Problems 1 and 2:

$$(x_1,\ x_2,\ x_3,\ x_4) \sim \mathcal{M}(n; 0.25\theta,\ 0.25(2+\theta),\ 0.5(1-2\theta),\ 0.5\theta)$$

Log-Likelihood

$$\begin{aligned} l = \sum_{i=1}^{n} log(f(x|\theta)) &= x_1 log(0.25\theta) + x_2 log(.25(2+\theta)) + x_3 log(0.5(1-2\theta)) + x_4 log(0.5\theta) + C_1 \\ &= x_1 log(\theta) + x_2 log(\theta) + x_3 log(2-4\theta) + x_4 log(\theta) + C_2 \end{aligned}$$

Where C_i are constants.
Let $x_2 = y_1 + y_2 \quad s.t.\ y_1 = 0.5,\ y_2 = 0.25\theta$. Then

$$y = (x_1,\ y_1,\ y_2,\ x_3,\ x_4) \sim \mathcal{M}(n; 0.25\theta,\ 0.5,\ 0.25\theta,\ 0.5(1-2\theta),\ 0.5\theta)$$

Then:

$$\begin{aligned} l_y = \sum_{i=1}^{n} log(f(x|\theta)) &= x_1 log(0.25\theta) + y_1 log(0.5) + y_2 log(0.25\theta) \\ &\quad + x_3 log(0.5(1-2\theta)) + x_4 log(0.5\theta) + C_3 \\ l_y &= (x_1 + y_2 + x_4) log(\theta) + x_3 log(1-2\theta) + C_4 \end{aligned}$$

E-Step

$$\begin{aligned} Q(\theta|\theta_K, x) = \mathbb{E}[l_y|\theta_K, x] &= \int_y l_y \mathbb{P}(y|\theta_K,\ x) dy \\ &= \sum_{y=1}^{x_2} [(y_2 + x_1 + x_4) log(\theta) + x_3 log(1-2\theta) + C_4] \mathbb{P}(y_2|\theta_K,\ x) \\ &= (\mathbb{E}[y_2|\theta_K, x] + x_1 + x_4) log(\theta) + x_3 log(1-2\theta) + C_4 \end{aligned}$$

Now, as $(y_1, y_2) \sim Binomial(x_2; \frac{2}{2+\theta_K},\ \frac{\theta_K}{2+\theta_K})$ at the Kth iteration:

$$\begin{aligned} \mathbb{E}[y_2|\theta_K, x] &= \sum_{y_2=0}^{x_2} y_2 \frac{x_2!}{y_2!(x_2-y_2)!} \left(\frac{2}{2+\theta_K}\right)^{x_2-y_2} \left(\frac{\theta_K}{2+\theta_K}\right)^{y_2} \\ &= \sum_{y_2=1}^{x_2} x_2 \frac{(x_2-1)!}{(y_2-1)!(x_2-y_2)!} \left(\frac{2}{2+\theta_K}\right)^{x_2-y_2} \left(\frac{\theta_K}{2+\theta_K}\right)^{y_2} \\ &= \left(\frac{\theta_K}{2+\theta_K}\right) x_2 \sum_{y_2-1=0}^{x_2-1} \frac{(x_2-1)!}{(y_2-1)!((x_2-1)-(y_2-1))!} \left(\frac{2}{2+\theta_K}\right)^{(x_2-1)-(y_2-1)} \left(\frac{\theta_K}{2+\theta_K}\right)^{y_2-1} \\ &= \left(\frac{\theta_K}{2+\theta_K}\right) x_2 \end{aligned}$$

Where the last equality follows from the summation being the CDF of a binomial distribution $F(k;n,p) = \mathbb{P}(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \begin{pmatrix} n \\ i \end{pmatrix} p^{n-i}(1-p)^i$ with $n = x_2 - 1 = k,\ i = y_2 - 1$ and $p = \frac{2}{2+\theta_K}$ over the entire space and is thus equal to one.

