ECSE 509 – Probability and Random Signals 2 – Fall 2020

Project Report

Alexander Fernandes

260960205

Project Description and Deliverables

Suppose that you are given a mixture of two exponential pdfs such that

$$f(x;\theta) = \pi_1 f_1(x;\lambda_1) + (1-\pi_1) f_2(x;\lambda_2) \tag{8}$$

where f_1 and f_2 are exponential distributions with parameters λ_1 and λ_2, respectively. In this assignment we want to estimate the parameter vector of this model $\theta = [\lambda_1, \lambda_2, \pi_1, \pi_2]^T$. Assume that $\lambda_1 = 1$, $\lambda_2 = 3$, and $\pi_1 = 0.25$.

(a) Plot the pdf of the exponential mixture.

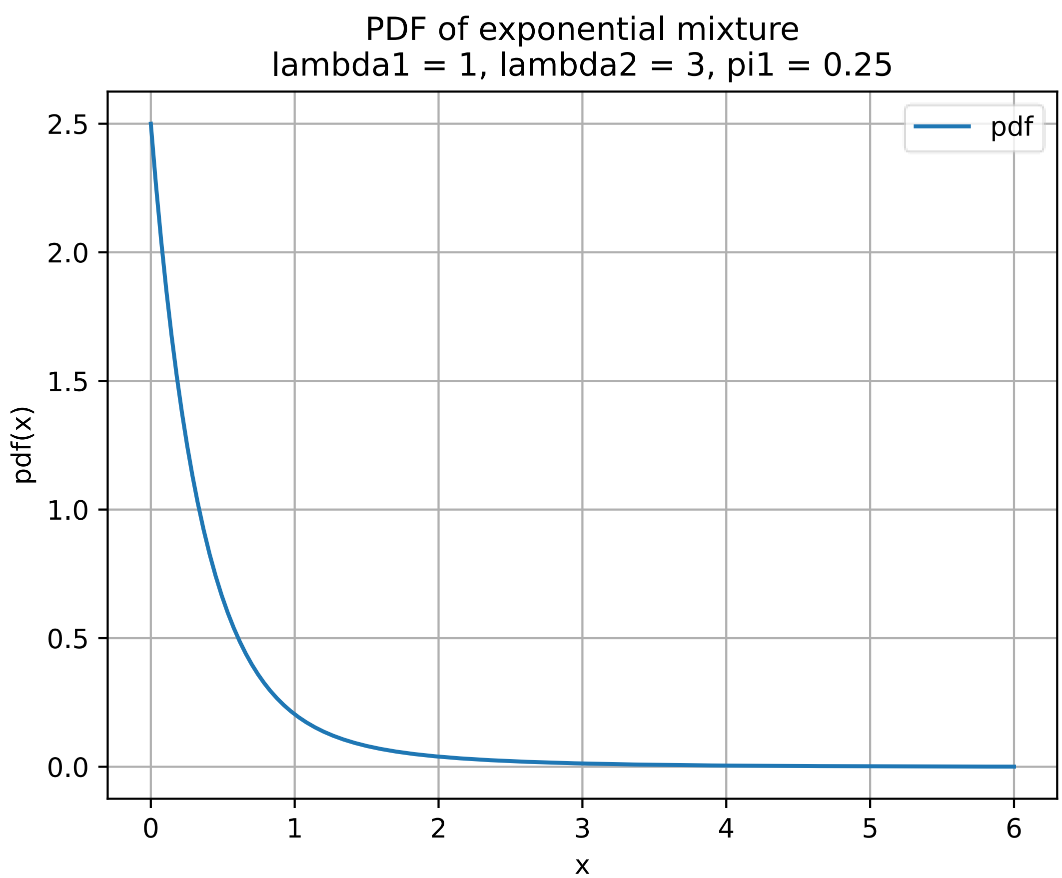


Figure 1 Plot of the pdf of the exponential mixture from x = 0 to 6

(b) Suppose that we observe n independent realizations of the above mixture. Derive the E and M update equations of the EM-algorithm for this problem. (Please include the derivation in your project report).

pdf exponential distribution:

$$\phi(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

pdf of exponential mixture distribution:

$$f(x;\theta) = \pi_1 \phi(x;\lambda_1) + \pi_2 \phi(x;\lambda_2)$$

$$f(x;\theta) = \pi_1 \lambda_1 e^{-\lambda_1 x} + \pi_2 \lambda_2 e^{-\lambda_2 x}$$

with $\theta^T = [\pi_1, \pi_2, \lambda_1, \lambda_2]$ parameters

and $\pi_1 + \pi_2 = 1$

likelihood function of complete distribution:

$$l(x;\theta) = \prod_{i=1}^{N} f(x_i;\theta) = \prod_{i=1}^{N} \left[\pi_1 \lambda_1 e^{-\lambda_1 x_i} + \pi_2 \lambda_2 e^{-\lambda_2 x_i}\right]$$

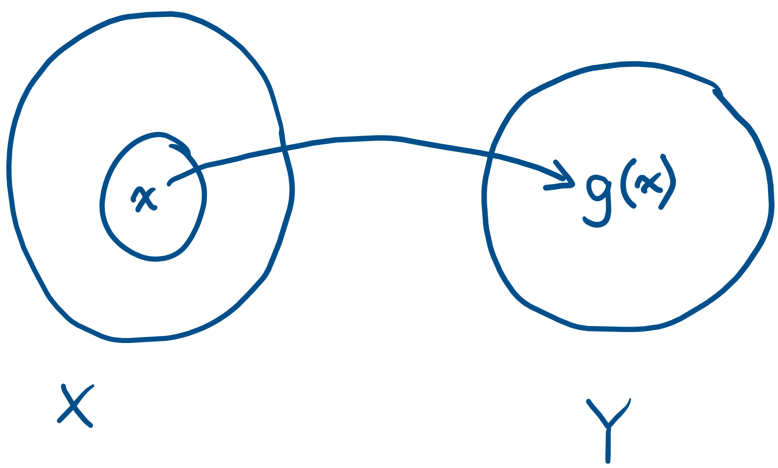
N observations from x distribution

log-likelihood function of complete distribution:

$$\boxed{L(\theta;x) = \sum_{i=1}^{N} \log\left[\pi_1 \lambda_1 e^{-\lambda_1 x_i} + \pi_2 \lambda_2 e^{-\lambda_2 x_i}\right]}$$

However it is difficult to work with $L(\theta;x)$ when deriving the EM algorithm:

We introduce $y = g(x)$ Y observed samples from X hidden samples



Introduce latent variable z to sample from individual pdf distributions within the mixture distribution:

$$z = \begin{cases} 0 & \text{with probability } \pi_1 \\ 1 & \text{with probability } \pi_2 \end{cases}$$

if $\underline{z} = 0$ sample from $\phi(x; \lambda_1)$

else $\underline{z} = 1$ sample from $\phi(x; \lambda_2)$

i.e.

$$f(y;\theta) = \int f(y, z; \theta)\, dz$$

$$f(y;\theta) = \int f(y|z;\theta)\, f(z;\theta)\, dz$$

$$f(y;\theta) = f(y|z=0;\theta)\, f(z=0;\theta) + f(y|z=1;\theta)\, f(z=1;\theta)$$

$$f(y;\theta) = \begin{cases} \pi_1 \phi(y;\lambda_1) & z = 0 \\ \pi_2 \phi(y;\lambda_2) & z = 1 \end{cases}$$

likelihood function:

$$\ell(y;\theta) = \prod_{i=1}^{N} f(y_i;\theta)$$

$$\ell(y;\theta) = \prod_{i=1}^{N} \left(\pi_1 \lambda_1 e^{-\lambda_1 y_i}\right)^{1-z} \left(\pi_2 \lambda_2 e^{-\lambda_2 y_i}\right)^{z}$$

log-likelihood function:

$$L(y,z;\theta) = \sum_{i=1}^{N} (1-z_i)\log\left(\pi_1 \lambda_1 e^{-\lambda_1 y_i}\right) + z_i \log\left(\pi_2 \lambda_2 e^{-\lambda_2 y_i}\right)$$

<u>Expectation step:</u>

$$Q(\theta|\theta^{(j)}) = E\left[L(y,z;\theta) \,\middle|\, Y;\theta^{(j)}\right]$$

$$Q(\theta|\theta^{(j)}) = \sum_{i=1}^{N} \left(1 - E[z_i | Y;\theta^{(j)}]\right) \log\left(\pi_1 \lambda_1 e^{-\lambda_1 y_i}\right)$$
$$+ E[z_i | Y;\theta^{(j)}] \log\left(\pi_2 \lambda_2 e^{-\lambda_2 y_i}\right)$$

$$Q(\theta|\theta^{(j)}) = \sum_{i=1}^{N} \underbrace{(1-\tau_i)}_{z=0}\log\left(\pi_1 \lambda_1 e^{-\lambda_1 y_i}\right) + \underbrace{\tau_i}_{z=1} \log\left(\pi_2 \lambda_2 e^{-\lambda_2 y_i}\right)$$

$$\tau_i = E\left(z_i = 1 \,\middle|\, Y;\theta^{(j)}\right)$$ $\leftarrow$ known as the recommender membership probability computed in the E-step

$$\tau_i = \frac{P(z_i = 1 \mid Y; \theta^{(j)})}{P(z_i = 0 \mid Y; \theta^{(j)}) + P(z_i = 1 \mid Y; \theta^{(j)})}$$

$$\boxed{\tau_i = \frac{\pi_2^{(j)} \lambda_2^{(j)} e^{-\lambda_2^{(j)} y_i}}{\pi_1^{(j)} \lambda_1^{(j)} e^{-\lambda_1^{(j)} y_i} + \pi_2^{(j)} \lambda_2^{(j)} e^{-\lambda_2^{(j)} y_i}}}$$

