Michael Leibert
Math 611
Homework 6

1. The following data are observations from the random variable $X = (1-W)\cdot Y_1 + W \cdot Y_2$, where W follows a Bernoulli ($p = 0.7$) distribution. The variables: $Y_1 \sim N(100, 20)$ and $Y_2 =\sim N(120, 25)$. W is independent of Y_1 and W is independent of Y_2. Code the EM algorithm for 12 iterations and present your outcomes in 3 columns, for the first and second mean and the mixture probability W respectively.

```r
x<-c(169.14353,135.73850,102.46566,80.91151,148.45425,144.68948,
        106.56257,104.83559,94.81216,109.47048,95.94150,123.84673,
        87.18401,104.73420,111.94364,119.69467,151.77627,81.80692,116.58660,98.28933)

mu_1 <- 100
mu_2 <- 120
sd_1<-sqrt(20)
sd_2<-sqrt(25)
tau_1 <- 0.3
tau_2 <- 1-tau_1
for( i in 1:12 ) {
  T_1 <- tau_1 * dnorm( x, mean=mu_1 ,sd=sd_1 )
  T_2 <- tau_2 * dnorm( x, mean=mu_2 ,sd=sd_2 )
  P_1 <- T_1 / (T_1 + T_2)
  P_2 <- T_2 / (T_1 + T_2) ## note: P_2 = 1 - P_1
  tau_1 <- mean(P_1)
  tau_2 <- mean(P_2)
  mu_1 <- sum( P_1 * x ) / sum(P_1)
  mu_2 <- sum( P_2 * x ) / sum(P_2)
  print( c(mu_1, mu_2, mean(P_1)) )}

## [1]  96.0448210 133.4018332   0.5074671
## [1]  98.206154 138.608392   0.598086
## [1]  99.4960354 141.4485353   0.6436841
## [1] 100.3847821 143.5734476   0.6744609
## [1] 101.0297441 145.4095110   0.6977308
## [1] 101.1785336 145.8058866   0.7027418
## [1] 101.2420837 145.9553753   0.7047344
## [1] 101.2753215 146.0340274   0.7057766
## [1] 101.2949962 146.0808567   0.7063943
## [1] 101.3074620 146.1106444   0.7067861
## [1] 101.3156909 146.1303597   0.7070448
## [1] 101.3212672 146.1437439   0.7072203
```

2. Consider the random variable Y, where $Y = \theta_1 \cdot Z_1^2 + \theta_2 \cdot Z_2^2$ where both Z_1 and Z_2 are standard normal independent random variables and $\theta_1 + \theta_2 = 1$.

a. Show analytically that the variable $W = \theta_1 \cdot Z_1^2$ is distributed as $Gamma\left(\frac{1}{2}, \frac{1}{2\theta_1}\right)$

Derive the mgf of W:

$$\begin{aligned}
M_W(t) &= E\left[e^{tW}\right] \\
&= E\left[\exp\left(t\theta_1 Z_1^2\right)\right] \\
&= \int_{-\infty}^{\infty} \exp\left[t\theta_1\ z_1^2\right]\ \frac{1}{\sqrt{2\pi}}\ \exp\left[-\frac{1}{2}\ z^2\right]\ \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[t\theta_1 z_1^2 - \frac{1}{2}\ z^2\right]\ \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-z_1^2\left(-t\theta_1 + \frac{1}{2}\right)\right]\ \mathrm{d}z \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\Big(-az_1^2\Big)\ \mathrm{d}z \\
&= \frac{1}{\sqrt{2\cancel{\pi}}}\frac{\sqrt{\cancel{\pi}}}{\sqrt{a}} \\
&= \frac{1}{\sqrt{2}\ \sqrt{-t\theta_1 + \frac{1}{2}}} \\
&= \frac{1}{\sqrt{1-2t\theta_1}}
\end{aligned}$$

let $a = \left(-t\theta_1 + \dfrac{1}{2}\right)$

Note: $\displaystyle\int_{-\infty}^{\infty} \exp\Big(-az_1^2\Big) = \frac{\sqrt{\pi}}{\sqrt{a}}$

