Monte Carlo Optimization

Jonathan Navarrete

June 3, 2017

Introduction

This section will cover topics to optimization problems, and solutions Monte Carlo methods provide. Topics covered include

1. Stochastic Search and Simulated Annealing
2. EM Algorithm and MC EM

Light bulb example

This exercise is taken from Flury and Zoppe, 2000, see Exercises in EM.

Below is the setup for the first exercise.

The First Exercise

Suppose there are two light bulb survival experiments. In the first, there are N bulbs whose exact lifetimes y_i for $i \in \{1, ..., N\}$ are recorded. The lifetimes have an exponential distribution, such that $y_i \sim Exp(\theta)$. In the second experiment, there are M bulbs. After some time $t > 0$, a researcher walks into the room and only records how many lightbulbs are still burning out of M bulbs. Depending on whether the lightbulbs are still burning or out, the results from the second experiment are right- or -left-censored. There are indicators $E_1, ..., E_M$ for each of the bulbs in the second experiment. If the bulb is still burning, $E_i = 1$, else $E_i = 0$.

Given this information, our task is to solve for an MLE estimator for θ.

Our first step in solving this is finding the joint likelihood for the observed and unobserved data (i.e. complete-data likelihood).

Let $X_1, ..., X_M$ be the (unobserved) lifetimes for the second experiment, and let $Z = \sum_{i=1}^{M} E_i$ be the number of light bulbs still burning. Thus, the observed data from both the experiments combined is $\mathcal{Y} = (Y_1, ..., Y_N, E_1, ..., E_M)$ and the unobserserved data is $\mathcal{X} = (X_1, ..., X_M)$.

The complete data log-likelihood is obtained by

$$\begin{aligned} L(\theta|X, Y) &= \prod_{i=1}^{N} \frac{1}{\theta} e^{y_i/\theta} \times \prod_{i=1}^{M} \frac{1}{\theta} e^{x_i/\theta} \\ &= \theta^{-N} e^{-N\bar{y}/\theta} \times \theta^{-M} e^{-\sum_{i=1}^{M} x_i/\theta} \end{aligned}$$

And log-likelihood is obtained by

$$\begin{aligned} log(L(\theta)) &= -N \times log(\theta) - N\bar{y}/\theta - M \times log(\theta) + \sum_{i=1}^{M} x_i/\theta \\ &= -N(log(\theta) + \bar{y}/\theta) - M \times log(\theta) + \sum_{i=1}^{M} x_i/\theta \end{aligned}$$

Or as written by Flury and Zoppe,

$$log^c(L(\theta|\mathcal{Y},\mathcal{X})) = -N(log(\theta)+\bar{Y}/\theta) - \sum_{i=1}^{M}(log(\theta)+X_i/\theta)$$

The next step, is to take the expectation of $log(L(\theta))$ with respect to observed data.

$$\begin{aligned}
E[log(L(\theta))|\mathcal{Y},\mathcal{X}] &= E[-N(log(\theta)+\bar{Y}/\theta) - \sum_{i=1}^{M}(log(\theta)+X_i/\theta)|\mathcal{Y},\mathcal{X}] \\
&= -N(log(\theta)+\bar{Y}/\theta) - E[\sum_{i=1}^{M}(log(\theta)+X_i/\theta)|\mathcal{Y},\mathcal{X}] \\
&= -N(log(\theta)+\bar{Y}/\theta) - M \times log(\theta) + E[\frac{1}{\theta}\sum_{i=1}^{M}X_i|\mathcal{Y},\mathcal{X}] \\
&= -N(log(\theta)+\bar{Y}/\theta) - M \times log(\theta) + \frac{1}{\theta}\sum_{i=1}^{M}E[X_i|\mathcal{Y},\mathcal{X}] \\
&= -N(log(\theta)+\bar{Y}/\theta) - M \times log(\theta) + \frac{1}{\theta}\sum_{i=1}^{M}E[X_i|E_i]
\end{aligned}$$

which is linear for unobserved X_i. But

(2)

$$E[X_i|\mathcal{Y}] = E[X_i|E_i] = \begin{cases} t+\theta & \text{if } E_i = 1 \\ \theta - t\frac{e^{-t/\theta}}{1-e^{-t/\theta}} & \text{if } E_i = 0 \end{cases}$$