Maximization Step

$$\Theta^{(j+1)} = \Theta_{max} = \underset{\Theta}{\arg\max}\ Q(\Theta \mid \Theta^{(j)})$$

$$\frac{\partial Q}{\partial \lambda_1} = \frac{\partial}{\partial \lambda_1} \left[\sum_{i=1}^{N} (1-\tau_i)\left[\log \pi_1 + \log \lambda_1 - \lambda_1 y_i\right] + \tau_i \left[\log \pi_2 + \log \lambda_2 - \lambda_2 y_i\right] \right]$$

$$\frac{\partial Q}{\partial \lambda_1} = \sum_{i=1}^{N} (1-\tau_i)\frac{1}{\lambda_1} - (1-\tau_i) y_i = 0$$

$$\frac{1}{\lambda_1} \sum_{i=1}^{N} (1-\tau_i) = \sum_{i=1}^{N} y_i$$

$$\boxed{\lambda_1^{(j+1)} = \frac{\sum_{i=1}^{N}(1-\tau_i)}{\sum_{i=1}^{N}(1-\tau_i)\,y_i}}$$

$$\frac{\partial Q}{\partial \lambda_2} = \sum_{i=1}^{N} \tau_i \frac{1}{\lambda_2} - \tau_i y_i = 0$$

$$\boxed{\lambda_2^{(j+1)} = \frac{\sum_{i=1}^{N}\tau_i}{\sum_{i=1}^{N}\tau_i\, y_i}}$$

$$\frac{\partial Q}{\partial \pi_1} = \sum_{i=1}^{N} (1-\tau_i)\frac{1}{\pi_1} - \tau_i \frac{1}{1-\pi_1} = 0$$

$$\frac{1-\pi_1}{\pi_1} = \frac{\sum_{i=1}^{N}\tau_i}{\sum_{i=1}^{N}(1-\tau_i)}$$

$$\frac{1}{\pi_1} = \frac{\sum_{i=1}^{N}\tau_i}{\sum_{i=1}^{N}(1-\tau_i)} + 1$$

$$\boxed{\pi_1^{(j+1)} = \frac{1}{\left(\frac{\sum_{i=1}^{N} \tau_i}{\sum_{i=1}^{N} (1-\tau_i)} + 1\right)}}$$

$$\boxed{\pi_2^{(j+1)} = 1 - \pi_1^{(j+1)}}$$

(c) Write a program which applies the EM algorithm you derived. You must include your source code (with comments so that it is easy to follow how it works) in your project report. *If the source code is not included, you will receive a mark of zero (0) for this assignment.* Assume that $n = 20$.

(d) Compute the log-likelihood of the incomplete set at each iteration and plot it to verify that it increased monotonically.

(e) For some iterations plot the estimate of the exponential mixture pdf:

$$\hat{f}(x) = \hat{\pi}_1 f_1(x; \hat{\lambda}_1) + \hat{\pi}_2 f_2(x; \hat{\lambda}_2). \tag{9}$$

For my program I decided to run three trial experiments:

One trial consists of the following:

1. randomly sample 20 latent variables z and 20 sample y observed variables from the corresponding distributions defined previously in my derivation for the EM algorithm
2. Run the EM algorithm independently using three different sets of initial parameters
   a. lambda1 = 0.5, lambda2 = 4, pi1 = 0.3
   b. lambda1 = 0.01, lambda2 = 10, pi1 = 0.1
   c. lambda1 = 0.1, lambda2 = 50, pi1 = 0.6
3. Plot the log-likelihood of the three initial parameters to compare
4. Plot the estimate of the exponential mixture pdf of the initial parameters, first iteration and final iteration

This procedure is conducted three times to obtain three trial experiments on the sampled data. I chose to run the experiments this way to compare three initial parameter estimates on three sampled data sets.

Trial 1

z = [1. 1. 1. 1. 1. 0. 0. 1. 0. 1. 1. 1. 0. 1. 1. 1. 0. 1. 1. 1.]

y = [0.15080982 0.53139416 0.15132354 0.42463518 0.01551725 0.29190298

0.10849508 0.22584108 0.28218577 0.29066336 0.00570947 0.05730676

0.65645223 0.29476604 0.09051243 0.16014269 1.41848097 0.72342676

0.0585217 0.11573235]

corresponding lambda1 = 1 / mean(y with z = 0) = 1.8132254280997984
corresponding lambda2 = 1 / mean(y with z = 1) = 4.550553101696321
corresponding pi1 = mean(z) = 0.25

Note: these statistics using the latent z variable are unknown to the EM algorithm and is shown for purposes of showing the underlying statistics of the sampled data that the EM algorithm is trying to estimate

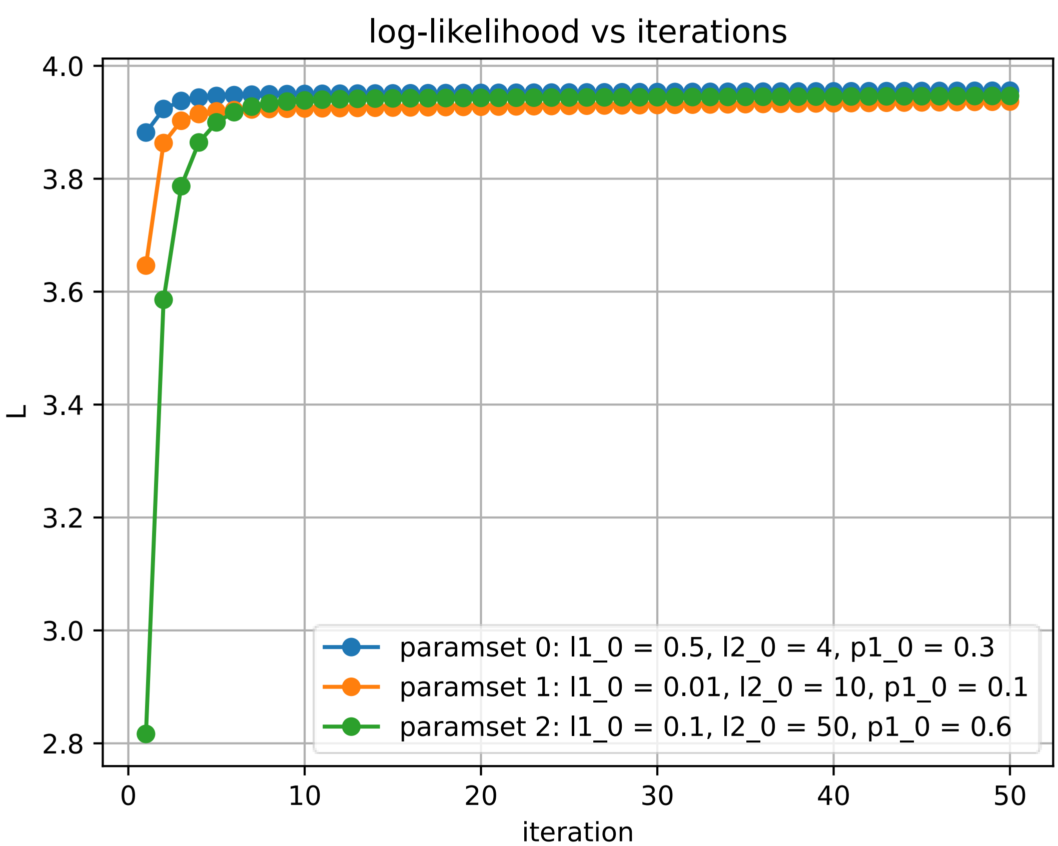


Figure 2: log-likelihood of iterations with trial 1 data

Set of initial parameters (a) with following EM algorithm values with trial 1 data

l1 [0.5 1.52114921 1.6532455 1.73803693 1.79602745 1.8373668
1.86773124 1.89056263 1.9080717 1.92174006 1.932593 1.94135748
1.94855869 1.95458147 1.95971096 1.96416023 1.96808951 1.97161994
1.97484338 1.97782978 1.98063243 1.98329207 1.98583985 1.98829959
1.99068954 1.9930237 1.99531277 1.99756497 1.99978657 2.00198239
2.00415613 2.00631059 2.00844794 2.01056981 2.01267746 2.01477185
2.01685371 2.01892359 2.0209819 2.02302897 2.02506503 2.02709026
2.02910481 2.03110877 2.03310222 2.03508522 2.03705781 2.03902004
2.04097192 2.04291348 2.04484474 2.04676572]

l2 [4. 4.34681903 4.16758958 4.073796 4.01821296 3.98295857
3.95974424 3.94419036 3.93376036 3.92688279 3.9225334 3.9200181
3.9188522 3.91868927 3.91927714 3.92042973 3.92200814 3.92390792
3.92604993 3.92837397 3.93083403 3.93339489 3.93602951 3.93871709
3.94144163 3.94419081 3.94695514 3.94972731 3.9525017 3.95527396
3.95804079 3.96079961 3.96354848 3.9662859 3.96901075 3.97172217
3.97441951 3.9771023 3.9797702 3.98242294 3.98506038 3.98768238
3.99028889 3.99287989 3.99545536 3.99801533 4.00055984 4.00308894
4.00560269 4.00810115 4.01058441 4.01305255]

p1 [0.3 0.1699746 0.17193753 0.17345054 0.17480049 0.17606936
0.17728951 0.17847684 0.17964043 0.18078597 0.1819173 0.18303707
0.18414718 0.18524902 0.18634361 0.1874317 0.18851386 0.1895905
0.19066193 0.1917284 0.19279008 0.19384708 0.19489951 0.19594743
0.19699089 0.19802992 0.19906455 0.20009479 0.20112064 0.20214212
0.20315922 0.20417194 0.20518027 0.20618423 0.20718379 0.20817897
0.20916975 0.21015612 0.2111381 0.21211567 0.21308884 0.21405759
0.21502194 0.21598189 0.21693742 0.21788855 0.21883527 0.21977759
0.22071552 0.22164905 0.2225782 0.22350296]