This is the Gaussian integral from HW3 #3

If $X \sim Gamma(n,k)$, the Mgf is:

$$\begin{aligned}
M_X(t) &= \int_0^{\infty} \exp(tx)\ \frac{k^n}{\Gamma(n)}\ x^{n-1}\ \exp(-kx)\ \mathrm{d}x \\
&= \frac{k^n}{\Gamma(n)} \int_0^{\infty} \exp\Big[-x\ (-t+k)\Big]\ x^{n-1}\ \mathrm{d}x \\
&= \frac{k^n}{\Gamma(n)} \int_0^{\infty} e^{-u}\ \cancel{\frac{1}{k-t}}\ u^{n-1}\ \left(\frac{1}{k-t}\right)^{n-\cancel{1}}\ \mathrm{d}u \\
&= \frac{k^n}{\Gamma(n)} \left(\frac{1}{k-t}\right)^n \underbrace{\int_0^{\infty} e^{-u}\ u^{n-1}\ \mathrm{d}u}_{\Gamma(n)} \\
&= \frac{1}{\cancel{\Gamma(n)}} \left(\frac{k}{k-t}\right)^n \cancel{\Gamma(n)} \\
&= \left(\frac{k}{k-t}\right)^n \qquad t < k.
\end{aligned}$$

$$\begin{aligned}
x &= \frac{u}{(k-t)} \\
u &= x(k-t) \\
\mathrm{d}u &= (k-t)\ \mathrm{d}x \\
\frac{1}{k-t}\ \mathrm{d}u &= \mathrm{d}x
\end{aligned}$$

Thus, if $Y \sim Gamma\left(\dfrac{1}{2}, \dfrac{1}{2\theta_1}\right)$

$$M_Y(t) = \left(\frac{\frac{1}{2\theta_1}}{\frac{1}{2\theta_1} - t}\right)^{\frac{1}{2}} = \left(\frac{\frac{1}{2\theta_1}}{\frac{1-2t\theta_1}{2\theta_1}}\right)^{\frac{1}{2}} = \left(\frac{1}{\cancel{2\theta_1}}\ \frac{\cancel{2\theta_1}}{1-2t\theta_1}\right)^{\frac{1}{2}} = \frac{1}{\sqrt{1-2t\theta_1}}$$

Therefore, $M_Y(t) = M_W(t)$ and they are the same distribution.

b. Use some numerical approach to derive the values of the MLE for the unknown parameters θ_1 and θ_2.

Consider if we set $\theta_1 = 0.4$. We generate estimates from a convolution of two Gamma RVs. Define: $Y = W_1 + W_2$.

```
rm(list = ls())
theta<- 0.4
W1<-rgamma(10000,shape=.5,scale= 1/(2*(1-theta)) )
W2<-rgamma(10000,shape=.5,scale= 1/ (2*(theta )) )
y <- W1+W2
```

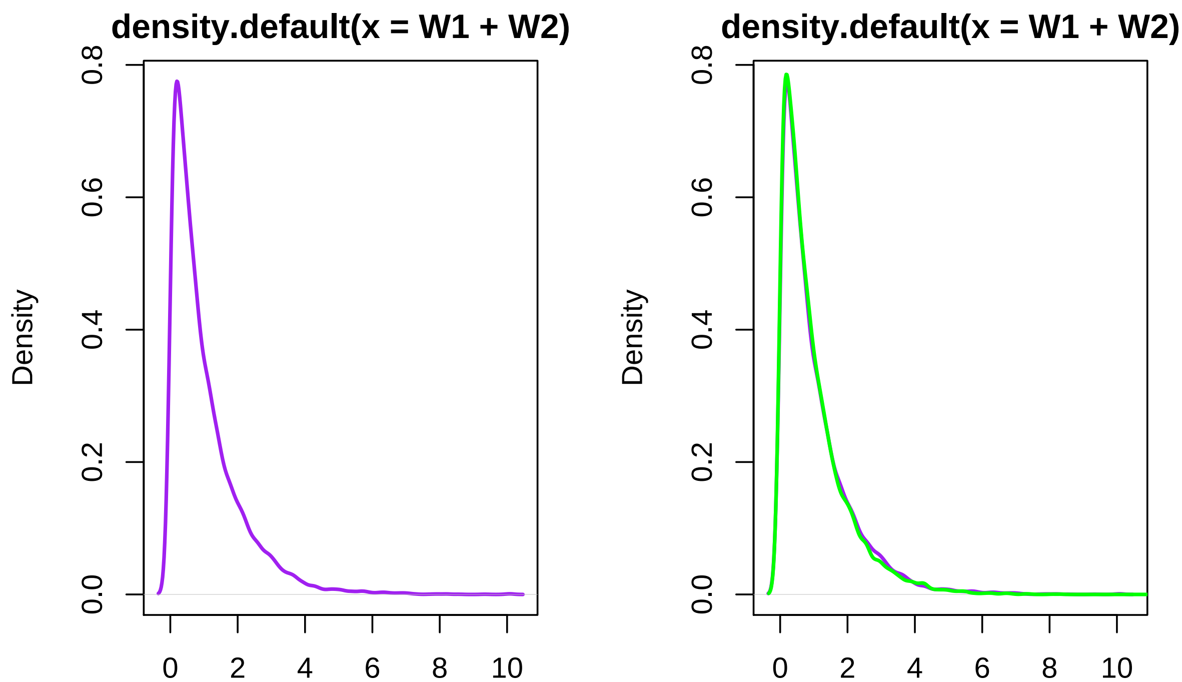
We then write the log-likehood function and attempt to optimize it.

```
f <- function(mu, x) {    sum(-dgamma(x,mu[1],mu[2],log = T))      }
out <- optim(par = c(1,1), fn=f,x=y,method = "L-BFGS-B" ,lower=c(0,0))
out

## $par
## [1] 0.9892464 0.9725908
##
## $value
## [1] 10169.43
##
## $counts
## function gradient
##        7        7
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
```