For the first case, $E_i = 1$, so

$$\begin{aligned}
E[x_i|x_i > t] &= E[x_i + t] \\
&= t + E[x_i] \\
&= t + \theta
\end{aligned}$$

For the second case, $E_i = 0$, then

$$\int_0^t P(X_i > x|X_i < t)\ dx = \int_0^t \frac{P(x < X_i < t)}{P(X_i < t)}\ dx$$

For the denominator, we get

$$\begin{aligned}
P(X_i < t) &= \int_0^t \frac{1}{\theta}e^{-x_i/\theta}dx \\
&= \frac{1}{\theta}(-\theta e^{-x_i/\theta})|_0^t \\
&= 1 - e^{-t/\theta}
\end{aligned}$$

and for the numerator we obtain

$$\begin{aligned}
P(x < X_i < t) &= \int_x^t \frac{1}{\theta}e^{-x_i/\theta}dx \\
&= \frac{1}{\theta}(-\theta e^{-x_i/\theta})|_0^t \\
&= e^{-x/\theta} - e^{-t/\theta}
\end{aligned}$$

Altogether, we obtain

$$
\begin{aligned}
\int_0^t P(X_i > x|X_i < t)\ dx &= \int_0^t \frac{P(x < X_i < t)}{P(X_i < t)}\ dx \\
&= \int_0^t \frac{e^{-x/\theta} - e^{-t/\theta}}{(1 - e^{-t/\theta})}\ dx \\
&= \frac{1}{(1 - e^{-t/\theta})} \int_0^t (e^{-x/\theta} - e^{-t/\theta})\ dx \\
&= \frac{1}{(1 - e^{-t/\theta})} (\int_0^t e^{-x/\theta} - \int_0^t e^{-t/\theta}\ dx) \\
&= \frac{1}{(1 - e^{-t/\theta})} (\theta(1 - e^{-t/\theta}) - x \times e^{-t/\theta}|_0^t) \\
&= \theta - t \times \frac{e^{-t/\theta}}{1 - e^{-t/\theta}}
\end{aligned}
$$

In order to calculate EM esimates for θ, we will plug in the expected values

$$
E[X_i|\mathcal{Y}] = E[X_i|E_i] = \begin{cases} t + \theta & \text{if } E_i = 1 \\ \theta - t\frac{e^{-t/\theta}}{1-e^{-t/\theta}} & \text{if } E_i = 0 \end{cases}
$$

into the log-likelihood

$$
\begin{aligned}
\log(L(\theta)) &= -N(log(\theta) + \bar{y}/\theta) - M \times log(\theta) + \sum_{i=1}^{M} x_i/\theta \\
&= -N \times log(\theta) - N\bar{y}/\theta - M \times log(\theta) + \sum_{i=1}^{M} x_i/\theta \\
&= -(N + M) \times log(\theta) - N\bar{y}/\theta + \sum_{i=1}^{M} x_i/\theta \\
&= -(N + M) \times log(\theta) - \frac{1}{\theta}(N\bar{y} + \sum_{i=1}^{M} x_i) \\
&= -(N + M)log(\theta) - \frac{1}{\theta}\big[N\bar{Y} + Z(t + \theta) + (M - Z)\big(\theta - t \times \frac{e^{-t/\theta}}{1 - e^{-t/\theta}}\big)\big]
\end{aligned}
$$

As we iterate through estimates of θ, we will use conditioned estimates of θ given previous estimates of θ. Such that the jth step consists of replacing X_i in (1) by its expected value (2), using the current numerical parameter value $\theta^{(j-1)}$.

(3)

$$
\log(L(\theta)) = -(N + M)log(\theta) - \frac{1}{\theta}[N\bar{Y} + Z(t + \theta^{(j-1)}) + (M - Z)(\theta^{(j-1)} - tp^{(j-1)})]
$$

where

$$
p^{(j)} = \frac{e^{-t/\theta^{(j)}}}{1 - e^{-t/\theta^{(j)}}}
$$

Once we take the derivative of the log-likelihood and set it to zero, we will come up with an estimate for θ