Set of initial parameters (b) with following EM algorithm values with trial 1 data

l1 [0.01 0.7898183 1.12625835 1.324535 1.44924756 1.53228907
1.58997447 1.63131311 1.66164708 1.68433179 1.70157206 1.71486845
1.72527037 1.73352741 1.74018383 1.74563986 1.75019272 1.75406483
1.75742362 1.76039569 1.76307713 1.76554103 1.76784315 1.77002605
1.77212228 1.77415672 1.77614843 1.77811197 1.78005849 1.78199649
1.78393246 1.78587132 1.78781679 1.7897717 1.79173812 1.7937176
1.79571126 1.79771987 1.79974396 1.80178385 1.80383971 1.80591159
1.80799944 1.81010311 1.81222242 1.81435713 1.81650696 1.81867159
1.82085068 1.82304387 1.82525079 1.82747103]

l2 [10. 4.18473989 3.80862446 3.68774061 3.63283138 3.60334444
3.58612518 3.57571582 3.56945653 3.56589688 3.56417825 3.56375989
3.56428465 3.56550779 3.56725666 3.56940668 3.57186634 3.57456745
3.57745863 3.58050081 3.58366405 3.58692525 3.5902665 3.59367383
3.59713628 3.6006452 3.60419374 3.6077764 3.61138877 3.61502725
3.61868888 3.62237119 3.62607211 3.62978984 3.63352284 3.63726973
3.6410293 3.6448004 3.64858201 3.65237316 3.65617292 3.65998042
3.66379483 3.66761532 3.6714411 3.67527141 3.67910548 3.68294257
3.68678195 3.69062291 3.69446474 3.69830675]

p1 [0.1 0.062043 0.06417212 0.06515373 0.06612144 0.06710512
0.06809272 0.06907896 0.07006293 0.07104548 0.072028 0.07301185
0.07399825 0.07498821 0.07598254 0.07698185 0.07798662 0.07899719
0.08001383 0.08103669 0.08206588 0.08310147 0.08414345 0.08519183
0.08624655 0.08730755 0.08837474 0.08944803 0.09052731 0.09161247
0.09270338 0.09379991 0.09490192 0.09600926 0.0971218 0.09823938
0.09936185 0.10048904 0.10162082 0.102757 0.10389744 0.10504197
0.10619042 0.10734263 0.10849843 0.10965765 0.11082013 0.11198569
0.11315418 0.11432541 0.11549923 0.11667547]

Set of initial parameters (c) with following EM algorithm values with trial 1 data

l1 [0.1 2.43848648 2.67069162 2.77017251 2.82674323 2.86340542
2.88886734 2.90725587 2.92084318 2.93101428 2.93867403 2.94444316
2.94876289 2.95195561 2.954262 2.95586498 2.95690557 2.95749383
2.95771652 2.95764261 2.95732729 2.95681501 2.95614174 2.95533667
2.9544236 2.95342198 2.95234772 2.95121389 2.95003122 2.94880851
2.94755305 2.9462708 2.94496671 2.94364484 2.94230854 2.9409606
2.93960328 2.93823849 2.93686779 2.93549246 2.93411356 2.93273199
2.93134846 2.92996359 2.92857787 2.92719173 2.92580549 2.92441946
2.92303386 2.9216489 2.92026474 2.91888152]

l2 [50. 15.91630633 11.52852942 9.51193037 8.36466802 7.63991047
7.15149492 6.80733527 6.55663835 6.36922673 6.2261702 6.11507878
6.02755595 5.95774225 5.90144504 5.85559927 5.81792254 5.78668823
5.76057246 5.73854851 5.71981235 5.70372918 5.68979444 5.67760473
5.66683603 5.65722707 5.64856648 5.64068278 5.63343662 5.6267145
5.62042387 5.61448916 5.60884863 5.6034518 5.59825735 5.59323147
5.58834647 5.58357967 5.5789125 5.57432974 5.56981892 5.56536981
5.56097403 5.55662474 5.55231628 5.54804406 5.54380429 5.53959385
5.53541021 5.53125127 5.52711531 5.52300092]

p1 [0.6 0.6907248 0.75062398 0.77214883 0.78193821 0.78683288
0.78933558 0.79054802 0.79101411 0.79102445 0.79074386 0.79027052
0.78966534 0.78896746 0.78820287 0.7873893 0.78653912 0.78566121
0.78476202 0.78384635 0.78291781 0.78197917 0.78103256 0.78007967
0.77912181 0.77816003 0.77719518 0.77622795 0.77525888 0.77428843
0.77331696 0.77234478 0.77137215 0.77039927 0.76942631 0.76845343
0.76748073 0.76650834 0.76553632 0.76456477 0.76359374 0.76262329
0.76165347 0.76068432 0.75971587 0.75874817 0.75778124 0.75681512
0.75584981 0.75488536 0.75392177 0.75295908]