The MLE estimates are both around 1, so it appears that $Y \sim Gamma(.98, .97)$. We can plot the densities of our sum of gammas and compare it to a $Gamma(.98, .97)$ density.

```
par(mar=c(2.1, 4.1, 2, 2.1),mfrow=c(1, 2))
plot(density( W1+W2 ) , col="purple",lwd=2)
plot(density( W1+W2 ) , col="purple",lwd=2)
lines(density(rgamma(10000,shape=0.9892464,scale=0.9725908) ) ,col="green",lwd=2)
```



The image on the left is only $Y = W_1 + W_2$, which had to be shown because the $Gamma(.98, .97)$ almost completely covers it.

For the MLE of θ_1 and θ_2,

$$0.9725908 = \frac{1}{2\theta_1}$$
$$1.945182 = \frac{1}{\theta_1}$$
$$0.5140908 = \theta_1 \qquad 0.4859092 = \theta_2$$

c. Demonstrate an EM algorithm that would approximate the values of the unknown parameters θ_1 and θ_2.

Suppose $\theta = 0.4$ and the latent Bernoulli variable that assigns group has probability $\tau = 0.6$ We generate group assignment.

```
rm(list = ls())
theta<- 0.4
tau <- 0.6
x <- y <- rep(0,10000)
for( i in 1:10000) {
 if( runif(1) < tau ) {
   x[i] <- rgamma(1,shape=.5,scale= 1/(2*(1-theta)) )
   y[i] <- "heads"
 } else {
   x[i] <- rgamma(1,shape=.5,scale= 1/ (2*(theta)) )
   y[i] <- "tails"  }}
```

We run the EM algorithm with initial guesses for scale and group assignment.

```
sc_1 <- .7

tau_1 <- 0.5
tau_2 <- 0.5

for( i in 1:25 ) {
  T_1 <- tau_1 * rgamma(x,shape=.5,scale= 1/(2*(1-sc_1)) )
  T_2 <- tau_2 * rgamma(x,shape=.5,scale=  1/ (2*(sc_1 )) )

  P_1 <- T_1 / (T_1 + T_2)
  P_2 <- T_2 / (T_1 + T_2)

  tau_1 <- mean(P_1)
  tau_2 <- mean(P_2)
  sc_1 <- sum( P_1 * x ) / sum(P_1)
  print( c(sc_1 , 1-sc_1 ) )}

## [1] 0.4990045 0.5009955
## [1] 0.5006477 0.4993523
## [1] 0.5021931 0.4978069
## [1] 0.5023254 0.4976746
## [1] 0.5027912 0.4972088
## [1] 0.5010614 0.4989386
## [1] 0.4976797 0.5023203
## [1] 0.5005366 0.4994634
## [1] 0.5074799 0.4925201
## [1] 0.5044033 0.4955967
## [1] 0.4944124 0.5055876
## [1] 0.4887653 0.5112347
## [1] 0.4946428 0.5053572
## [1] 0.5010352 0.4989648
## [1] 0.5038926 0.4961074
## [1] 0.5002886 0.4997114
## [1] 0.4986313 0.5013687
## [1] 0.4967634 0.5032366
## [1] 0.4983397 0.5016603
## [1] 0.5012589 0.4987411
## [1] 0.5057447 0.4942553
## [1] 0.5102476 0.4897524
## [1] 0.5060056 0.4939944
## [1] 0.4982798 0.5017202
## [1] 0.5043396 0.4956604
```

We see that our EM-algorithm estimates are close to the numerical approach we took in 2b.

