true value of μ_2 and vice-versa. Similarly, σ_1 seems to converge to 0.4, which is the true value of σ_2 and vice-versa.

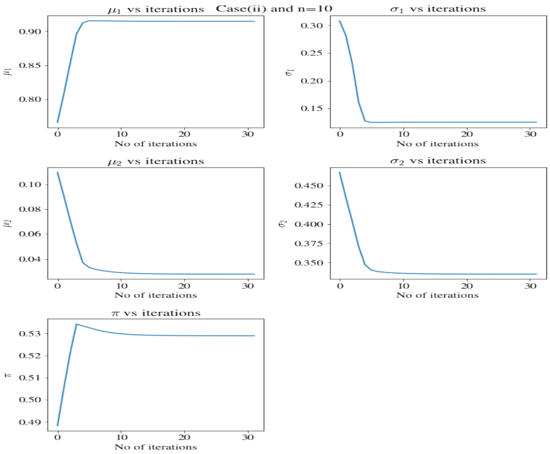
This is because, there might be multiple local maximas while estimating the values of various parameters. If the initialized values are closer to the true value, the only maxima in the vicinity would be the global maxima, which explains why there is no problem of convergence for Case-(a). But in this case, since the initialized values are quite far from the actual value, it reaches a local maxima and not the global maxima.

3.4 Plots and inferences for Case (c)

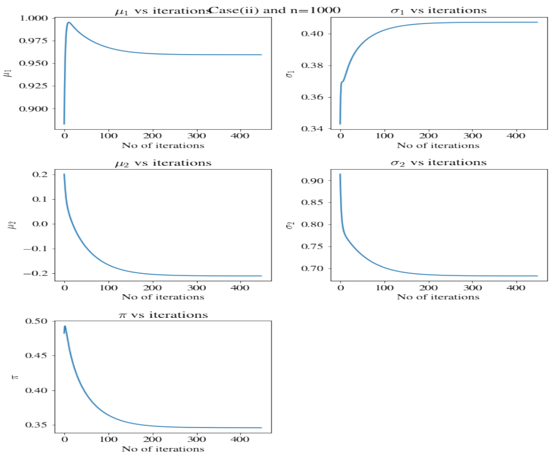
The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.0, \mu_2 = 1.0, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.1$$
$$\textbf{Initialized values} : \mu_1 = 0.1, \mu_2 = 0.8, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.45$$

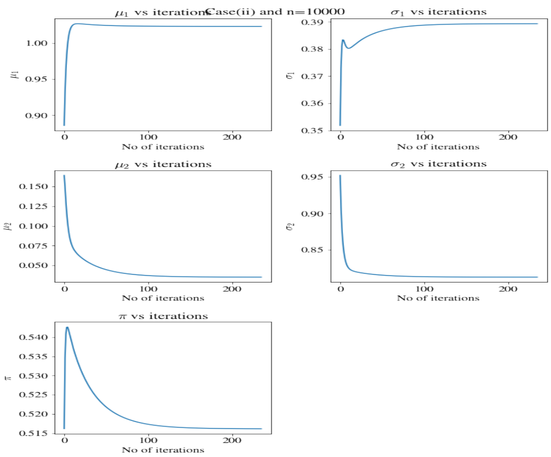
Here, the initialized value of π is far from the actual value, whereas all the other parameters are initialized closer to the actual value. From the table, we can observe that the values of all the variables converge to their actual values, but take a significantly higher number of iterations to do the same.