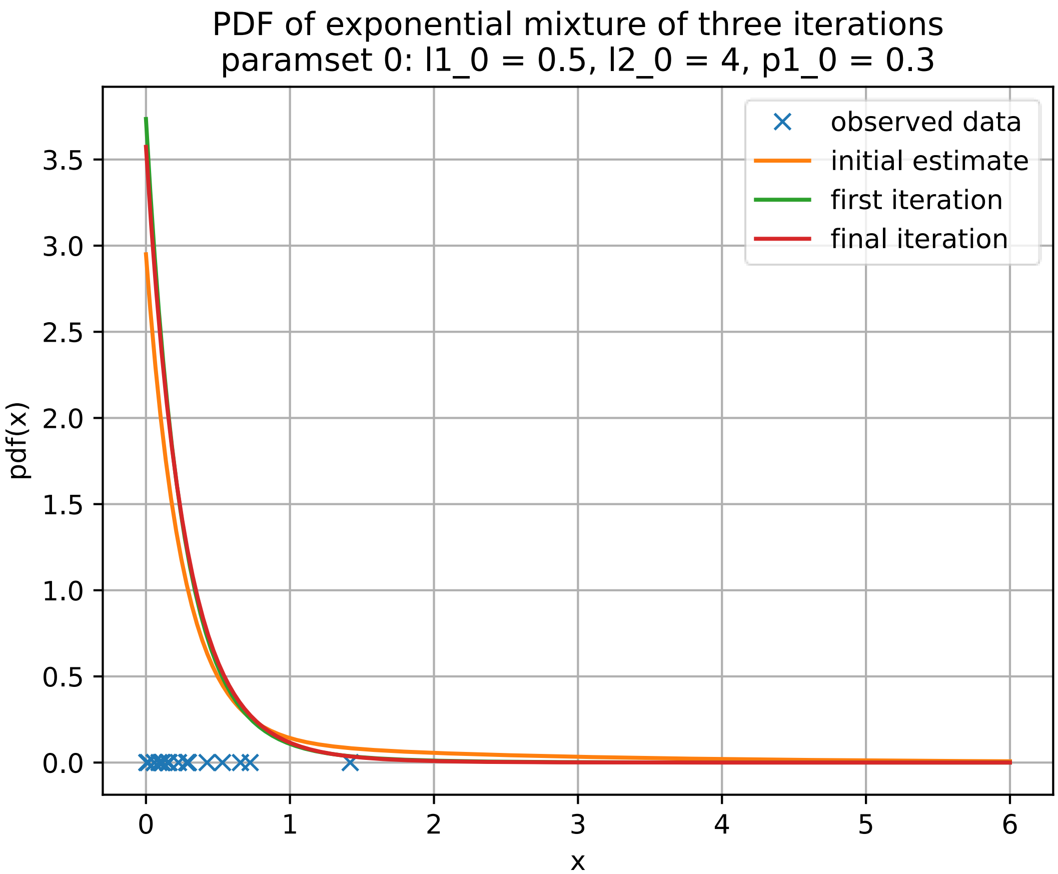


Figure 3: Plot iterations using the initial parameter set (a) with trial 1 data

l1[0] 0.5

l1[1] 1.5211492104250643

l1[-1] 2.0467657198431883

l2[0] 4.0

l2[1] 4.3468190255369725

l2[-1] 4.013052545654996

p1[0] 0.3

p1[1] 0.16997459706953924

p1[-1] 0.22350295671422243

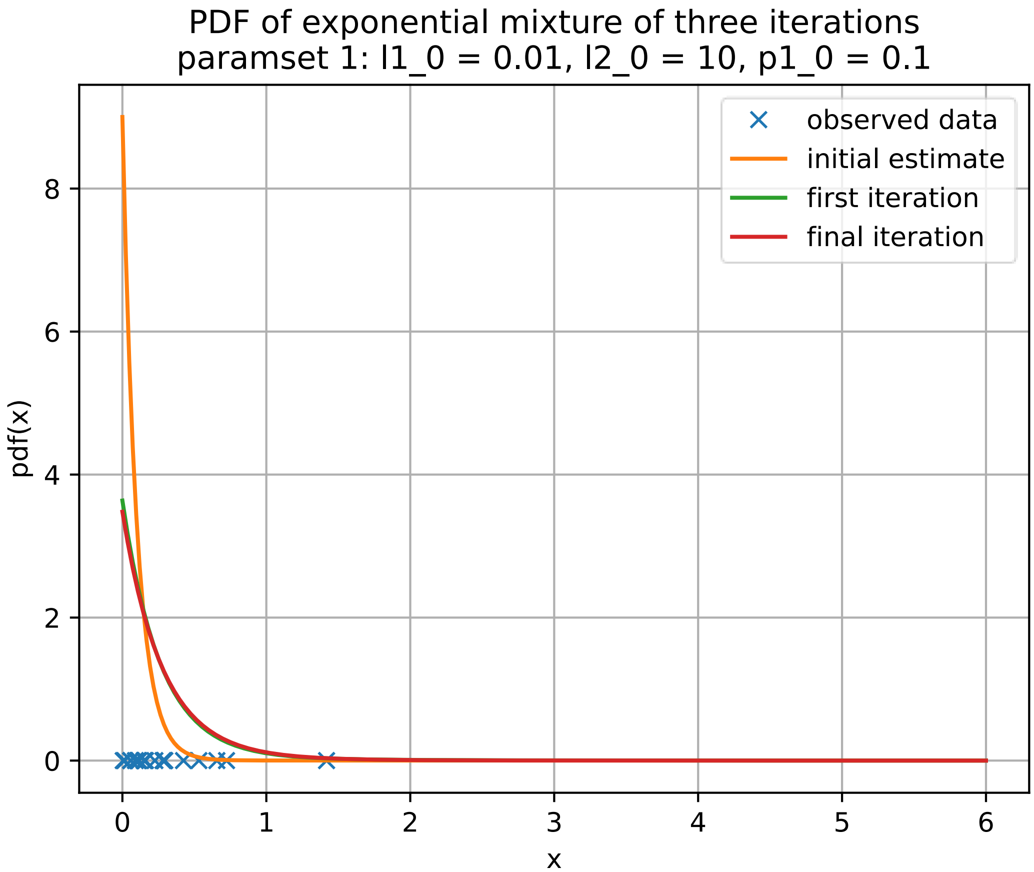


Figure 4: Plot iterations using the initial parameter set (b) with trial 1 data

l1[0] 0.01

l1[1] 0.7898183037535644

l1[-1] 1.8274710266149985

l2[0] 10.0

l2[1] 4.184739886392114

l2[-1] 3.6983067504072964

p1[0] 0.1

p1[1] 0.062043003841997 19

p1[-1] 0.11667546678477289

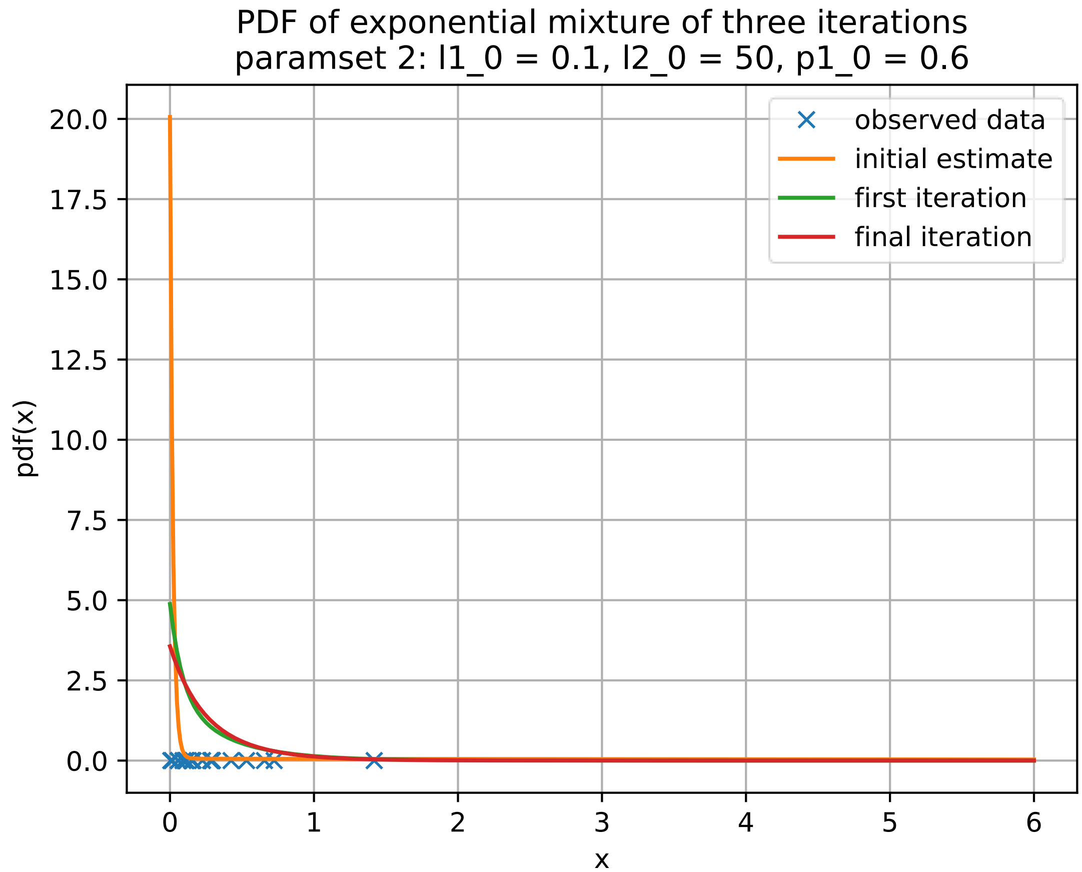


Figure 5: Plot iterations using the initial parameter set (c) with trial 1 data

l1[0] 0.1

l1[1] 2.4384864821364958

l1[-1] 2.918881522274517

l2[0] 50.0

l2[1] 15.916306331712198

l2[-1] 5.523000920786463

p1[0] 0.6

p1[1] 0.6907248002618296

p1[-1] 0.7529590810229062

Trial 2

latent variable z [1. 0. 1. 0. 1. 1. 0. 1. 1. 1. 1. 0. 1. 1. 1. 1. 0. 1. 0. 1.]

observed sample y [1.11492269 1.50577472 0.14934126 1.51995324 0.61153304 0.22057551

0.98009025 0.45943436 0.01333259 0.40912103 0.32575831 0.628392

0.15246232 0.02150881 0.05534196 0.33708163 0.13160242 0.72365789

0.910331 0.85572887]

corresponding lambda1 = 1 / mean(y with z = 0) = 1.0570557064833968
corresponding lambda2 = 1 / mean(y with z = 1) = 2.568901479452905
corresponding pi1 = mean(z) = 0.3

Note: these statistics using the latent z variable are unknown to the EM algorithm and is shown for purposes of showing the underlying statistics of the sampled data that the EM algorithm is trying to estimate

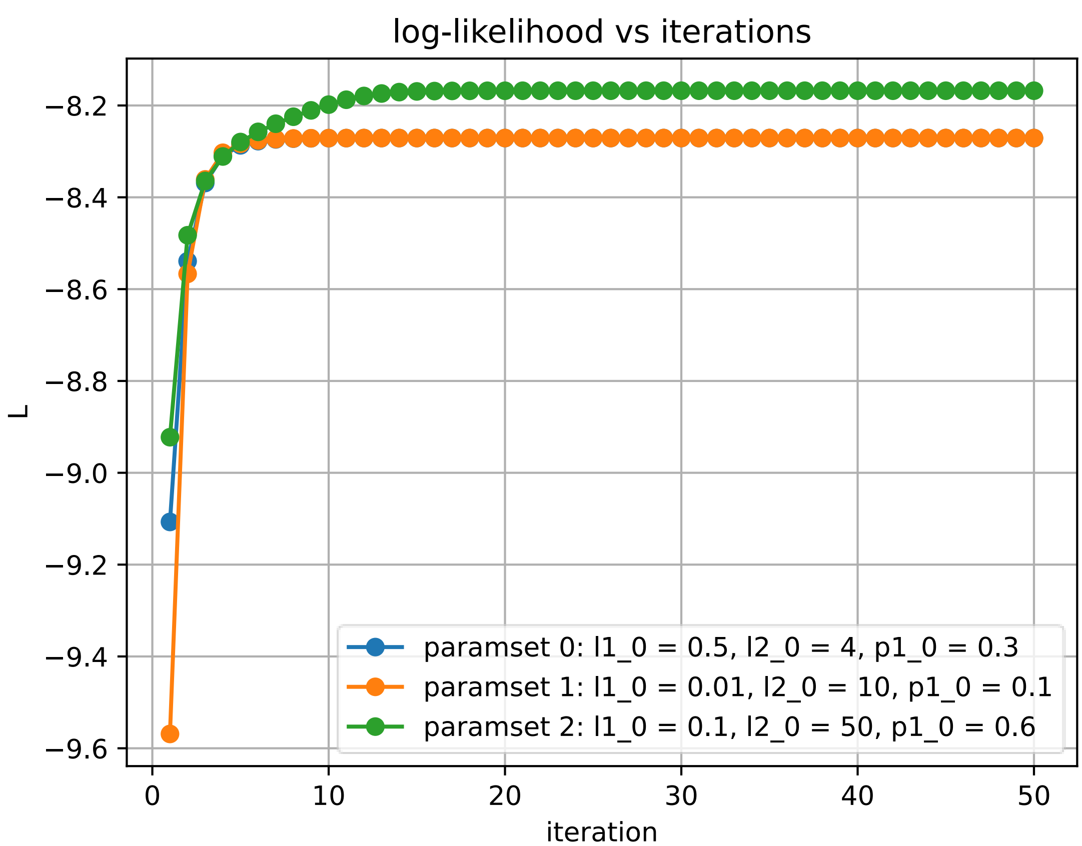


Figure 6: log-likelihood of iterations with trial 2 data

Set of initial parameters (a) with following EM algorithm values with trial 2 data

l1 [0.5 1.05541994 1.28235759 1.44483063 1.55881548 1.63679137
1.68950971 1.72498924 1.74882905 1.76484055 1.77559385 1.78281633
1.78766782 1.79092696 1.79311654 1.79458763 1.79557604 1.79624015
1.79668637 1.7969862 1.79718767 1.79732303 1.79741399 1.79747511
1.79751618 1.79754377 1.79756232 1.79757477 1.79758315 1.79758877
1.79759255 1.79759509 1.7975968 1.79759794 1.79759871 1.79759923
1.79759958 1.79759981 1.79759997 1.79760008 1.79760015 1.79760019
1.79760023 1.79760025 1.79760026 1.79760027 1.79760028 1.79760028
1.79760029 1.79760029 1.79760029 1.79760029]

l2 [4. 2.71937199 2.25756161 2.0529745 1.94952279 1.89203376
1.85793957 1.83683397 1.82339858 1.81468835 1.80897335 1.80519374
1.80268091 1.80100442 1.79988327 1.79913234 1.79862885 1.79829102
1.79806424 1.79791195 1.79780967 1.79774097 1.79769481 1.7976638
1.79764296 1.79762896 1.79761956 1.79761324 1.79760899 1.79760614
1.79760422 1.79760293 1.79760207 1.79760148 1.79760109 1.79760083
1.79760065 1.79760054 1.79760046 1.7976004 1.79760037 1.79760034
1.79760033 1.79760031 1.79760031 1.7976003 1.7976003 1.7976003
1.7976003 1.79760029 1.79760029 1.79760029]

p1 [0.3 0.32524807 0.33646651 0.33751611 0.33718801 0.33687853
0.33670424 0.33661755 0.33657643 0.33655734 0.33654858 0.33654458
0.33654277 0.33654194 0.33654157 0.3365414 0.33654132 0.33654129
0.33654127 0.33654127 0.33654126 0.33654126 0.33654126 0.33654126
0.33654126 0.33654126 0.33654126 0.33654126 0.33654126 0.33654126
0.33654126 0.33654126 0.33654126 0.33654126 0.33654126 0.33654126
0.33654126 0.33654126 0.33654126 0.33654126 0.33654126 0.33654126
0.33654126 0.33654126 0.33654126 0.33654126 0.33654126 0.33654126
0.33654126 0.33654126 0.33654126 0.33654126]