3. An office has three machines that each break with probability 0.1 each day, but when there is at least one broken, then there is probability 0.5 that the repairman can fix one of them for use the next day. If we ignore the possibility of two machines breaking on the same day.

a. Determine the transition probability matrix that models the number of working machines.

$$
\begin{array}{c c}
 & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\
\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.05 & 0.5 & 0.45 & 0 \\ 0 & 0.1 & 0.5 & 0.4 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}
\end{array}
$$

At state zero, there is probability 0.5 that the repairman can fix one of the machines.

$$\begin{aligned}&\text{to state 0: } P(fail) = 0.5\\&\text{to state 1: } P(fix) = 0.5\end{aligned}$$

At state one, $P(break|1 \text{ active machine}) = 0.1$ and $P(working|1 \text{ active machine}) = 0.9$

$$\begin{aligned}&\text{to state 0: } P(break|1) \cap P(fail) = 0.1 \cdot 0.5 = 0.05\\&\text{to state 1: } P(break|1) \cap P(fix) \cup P(working|1) \cap P(fail) = 0.1 \cdot 0.5 + 0.9 \cdot 0.5 = 0.5\\&\text{to state 2: } P(working|1) \cap P(fix) = 0.9 \cdot 0.5 = 0.45\end{aligned}$$

At state two, $P(\text{Machine 1 breaks OR Machine 2 breaks}) = P(break|2) = 0.2$; $P(working|2) = 0.8$

$$\begin{aligned}&\text{to state 1: } P(break|2) \cap P(fail) = 0.2 \cdot 0.5 = 0.1\\&\text{to state 2: } P(break|2) \cap P(fix) \cup P(working|2) \cap P(fail) = 0.2 \cdot 0.5 + 0.8 \cdot 0.5 = 0.5\\&\text{to state 3: } P(working|2) \cap P(fix) = 0.8 \cdot 0.5 = 0.4\end{aligned}$$

At state three, $P(\text{Machine 1 breaks OR Machine 2 breaks OR Machine 3 breaks}) = P(break|3) = 0.3$; $P(working|3) = 0.7$

b. Using a recursive formula, derive the stationary distribution.

$$\text{if } \pi(0) = c \text{ then}$$

$$\begin{aligned}\pi(1) &= \pi(0) \cdot \frac{p_0}{q_1} = c \cdot \frac{0.5}{0.05} = 10c\\ \pi(2) &= \pi(1) \cdot \frac{p_1}{q_2} = 10c \cdot \frac{0.45}{0.1} = 45c\\ \pi(3) &= \pi(2) \cdot \frac{p_2}{q_3} = 45c \cdot \frac{0.4}{0.3} = 60c\end{aligned}$$

The sum of the $\pi's$ is $116c$, so if we let $c = \frac{1}{116}$ then we get

$$\pi(3) = \frac{60}{116} \quad \pi(2) = \frac{45}{116}, \quad \pi(1) = \frac{10}{116}, \quad \pi(0) = \frac{1}{116}$$

4. In the Ehrenfest chain example presented in class, show that the binomial stationary model satisfies the detailed balanced condition.

$$\begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 3/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 2/3 & 0 & 1/3 \\ 0 & 0 & 3/3 & 0 \end{pmatrix} \end{array}$$

Setting $\pi(0) = c$ we have

$$\pi(1) = 3c, \quad \pi(2) = \pi(1) = 3c \quad \pi(3) = \frac{\pi(2)}{3} = c$$

The sum of the π's is $8c$, so we pick $c = \frac{1}{8}$ to get

$$\pi(0) = \frac{1}{8}, \quad \pi(1) = \frac{3}{8} \quad \pi(2) = \frac{3}{8} \quad \pi(3) = \frac{1}{8}$$

We see in general that the binomial distribution with $p = \frac{1}{2}$ is the stationary distribution:

$$\pi(x) = \frac{1}{2^n}\binom{n}{x}.$$

To verify that this satisfied the detailed balance condition:

$$\begin{aligned}
\pi(x)p(x, x+1) &= 2^{-n}\ \frac{n!}{x!(n-x)!} \cdot \frac{n-x}{n} \\
&= 2^{-n}\ \frac{n!}{(x+1)!(n-x-1)!} \cdot \frac{x+1}{n} \\
&= \pi(x+1)p(x+1, x)
\end{aligned}$$