$$Q(\theta|\theta_K, x) = \left(\left(\frac{\theta_K}{2+\theta_K}\right) x_2 + x_1 + x_4\right) log(\theta) + x_3 log(1-2\theta) + C_4$$

M-Step

Taking the first-order condition of $Q(\theta|\theta_K, x)$ with respect to θ and setting equal to zero so as to obtain $\theta_{k+1} = \arg\max_{\theta} Q\left(\theta|\theta_K, x_o\right)$ yields:

$$\begin{aligned}
\frac{dQ(\theta|\theta_K, x)}{d\theta} &= \frac{\left(\frac{\theta_K}{2+\theta_K}\right) x_2 + x_1 + x_4}{\theta^*} = \frac{x_3}{0.5 - \theta^*} \\
&= 2\left[\left(\frac{\theta_K}{2+\theta_K}\right) x_2 + x_1 + x_3 + x_4\right] \\
&\Rightarrow \theta^* = \theta_{K+1} = \frac{\left(\frac{\theta_K}{2+\theta_K}\right) x_2 + x_1 + x_4}{2\left[\left(\frac{\theta_K}{2+\theta_K}\right) x_2 + x_1 + x_3 + x_4\right]}
\end{aligned}$$

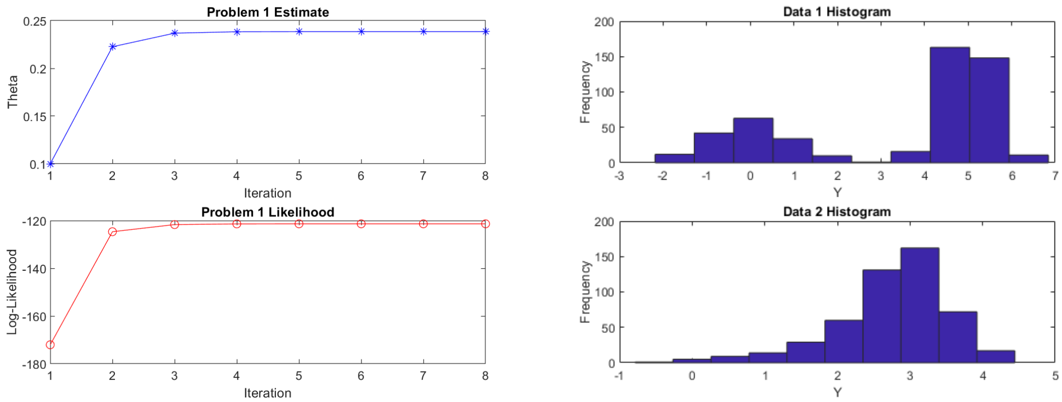
MATLAB Problems 1 and 2

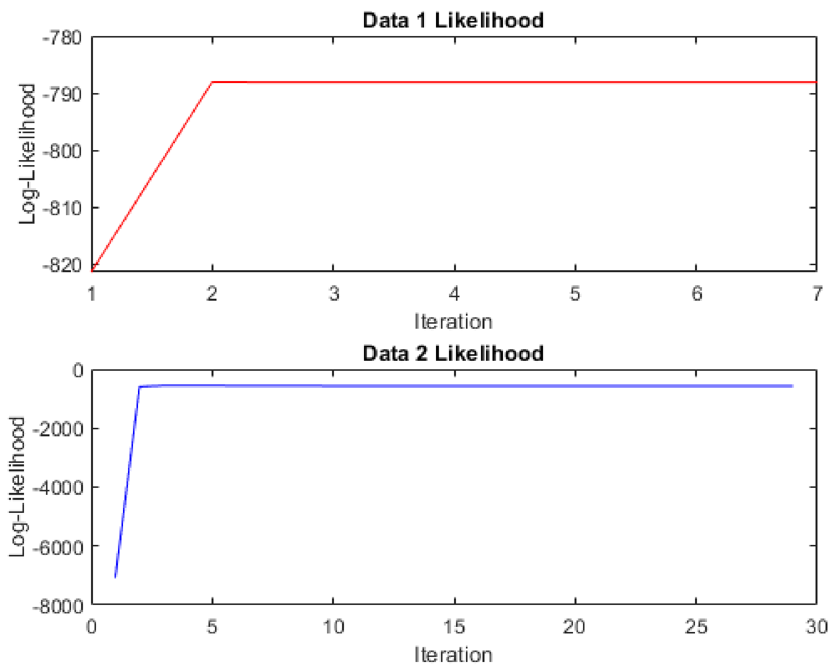
Output:

```
---------------------------------------------------------------
Oscar Martinez          Homework 6: Problems 1 and 2          STA 5106
---------------------------------------------------------------
-----------Problem 1----------

theta =

Columns 1 through 5

0.1000    0.2226    0.2369    0.2384    0.2385

Columns 6 through 8

0.2385    0.2385    0.2385

-----------Problem 2----------
----------Dataset 1---------

m =
```

```
7

mu = (0.023, 4.970), sigma = (0.914, 0.522), alpha = (0.324, 0.676)

----------Dataset 2---------

m =

29

mu = (0.550, 2.818), sigma = (0.722, 0.700), alpha = (0.027, 0.973)
```