Set of initial parameters (b) with following EM algorithm values with trial 2 data

l1 [0.01 0.83505844 1.15940451 1.37970307 1.52396018 1.61735652
1.67818891 1.71814182 1.74456228 1.76212171 1.77383299 1.7816627
1.78690592 1.79042097 1.79277923 1.79436219 1.7954251 1.79613898
1.7966185 1.79694065 1.79715708 1.79730249 1.79740019 1.79746584
1.79750995 1.79753959 1.7975595 1.79757289 1.79758188 1.79758792
1.79759198 1.79759471 1.79759654 1.79759777 1.7975986 1.79759915
1.79759953 1.79759978 1.79759995 1.79760006 1.79760014 1.79760019
1.79760022 1.79760025 1.79760026 1.79760027 1.79760028 1.79760028
1.79760029 1.79760029 1.79760029 1.79760029]

l2 [10. 2.76991551 2.18326275 1.98908524 1.90557407 1.86289283
1.83868538 1.82407368 1.81491103 1.80902588 1.80518725 1.80265828
1.80098113 1.79986404 1.79911781 1.79861836 1.79828365 1.79805914
1.79790846 1.7978073 1.79773936 1.79769372 1.79766306 1.79764247
1.79762863 1.79761933 1.79761309 1.79760889 1.79760607 1.79760417
1.7976029 1.79760204 1.79760147 1.79760108 1.79760082 1.79760065
1.79760053 1.79760045 1.7976004 1.79760037 1.79760034 1.79760033
1.79760031 1.79760031 1.7976003 1.7976003 1.7976003 1.7976003
1.79760029 1.79760029 1.79760029 1.79760029]

p1 [0.1 0.23344358 0.2429458 0.24117784 0.23986927 0.23925468
0.23898269 0.23886296 0.23881001 0.23878647 0.23877595 0.23877124
0.23876912 0.23876817 0.23876774 0.23876754 0.23876746 0.23876742
0.2387674 0.23876739 0.23876739 0.23876739 0.23876738 0.23876738
0.23876738 0.23876738 0.23876738 0.23876738 0.23876738 0.23876738
0.23876738 0.23876738 0.23876738 0.23876738 0.23876738 0.23876738
0.23876738 0.23876738 0.23876738 0.23876738 0.23876738 0.23876738
0.23876738 0.23876738 0.23876738 0.23876738 0.23876738 0.23876738
0.23876738 0.23876738 0.23876738 0.23876738]

Set of initial parameters (c) with following EM algorithm values with trial 2 data

l1 [0.1 1.50063091 1.59977923 1.64269115 1.66767787 1.6842823
1.69613472 1.70495221 1.7116653 1.71683032 1.72081177 1.72387777
1.72625356 1.7281384 1.72968858 1.73099786 1.73210777 1.73303658
1.73380009 1.73441779 1.73491153 1.73530287 1.73561133 1.73585358
1.73604339 1.7361919 1.736308 1.73639872 1.73646959 1.73652494
1.73656817 1.73660193 1.7366283 1.7366489 1.73666499 1.73667756
1.73668737 1.73669504 1.73670103 1.73670571 1.73670936 1.73671222
1.73671445 1.73671619 1.73671755 1.73671862 1.73671945 1.7367201
1.73672061 1.736721 1.73672131 1.73672155]

l2 [50. 20.22174127 17.63405158 17.05880569 17.28866319 18.0207938
19.1624328 20.69721061 22.64174389 25.01707367 27.80966599 30.91496078
34.09335516 37.01178014 39.39207929 41.13837303 42.32377911 43.09232185
43.58138769 43.89314771 44.09499587 44.22881753 44.31998498 44.38379794
44.42957268 44.46309564 44.48805759 44.50688483 44.52122241 44.5322186
44.54069564 44.54725493 44.55234387 44.55629959 44.55937866 44.56177773
44.5636483 44.56510756 44.56624636 44.56713533 44.56782942 44.56837143
44.56879473 44.56912535 44.56938359 44.56958532 44.56974289 44.56986599
44.56996215 44.57003728 44.57009597 44.57014182]

p1 [0.6 0.82155454 0.878973 0.90464189 0.92000848 0.9304623
0.93807367 0.94383335 0.94828507 0.95175716 0.95446515 0.95656643
0.95819321 0.95946595 0.96048585 0.96132177 0.96201203 0.96257841
0.96303765 0.96340574 0.96369812 0.96392888 0.96411024 0.96425238
0.96436361 0.96445055 0.96451847 0.96457152 0.96461295 0.9646453
0.96467056 0.96469029 0.9647057 0.96471773 0.96472713 0.96473447
0.96474021 0.96474468 0.96474818 0.96475092 0.96475305 0.96475472
0.96475602 0.96475704 0.96475784 0.96475846 0.96475894 0.96475932
0.96475962 0.96475985 0.96476003 0.96476017]