(a) N=10

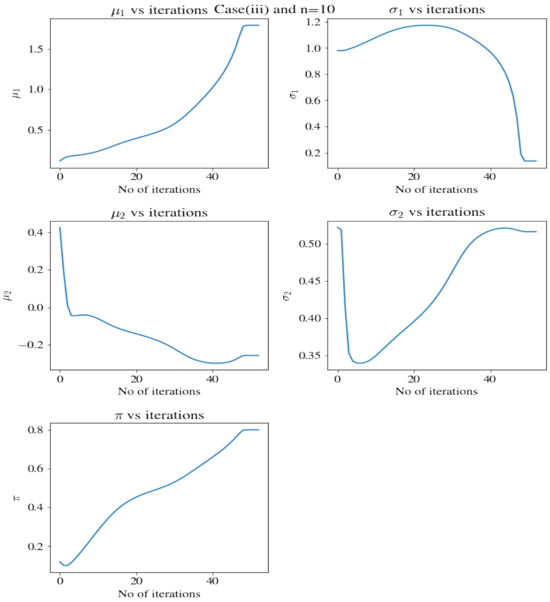


(b) N=1000

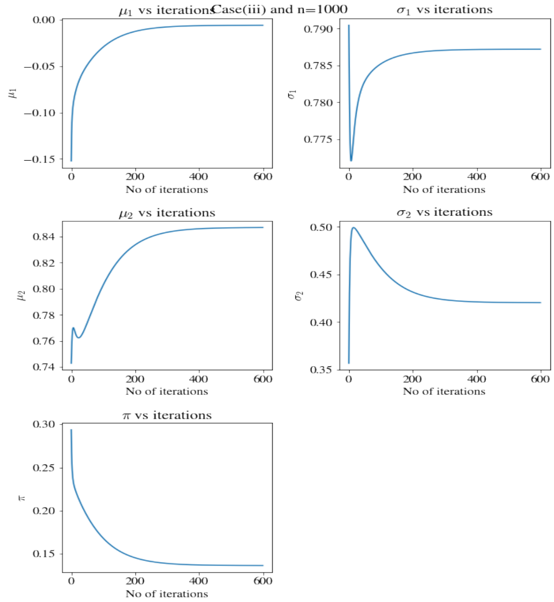


(c) N=10000

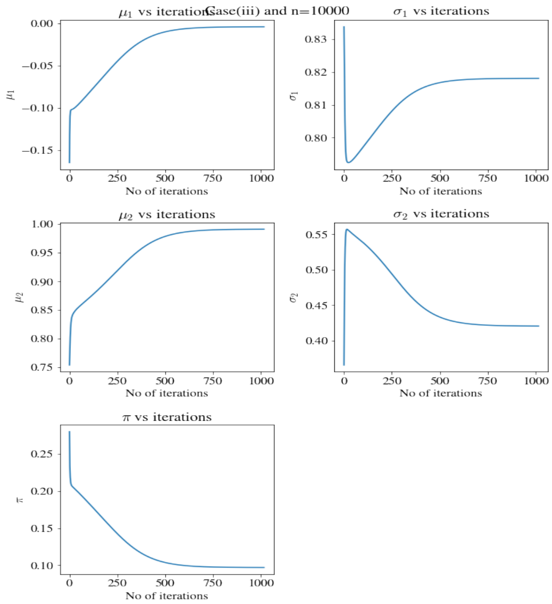
Figure 5: Learning curves for Case-(b)



(a) N=10



(b) N=1000



(c) N=10000

Figure 6: Learning curves for Case-(c)

4 Question3

4.1 Table of values

INIT	PARAM	TRUE VAL	N=5000
[0.45,0.5,0.5,0.9,0.3]	π	0.5	0.5242
	μ_1	0.0	-0.0722
	μ_2	1.0	0.9992
	σ_1	0.8	0.7817
	σ_2	0.4	0.3983
[0.45,0.5,0.5,0.9,0.3]	π	0.5	0.528
	μ_1	0.2	0.1577
	μ_2	0.8	0.7901
	σ_1	0.8	0.7936
	σ_2	0.4	0.4274
[0.45,0.5,0.5,0.9,0.3]	π	0.5	0.5118
	μ_1	0.4	0.3701
	μ_2	0.6	0.5965
	σ_1	0.8	0.7857
	σ_2	0.4	0.4058
[0.45,0.5,0.5,0.9,0.3]	π	0.5	0.5038
	μ_1	0.48	0.4936
	μ_2	0.52	0.5041
	σ_1	0.8	0.7949
	σ_2	0.4	0.3871