Figures:





Code:

```
 clc
clear

%  % Diary
%  dfile ='MATLAB_Output_OM.txt';
%  if exist(dfile, 'file') ; delete(dfile); end
%  diary(dfile)
%  diary on
%  diary MATLAB_Output_OM.txt

%Introduction
fprintf('--------------------------------------------------------------\n');
fprintf('Oscar Martinez \t Homework 6: Problems 1 and 2 \t STA 5106\n');
fprintf('--------------------------------------------------------------\n');

%-----Problem 1:-----
fprintf('-----------Problem 1----------\n');

n = 100;
x = [6 52 28 14];

% EM procedure
theta0 = 0.1;  % initial guess

th = @(x) x/(2+x);
theta(1) = theta0;
nth= theta(1)+1;
i = 1;
while (abs(theta(i)-nth)>1e-6)
nth = theta(i);
theta(i+1) = ( (th(theta(i)))*x(2)+x(4)+x(1) )/ ...
(2*( (th(theta(i)))*x(2)+x(1)+x(3)+x(4) ));
i = i+1;
end

theta

% plot the estimate
figure(1)
subplot(211);
plot(theta, 'b-*');
```

```
xlabel('Iteration');
ylabel('Theta');
title('Problem 1 Estimate')

% compute the log-likelihood
K = length(theta);
for k = 1:K
ll(k) = (x(1)+x(2)+x(4))*log(theta(k)) + x(3)*log(1-2*theta(k));
end

% plot the log-likelihood
subplot(212);
plot(ll, 'ro-');
xlabel('Iteration');
ylabel('Log-Likelihood');
title('Problem 1 Likelihood')

%-----Problem 2:-----
fprintf('------------Problem 2-----------\n');


fprintf('----------Dataset 1----------\n');
clear

load ('hw6_2_data1.mat')
Y1=Y;

n=length(Y);
figure(2)
subplot(211);
hist(Y);
xlabel('Y');
ylabel('Frequency');
title('Data 1 Histogram')

alpha(1,:) = [0.5 0.5];
mu(1,:) = [0 5];
sigma(1,:) = [1 0.5];

u = 1;
m = 1;
while (u > 1e-6)
for i = 1:n
temp = alpha(m,:).*sigma(m,:).^(-1).*exp(-(Y(i)-mu(m,:)).^2./(2*sigma(m,:)));
```

```
P(:,i,m) = temp'./sum(temp);
end
alpha(m+1,:) = mean(P(:,:,m),2)';
mu(m+1,:) = sum(ones(2,1)*Y.*P(:,:,m),2)'./sum(P(:,:,m),2)';
sigma(m+1,:) = sqrt(sum((ones(2,1)*Y-mu(m+1,:)'*ones(1,n)).^2.*P(:,:,m),2)' ...
./sum(P(:,:,m),2)');
u = max([norm(alpha(m+1,:)-alpha(m,:)) norm(mu(m+1,:)-mu(m,:)) ...
norm(sigma(m+1,:)-sigma(m,:))]);

m=m+1;
end

m

% compute the log-likelihood
K = length(alpha);
for k = 1:K
f1 = normpdf(Y,mu(k,1),sigma(k,1));
f2 = normpdf(Y,mu(k,2),sigma(k,2));
ll(k) = sum(log(alpha(k,1)*f1+alpha(k,2)*f2));
end

% plot the log-likelihood