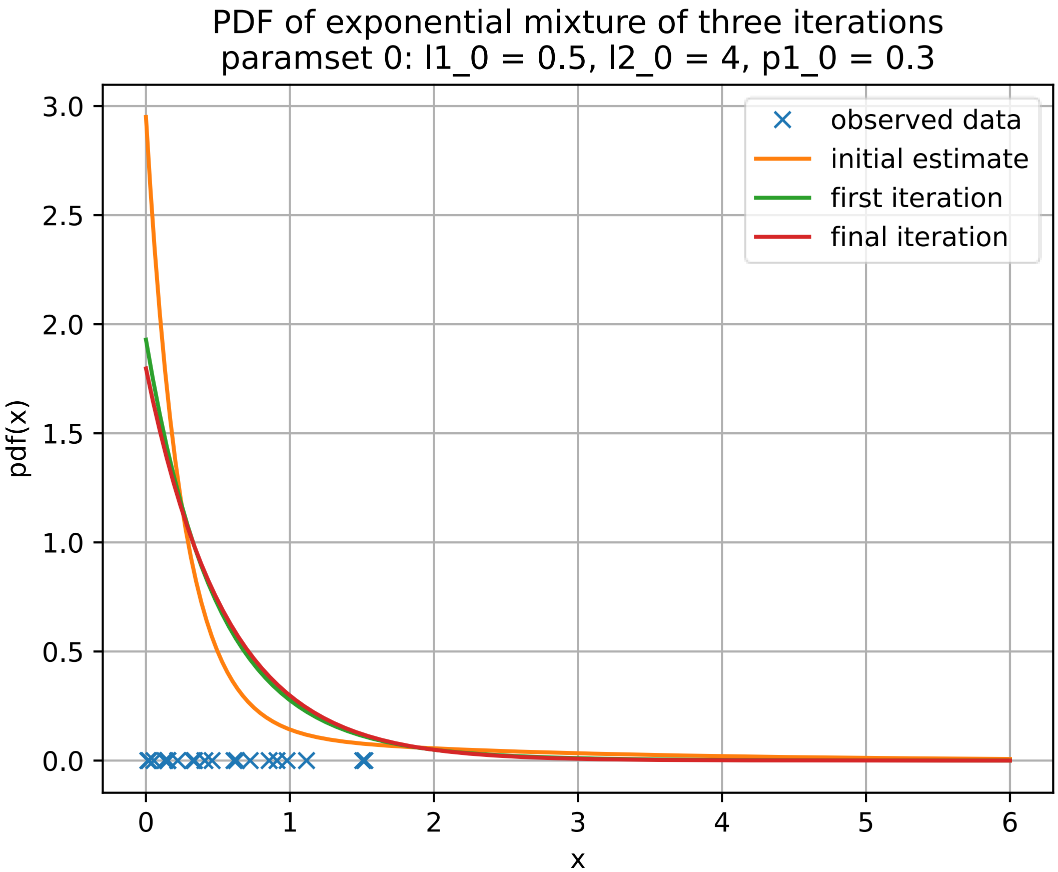


Figure 7: Plot iterations using the initial parameter set (a) with trial 2 data

l1[0] 0.5

l1[1] 1.0554199411041836

l1[-1] 1.7976002906518882

l2[0] 4.0

l2[1] 2.7193719896591984

l2[-1] 1.7976002934120388

p1[0] 0.3

p1[1] 0.32524806968525966

p1[-1] 0.33654126157938324

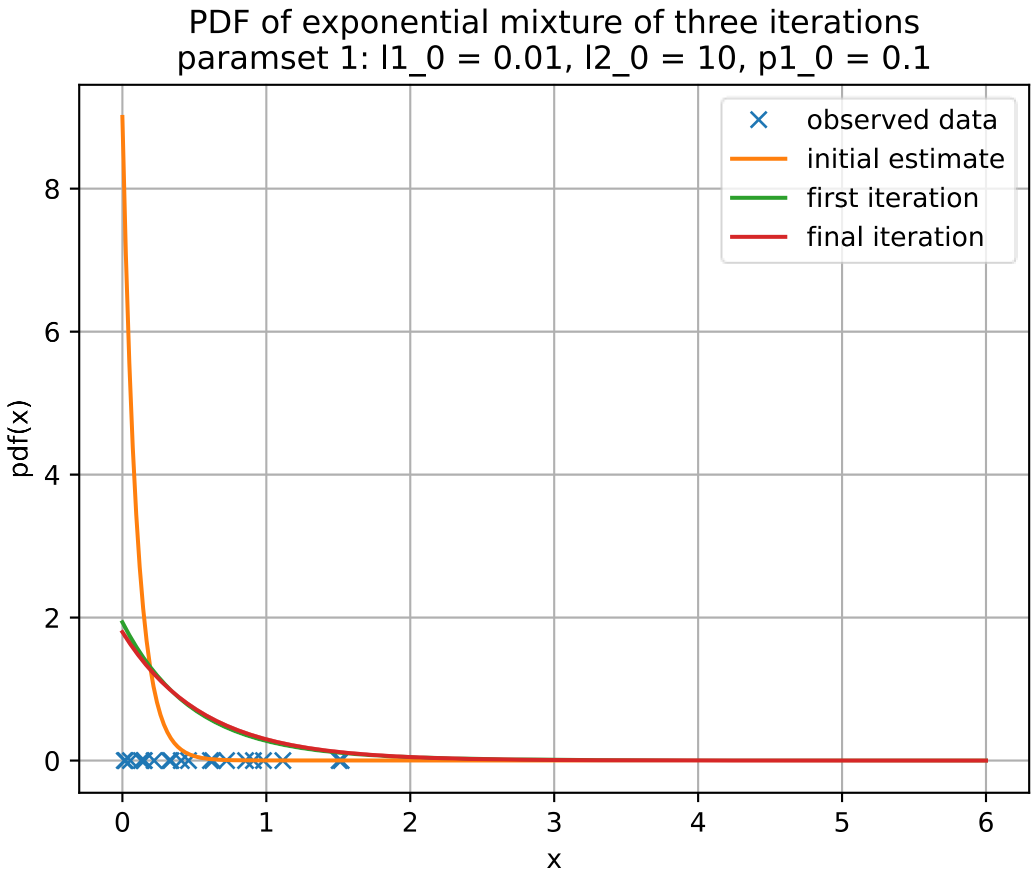


Figure 8: Plot iterations using the initial parameter set (b) with trial 2 data

l1[0] 0.01

l1[1] 0.8350584434102448

l1[-1] 1.7976002905163444

l2[0] 10.0

l2[1] 2.769915512291139

l2[-1] 1.797600293100036

p1[0] 0.1

p1[1] 0.23344358116394429

p1[-1] 0.2387673840085189

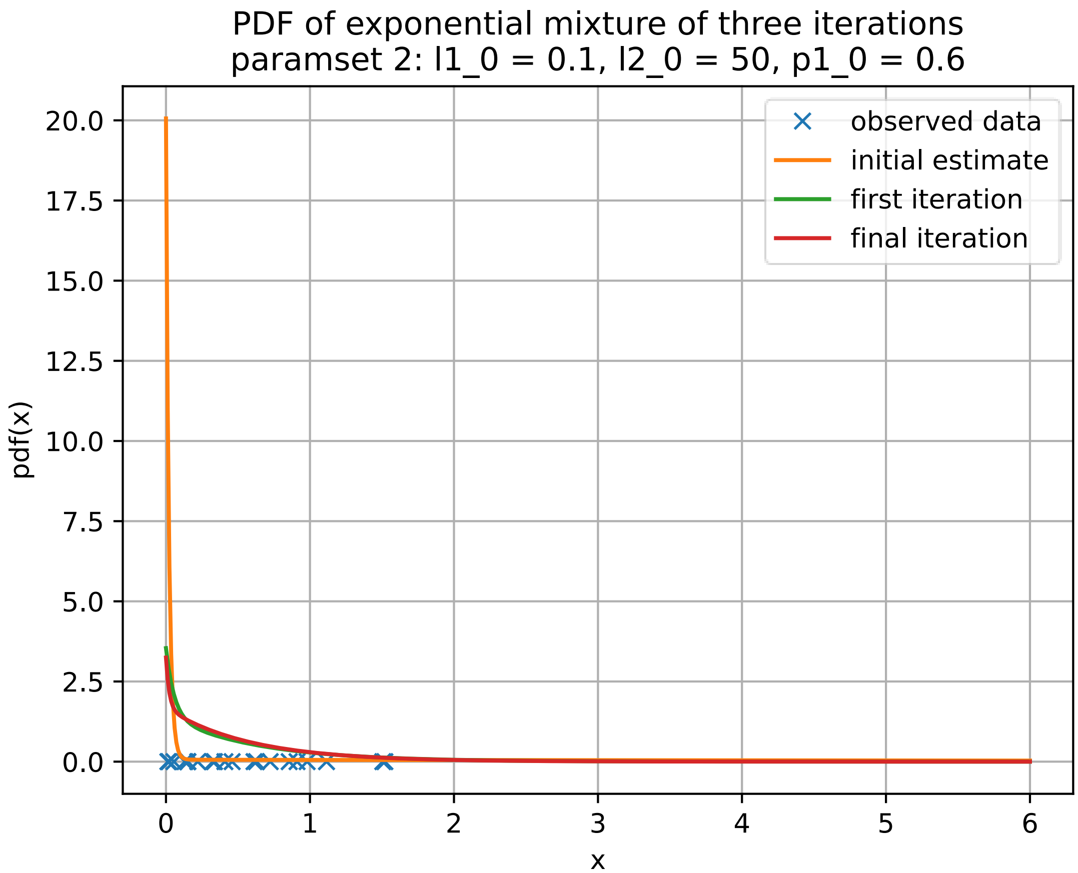


Figure 9: Plot iterations using the initial parameter set (c) with trial 2 data

l1[0] 0.1

l1[1] 1.5006309072881505

l1[-1] 1.7367215531178295

l2[0] 50.0

l2[1] 20.221741271537535

l2[-1] 44.57014181584588

p1[0] 0.6

p1[1] 0.8215545430351496

p1[-1] 0.9647601718131474

Trial 3

latent variable z [1. 1. 1. 1. 1. 1. 0. 0. 1. 0. 1. 1. 1. 1. 1. 1. 0. 1. 1. 1.]

observed sample y [0.63410089 0.17974152 0.71293181 0.3688489 0.93324476 0.4504294

 0.49755298 0.98283416 0.38372142 1.64645815 0.11902371 0.10463851

 0.00261723 0.77614421 1.10686012 0.03336725 0.92115078 0.1462741

 0.06931131 0.33561161]

corresponding lambda1 = 1 / mean(y with z = 0) = 0.9881432518760278
corresponding lambda2 = 1 / mean(y with z = 1) = 2.5169632564833613
corresponding pi1 = mean(z) = 0.2

Note: these statistics using the latent z variable are unknown to the EM algorithm and is shown for purposes of showing the underlying statistics of the sampled data that the EM algorithm is trying to estimate

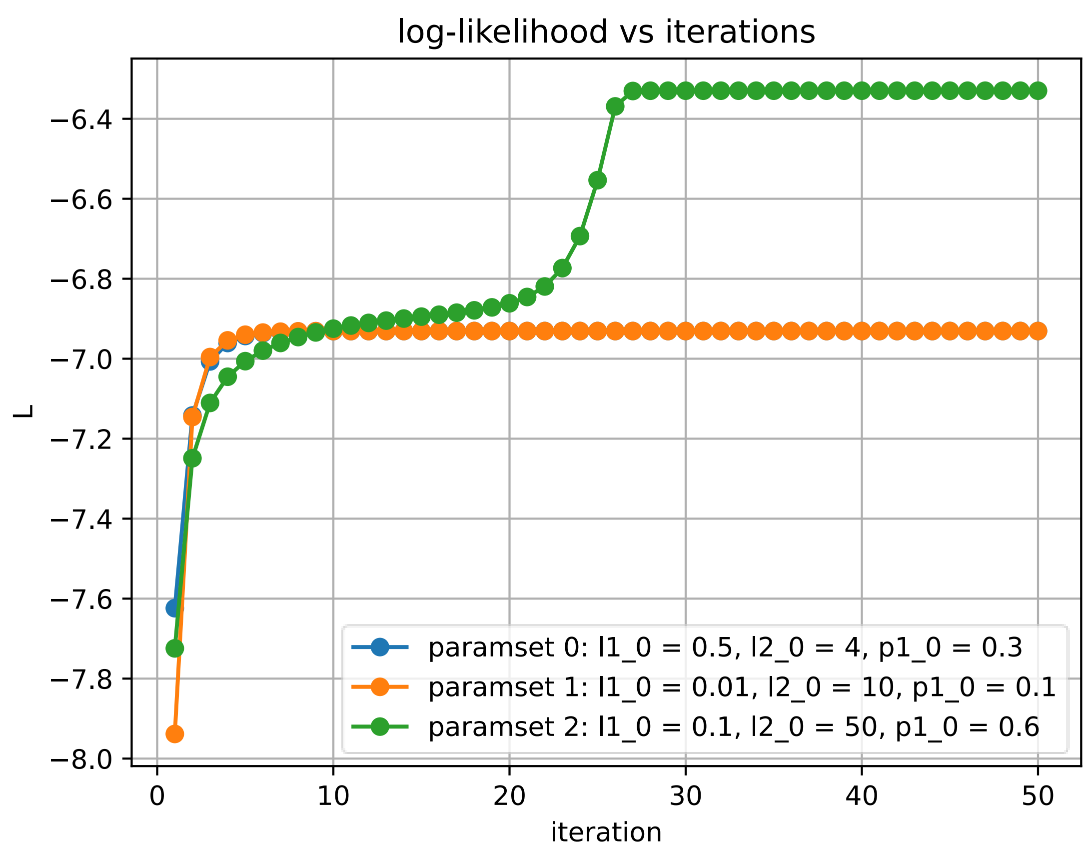


Figure 10: log-likelihood of iterations with trial 3 data

Set of initial parameters (a) with following EM algorithm values with trial 3 data