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} ln(L(\theta)) &= 0 \\
0 &= -\frac{(N+M)}{\theta} + \frac{1}{\theta^2}\big[N\bar{Y} + Z(t+\theta) + (M-Z)\big(\theta - t \times \frac{e^{-t/\theta}}{1-e^{-t/\theta}}\big)\big] \\
\frac{(N+M)}{\theta} &= \frac{1}{\theta^2}\big[N\bar{Y} + Z(t+\theta) + (M-Z)\big(\theta - t \times \frac{e^{-t/\theta}}{1-e^{-t/\theta}}\big)\big] \\
\theta &= \big[N\bar{Y} + Z(t+\theta) + (M-Z)\big(\theta - t \times \frac{e^{-t/\theta}}{1-e^{-t/\theta}}\big)\big] \ / \ (N+M)
\end{aligned}
$$

Thus, for each jth M-step, we will calculate

$$
\begin{aligned}
\theta^{(j)} &= f(\theta^{(j-1)}) \\
\theta &= \big[N\bar{Y} + Z(t+\theta^{(j-1)}) + (M-Z)\big(\theta^{(j-1)} - t \times \frac{e^{-t/\theta^{(j-1)}}}{1-e^{-t/\theta^{(j-1)}}}\big)\big] \ / \ (N+M)
\end{aligned}
$$

```
set.seed(5678)
theta = 5
rate = 1/theta

t = 5
N = 100
M = 50
y = rexp(n = N, rate = rate)
x = rexp(n = M, rate = rate)
x = sort(x)
E = as.integer(x > t)

N.ybar = sum(y)
Z = sum(E)
t = 5

theta.j = 0.1
theta.jp1 = 0.5
for(i in 1:10){
  theta.j = theta.jp1
  p = (exp(-t/theta.j)/(1-exp(-t/theta.j)))
  theta.jp1 = (N.ybar + Z*( t + theta.j) + (M-Z)*(theta.j - t*p) ) / (N+M)
  print(theta.jp1)
}
```

```
## [1] 4.624345
## [1] 5.366158
## [1] 5.445061
## [1] 5.45323
## [1] 5.454073
## [1] 5.45416
## [1] 5.454169
## [1] 5.45417
## [1] 5.45417
## [1] 5.45417
```

```
## compare against MLE from observed data
mean(y)
```

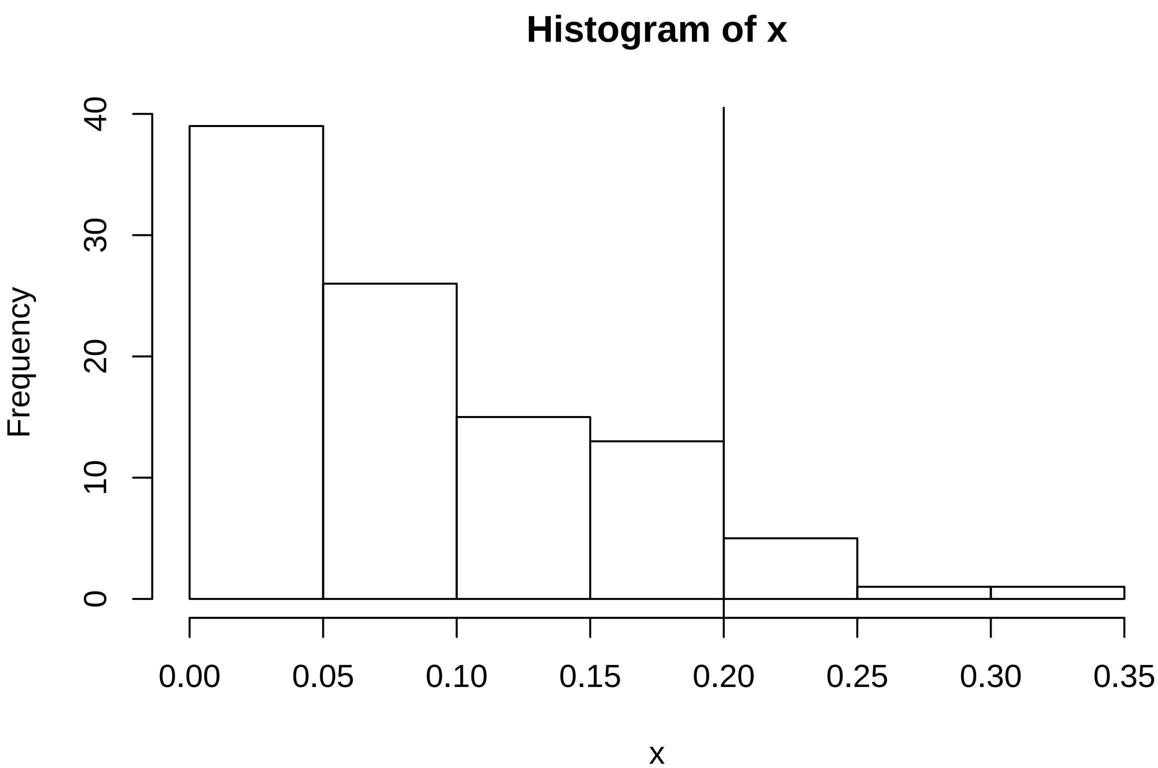
```
## [1] 6.036602
```

```
## note, results will vary if you remove seed
```