figure(3)
subplot(211);
plot(ll, 'r-');
xlabel('Iteration');
ylabel('Log-Likelihood');
title('Data 1 Likelihood')


fprintf('mu = (%4.3f, %4.3f), sigma = (%4.3f, %4.3f), alpha = (%4.3f, %4.3f) \n', mu(end,1), mu(end,2), sigma(end,1), sigma(end,2), alpha(end,1), alpha(end,2))

fprintf('\n----------Dataset 2----------\n');

load ('hw6_2_data2.mat')
Y2=Y;
n=length(Y);
figure(2)
subplot(212);
hist(Y);
xlabel('Y');
```

```
ylabel('Frequency');
title('Data 2 Histogram')

alpha(1,:) = [0.5 0.5];
mu(1,:) = [1 4];
sigma(1,:) = [0.3 0.1];

u = 1;
m = 1;
while (u > 1e-3)
for i = 1:n
temp = alpha(m,:).*sigma(m,:).^(-1).*exp(-(Y(i)-mu(m,:)).^2./(2*sigma(m,:)));
P(:,i,m) = temp'./sum(temp);
end
alpha(m+1,:) = mean(P(:,:,m),2)';
mu(m+1,:) = sum(ones(2,1)*Y.*P(:,:,m),2)'./sum(P(:,:,m),2)';
sigma(m+1,:) = sqrt(sum((ones(2,1)*Y-mu(m+1,:)'*ones(1,n)).^2.*P(:,:,m),2)' ...
./sum(P(:,:,m),2)');
u = max([norm(alpha(m+1,:)-alpha(m,:)) norm(mu(m+1,:)-mu(m,:)) ...
norm(sigma(m+1,:)-sigma(m,:))]);

m=m+1;
end

m
fprintf('mu = (%4.3f, %4.3f), sigma = (%4.3f, %4.3f), alpha = (%4.3f, %4.3f) \n', mu(end,1), mu(end,2), sigma(end,1), sigma(end,2), alpha(end,1), alpha(end,2))


% compute the log-likelihood
K = length(alpha);
for k = 1:K
f1 = normpdf(Y,mu(k,1),sigma(k,1));
f2 = normpdf(Y,mu(k,2),sigma(k,2));
ll(k) = sum(log(alpha(k,1)*f1+alpha(k,2)*f2));
end

% plot the log-likelihood
figure(3)
subplot(212);
plot(ll, 'b-');
xlabel('Iteration');
```

```
ylabel('Log-Likelihood');
title('Data 2 Likelihood')

diary off
```

Problem 3

```
In [3]: from numpy import *
from matplotlib import pyplot
#Problem 3
print("---Problem 3---")
n = 100
x = [6, 52, 28, 14]

# EM procedure
theta = zeros(20)
theta[0] = 0.1
nth = (theta[0]/(2+theta[0])*x[1] + x[0]+ x[3])/2*((theta[0]/(2+theta[0]))*x[1]+x[0]+x
i = 0;
while (abs(theta[i]-nth)>1e-6):
nth = theta[i]
theta[i+1] = (theta[i]/(2+theta[i])*x[1]+ x[0]+x[3])/2*((theta[i]/(2+theta[i]))*x[1]+x
i = i+1
print(theta)
```

```
---Problem 3---
[1.00000000e-01 9.52562358e+01 2.01906463e+03 2.08465655e+03
2.08476161e+03 2.08476177e+03 2.08476177e+03 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00
0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
```

```
In [ ]:
```