l1 [0.5 1.13530795 1.38501577 1.55847513 1.67714523 1.7572949
1.81120717 1.84745206 1.87183265 1.88824453 1.89929929 1.90674927
1.91177172 1.9151585 1.91744271 1.91898347 1.92002285 1.92072403
1.92119709 1.92151624 1.92173157 1.92187684 1.92197486 1.92204099
1.92208561 1.92211572 1.92213603 1.92214973 1.92215898 1.92216522
1.92216943 1.92217227 1.92217418 1.92217547 1.92217635 1.92217694
1.92217733 1.9221776 1.92217778 1.9221779 1.92217799 1.92217804
1.92217808 1.9221781 1.92217812 1.92217813 1.92217814 1.92217815
1.92217815 1.92217815 1.92217815 1.92217815]

l2 [4. 2.7436492 2.32157259 2.14369525 2.05470199 2.00508593
1.97547035 1.95701721 1.94520039 1.93749836 1.93241995 1.92904582
1.92679264 1.9252829 1.92426901 1.92358708 1.92312795 1.92281862
1.92261011 1.92246952 1.92237471 1.92231075 1.92226761 1.92223851
1.92221887 1.92220563 1.92219669 1.92219066 1.92218659 1.92218385
1.922182 1.92218075 1.9221799 1.92217934 1.92217895 1.92217869
1.92217852 1.9221784 1.92217832 1.92217827 1.92217823 1.92217821
1.92217819 1.92217818 1.92217817 1.92217817 1.92217816 1.92217816
1.92217816 1.92217816 1.92217816 1.92217816]

p1 [0.3 0.30167137 0.30727622 0.30689812 0.30625894 0.30588676
0.30570169 0.30561413 0.3055735 0.30555482 0.30554627 0.30554236
0.30554058 0.30553977 0.3055394 0.30553923 0.30553915 0.30553911
0.3055391 0.30553909 0.30553909 0.30553909 0.30553909 0.30553909
0.30553909 0.30553909 0.30553909 0.30553909 0.30553909 0.30553909
0.30553909 0.30553909 0.30553909 0.30553909 0.30553909 0.30553909
0.30553909 0.30553909 0.30553909 0.30553909 0.30553909 0.30553909
0.30553909 0.30553909 0.30553909 0.30553909 0.30553909 0.30553909
0.30553909 0.30553909 0.30553909 0.30553909]

Set of initial parameters (b) with following EM algorithm values with trial 3 data

l1 [0.01 0.88547278 1.24691333 1.48489332 1.63666251 1.73396012
1.79718672 1.8387471 1.86629343 1.88465631 1.89694528 1.90519121
1.91073416 1.91446465 1.91697738 1.91867079 1.91981246 1.92058235
1.92110161 1.92145188 1.92168817 1.92184757 1.92195512 1.92202767
1.92207663 1.92210965 1.92213194 1.92214697 1.92215712 1.92216396
1.92216858 1.92217169 1.9221738 1.92217521 1.92217617 1.92217682
1.92217725 1.92217755 1.92217774 1.92217788 1.92217797 1.92217803
1.92217807 1.9221781 1.92217812 1.92217813 1.92217814 1.92217814
1.92217815 1.92217815 1.92217815 1.92217815]

l2 [10. 2.73578589 2.22982213 2.07504378 2.00910687 1.97516121
1.95573939 1.94392689 1.93647184 1.93165677 1.92850036 1.92641118
1.92501956 1.92408866 1.92346419 1.92304449 1.92276206 1.92257184
1.92244365 1.92235723 1.92229895 1.92225964 1.92223313 1.92221524
1.92220318 1.92219504 1.92218955 1.92218584 1.92218334 1.92218165
1.92218052 1.92217975 1.92217923 1.92217888 1.92217865 1.92217849
1.92217838 1.92217831 1.92217826 1.92217822 1.9221782 1.92217819
1.92217818 1.92217817 1.92217817 1.92217816 1.92217816 1.92217816
1.92217816 1.92217816 1.92217816 1.92217816]

p1 [0.1 0.20255894 0.20303823 0.20010074 0.19873178 0.19815416
0.19790828 0.19780168 0.19775478 0.19773391 0.19772456 0.19772035
0.19771845 0.19771759 0.1977172 0.19771702 0.19771694 0.1977169
0.19771688 0.19771688 0.19771687 0.19771687 0.19771687 0.19771687
0.19771687 0.19771687 0.19771687 0.19771687 0.19771687 0.19771687
0.19771687 0.19771687 0.19771687 0.19771687 0.19771687 0.19771687
0.19771687 0.19771687 0.19771687 0.19771687 0.19771687 0.19771687
0.19771687 0.19771687 0.19771687 0.19771687 0.19771687 0.19771687
0.19771687 0.19771687 0.19771687 0.19771687]

Set of initial parameters (c) with following EM algorithm values with trial 3 data

l1 [0.1 1.55395188 1.66713546 1.71976847 1.7519066 1.77405619
1.7904043 1.80301161 1.81303488 1.82118491 1.82792793 1.83358508
1.83838594 1.84249976 1.84605443 1.84914868 1.85185989 1.85424899
1.856363 1.85823493 1.8598796 1.86128215 1.86237371 1.86299721
1.86292058 1.86196537 1.85939444 1.8541355 1.85163771 1.85092709
1.85073148 1.85067813 1.85066362 1.85065968 1.8506586 1.85065831
1.85065823 1.85065821 1.85065821 1.8506582 1.8506582 1.8506582
1.8506582 1.8506582 1.8506582 1.8506582 1.8506582 1.8506582
1.8506582 1.8506582 1.8506582 1.8506582]

l2 [50. 16.52762557 13.931538 13.16570086 13.02663553
13.21393354 13.61074783 14.15815707 14.82053228 15.57422127
16.40396882 17.3021255 18.26883169 19.31268282 20.45208253
21.71794636 23.15893351 24.85137095 26.91823588 29.56664757
33.16541383 38.41343392 46.71821689 61.04523893 87.86542873
143.88178146 274.38134601 378.5592148 381.92474388 381.93622483
381.93508697 381.93475388 381.93466291 381.93463816 381.93463143
381.9346296 381.9346291 381.93462897 381.93462893 381.93462892
381.93462892 381.93462892 381.93462892 381.93462892 381.93462892
381.93462892 381.93462892 381.93462892 381.93462892 381.93462892
381.93462892 381.93462892]

p1 [0.6 0.78855217 0.84927964 0.87887593 0.89765312 0.91099045
0.92106174 0.92895728 0.9353046 0.94050144 0.94481792 0.94844672
0.95153018 0.95417609 0.95646784 0.95847098 0.9602379 0.96181089
0.96322393 0.96450302 0.96566408 0.96670605 0.96759537 0.96824233
0.96850387 0.96826402 0.96711435 0.96442704 0.96312303 0.96275161
0.96264937 0.96262148 0.9626139 0.96261184 0.96261128 0.96261112
0.96261108 0.96261107 0.96261107 0.96261107 0.96261107 0.96261107
0.96261107 0.96261107 0.96261107 0.96261107 0.96261107 0.96261107

0.96261107 0.96261107 0.96261107 0.96261107]