4.2 Plots for Case (i)

The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values}: \mu_1 = 0.0, \mu_2 = 1.0, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.5$$
$$\textbf{Initialized values}: \mu_1 = 0.5, \mu_2 = 0.5, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.45$$

The learning curve is plotted below:

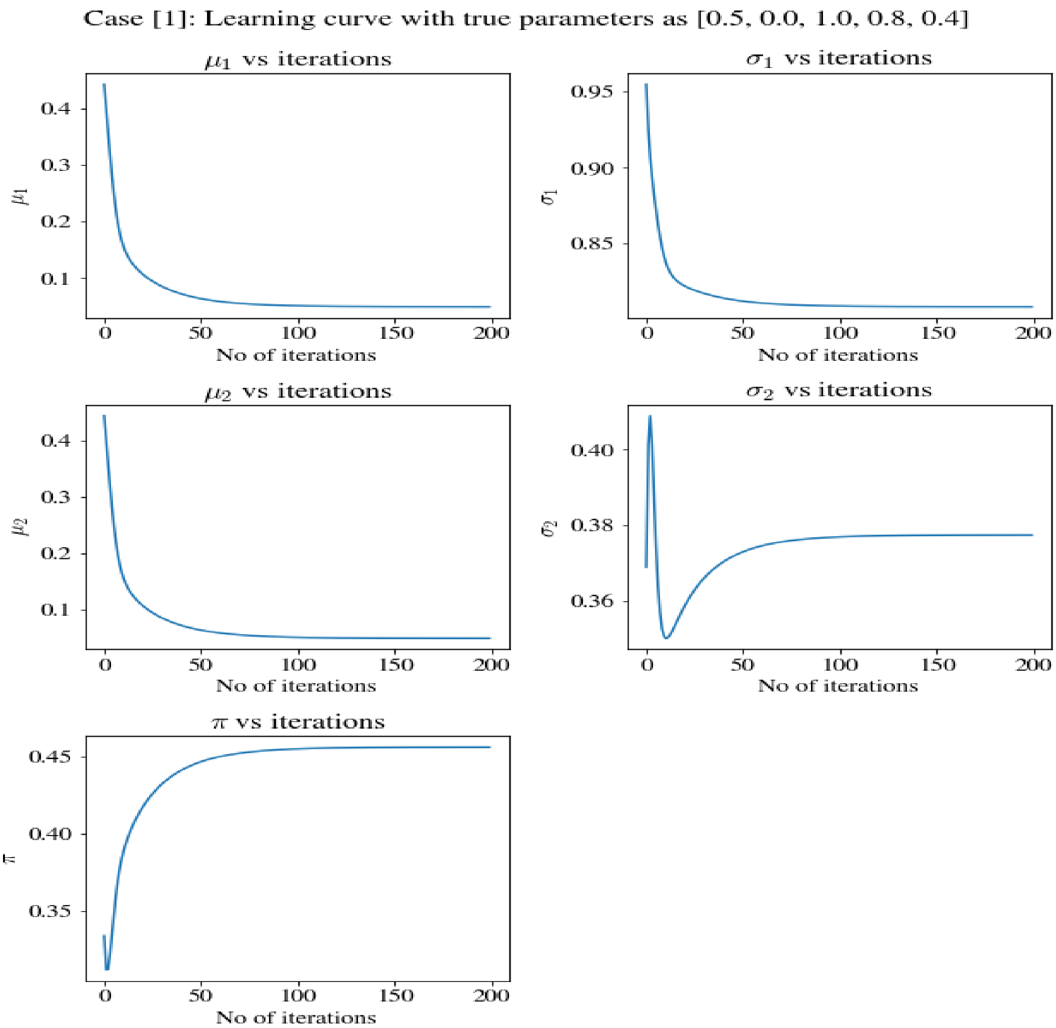


Figure 7: Learning curves for Case-(i)