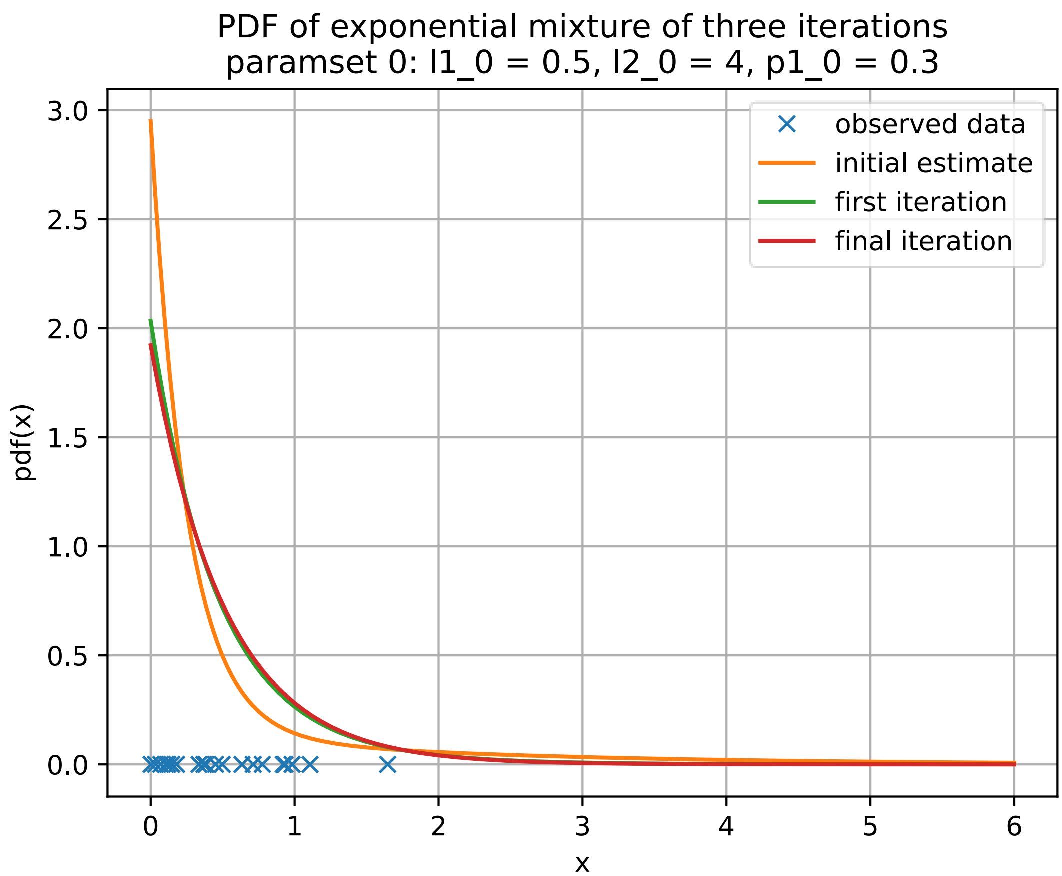


Figure 11: Plot iterations using the initial parameter set (a) with trial 3 data

l1[0] 0.5

l1[1] 1.135307946908389

l1[-1] 1.9221781536691063

l2[0] 4.0

l2[1] 2.7436491972501775

l2[-1] 1.9221781569104766

p1[0] 0.3

p1[1] 0.30167137149960316

p1[-1] 0.3055390856148512

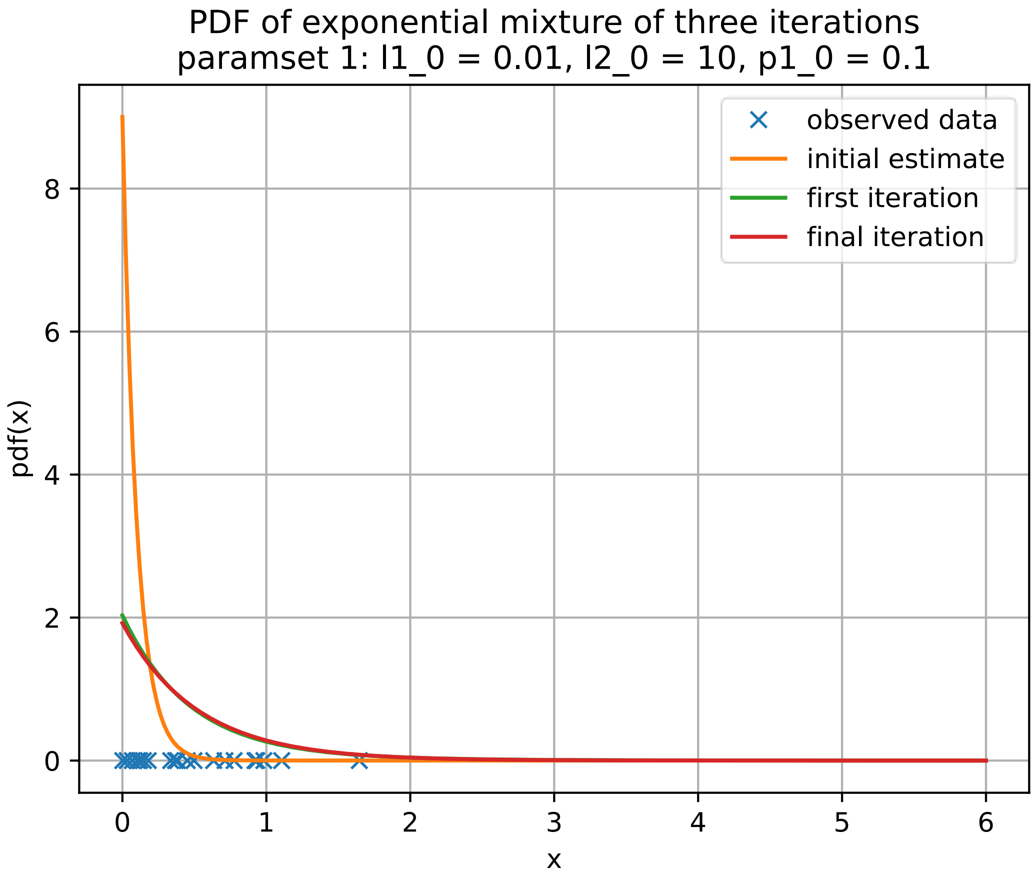


Figure 12: Plot iterations using the initial parameter set (b) with trial 3 data

l1[0] 0.01

l1[1] 0.8854727800628595

l1[-1] 1.9221781534506024

l2[0] 10.0

l2[1] 2.735785885501221

l2[-1] 1.9221781565287033

p1[0] 0.1

p1[1] 0.20255894145550116

p1[-1] 0.19771687050831116

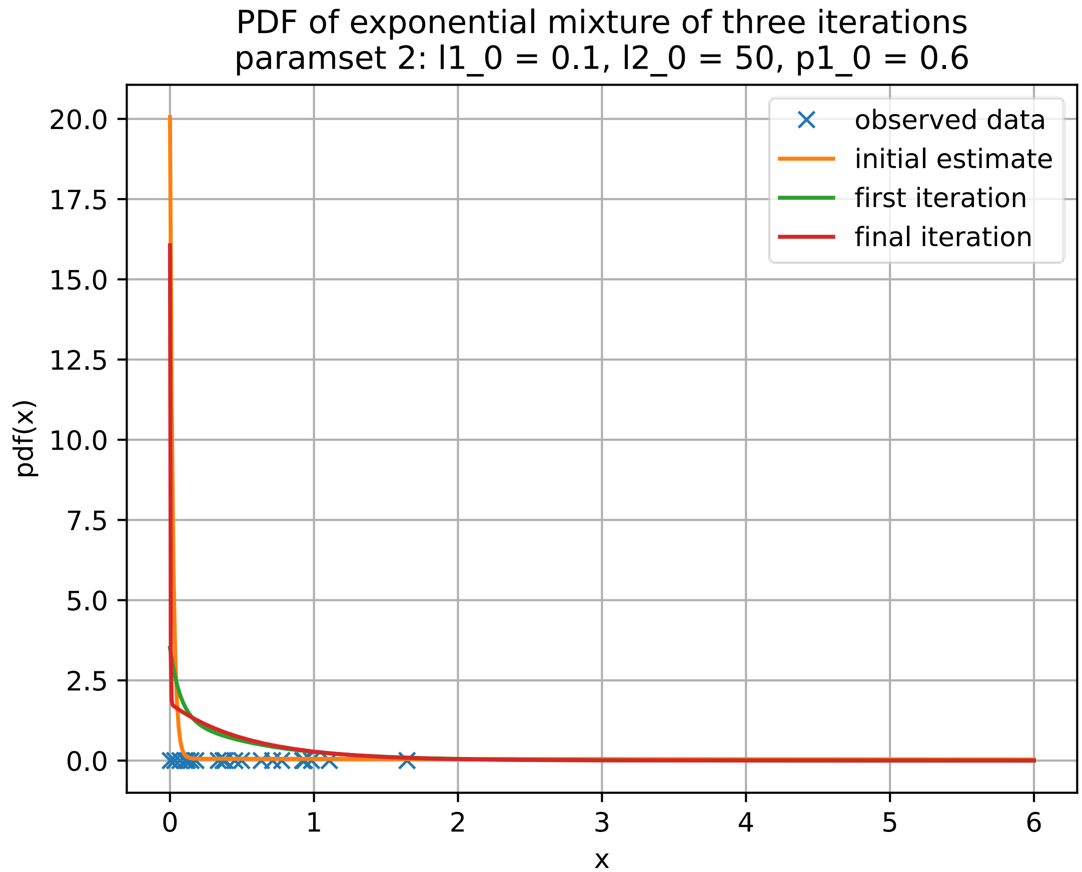


Figure 13: Plot iterations using the initial parameter set (c) with trial 3 data

l1[0] 0.1

l1[1] 1.5539518830554906

l1[-1] 1.850658203398834

l2[0] 50.0

l2[1] 16.527625570085604

l2[-1] 381.9346289150694

p1[0] 0.6

p1[1] 0.7885521685752765

p1[-1] 0.962611067828696

Based on the results, the EM algorithm was successfully derived with the corresponding log-likelihood function increasing monotonically.

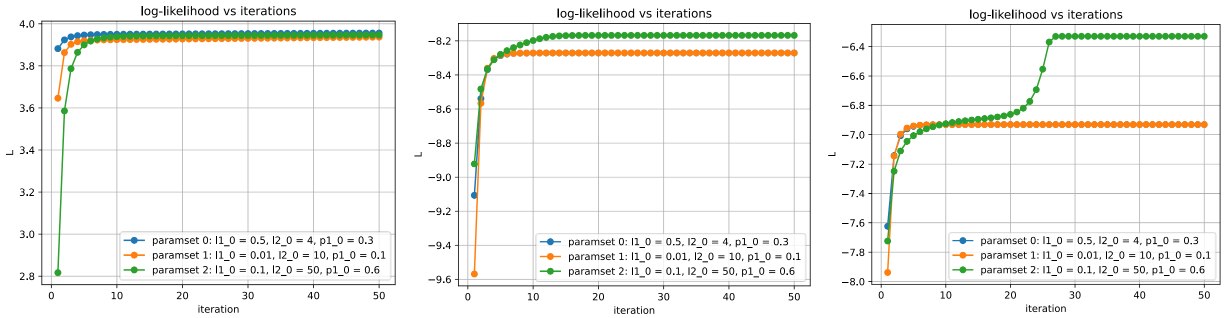


Figure 14: comparison of log-likelihood iterations, trial 1 (left) trial 2 (center) trial 3 (right) sampled data

By comparing the log-likelihood vs iteration plots of the three trial experiments using separate sampled data, we see that two parameter sets converge to the same log-likelihood value while the parameter set [lambda1 = 0.1, lambda2 = 50, pi1 = 0.5] converges to a higher log-likelihood for trial 2 and 3.

This shows evidence that using different initial parameters can lead to different log-likelihood convergence values. Although we cannot be certain that the highest log-likelihood value shown in Figure 14 is the global maximum, we can see that with different initial parameters, the log-likelihood can converge to different local maximums.