4.3 Plots for Case (ii)

The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.2, \mu_2 = 0.8, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.5$$
$$\textbf{Initialized values} : \mu_1 = 0.5, \mu_2 = 0.5, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.45$$

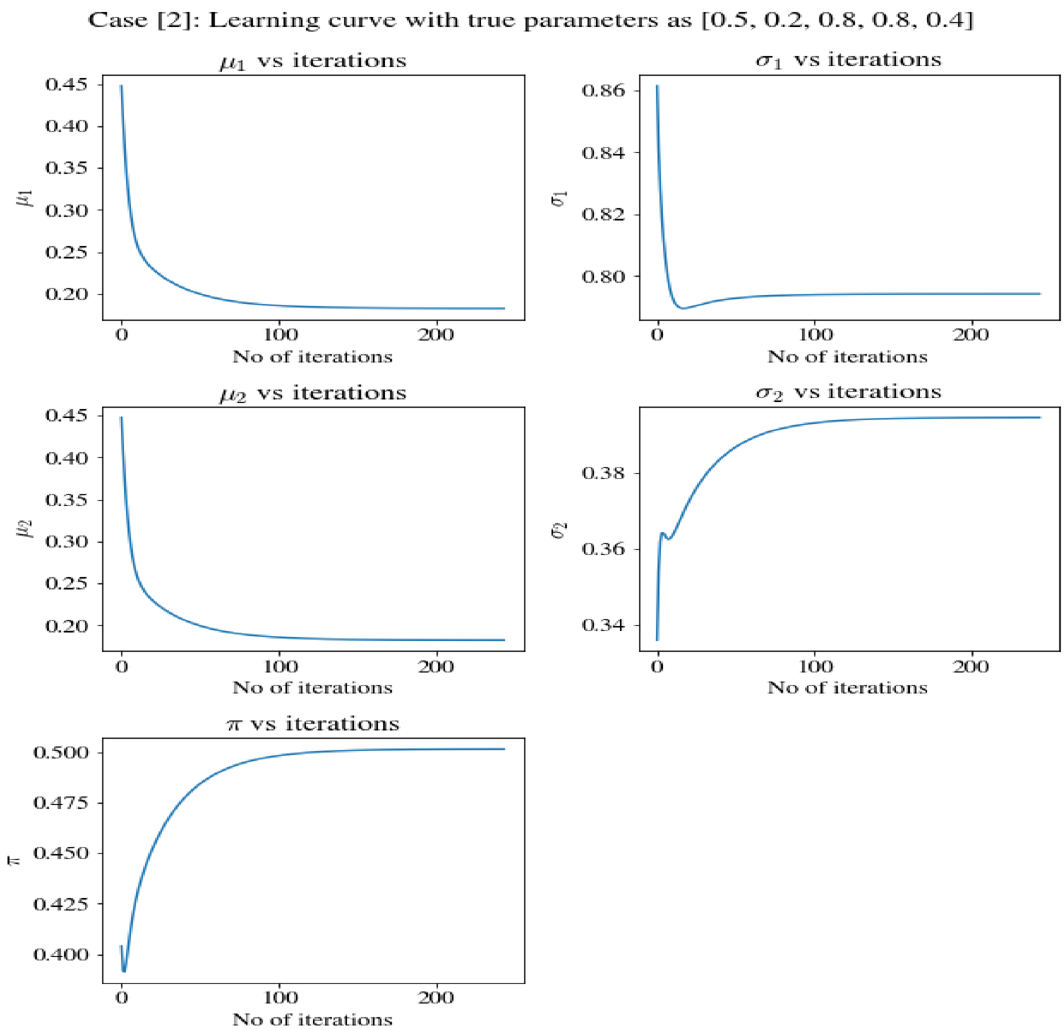


Figure 8: Learning curves for Case-(ii)

4.4 Plots for Case (iii)

The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values} : \mu_1 = 0.4, \mu_2 = 0.6, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.5$$
$$\textbf{Initialized values} : \mu_1 = 0.5, \mu_2 = 0.5, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.45$$

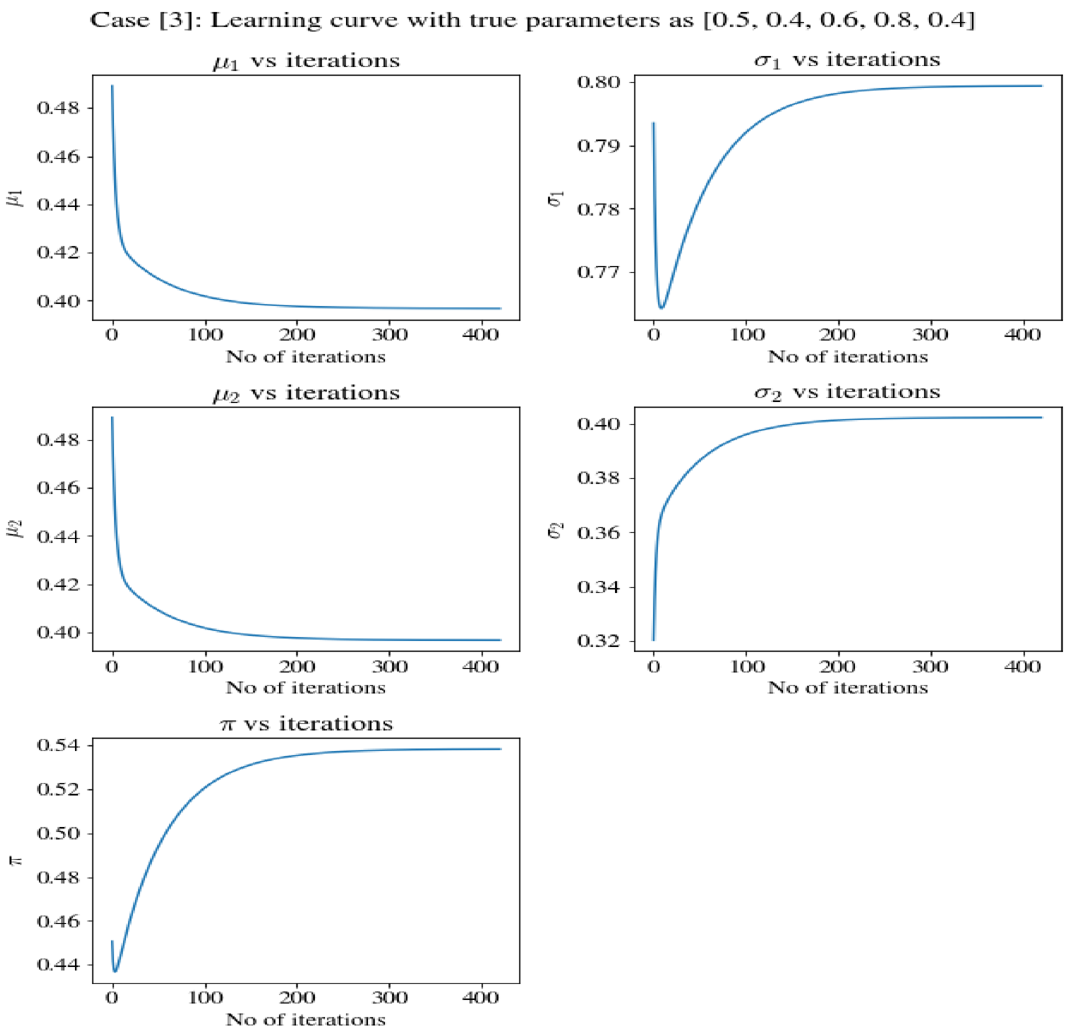


Figure 9: Learning curves for Case-(iii)

4.5 Plots and inferences for Case (iv)

The true values and the initialized values of the various parameters in this case are as follows:

$$\textbf{True values}: \mu_1 = 0.48, \mu_2 = 0.52, \sigma_1 = 0.8, \sigma_2 = 0.4, \pi = 0.5$$
$$\textbf{Initialized values}: \mu_1 = 0.5, \mu_2 = 0.5, \sigma_1 = 0.9, \sigma_2 = 0.3, \pi = 0.45$$

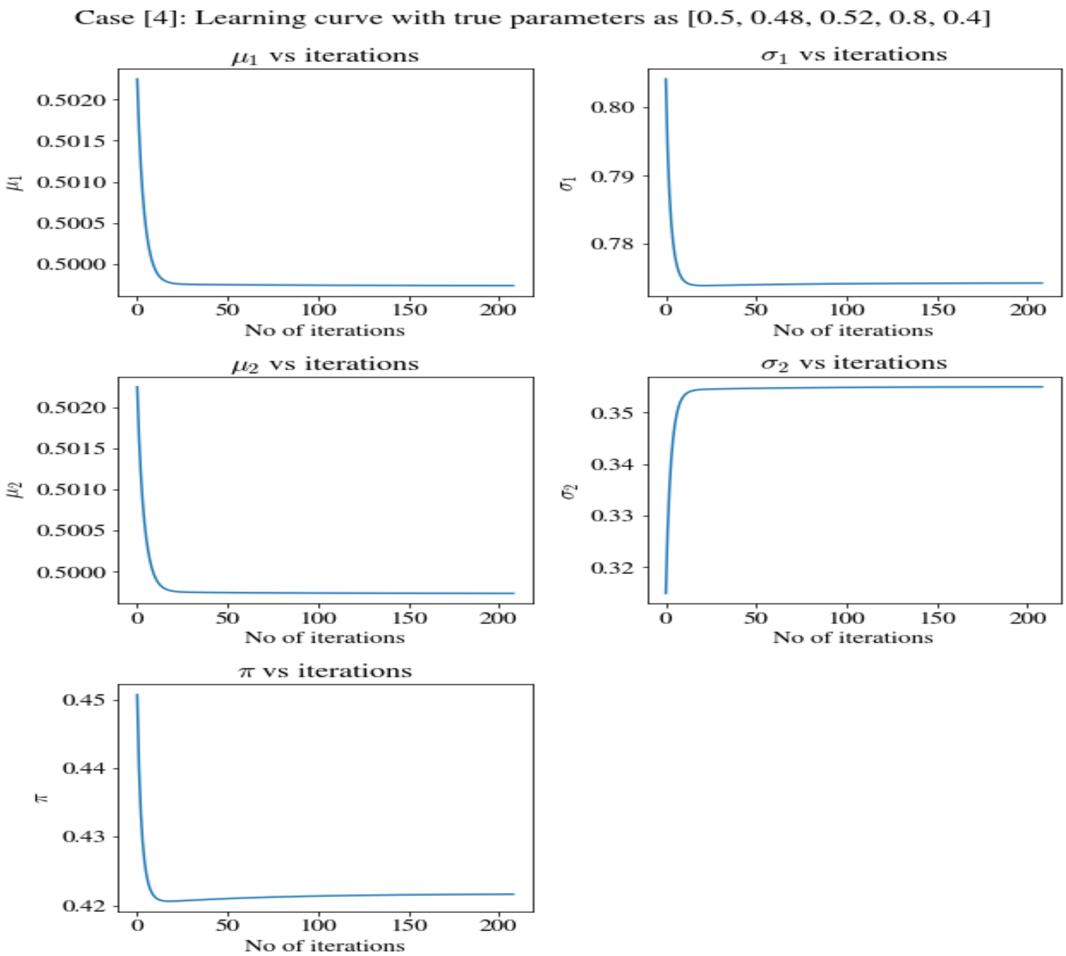


Figure 10: Learning curves for Case-(iv)

As we can see from the table, the error in estimating the mean of both the Gaussian distributions (roughly) increases as the means of the two distributions come closer.