Censored Exponential Data

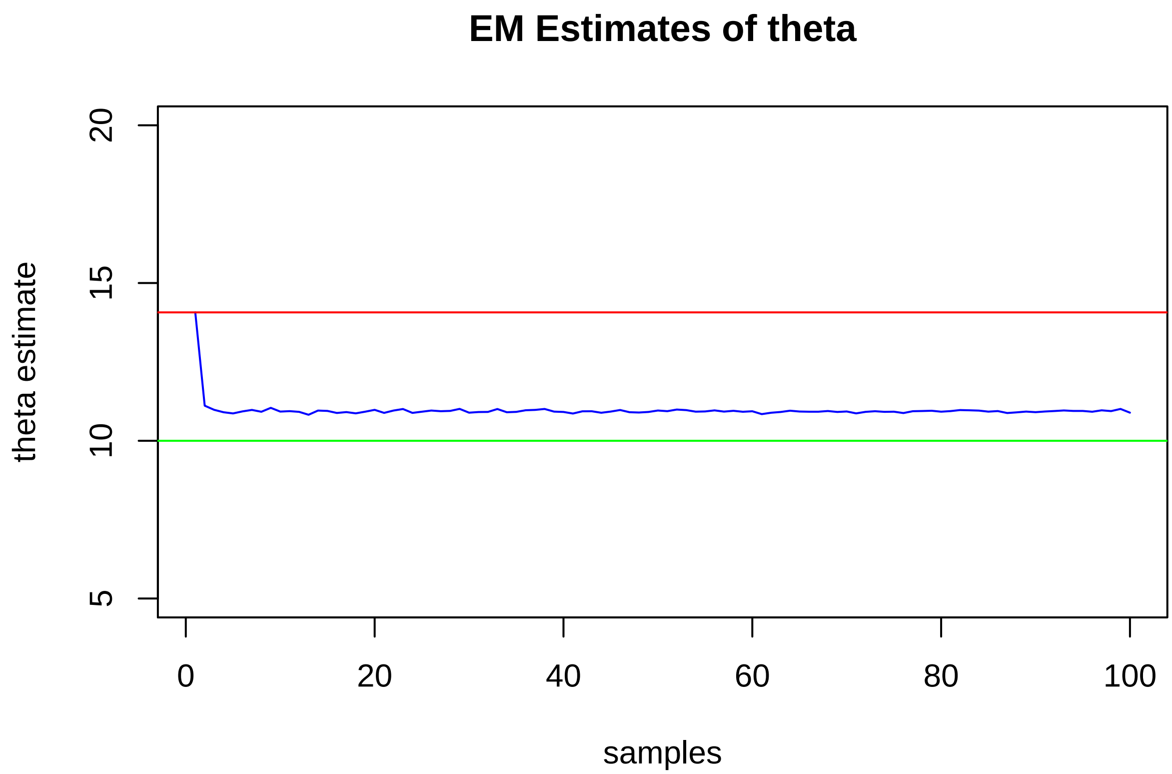
The following is an example from *Computational Statistics* by Givens and Hoeting. Example 4.7.

```
set.seed(456789)
truetheta=10
t = 0.2 ##censoring time
n = 100
x = round(rexp(n, truetheta), 4) #R uses rate
hist(x)
abline(v = t)
```




```
## Determine observed vs. missing data
y = x[x < t]
y=sort(y)
r=sum(x < t) ## observed data
m = n-r



N = 100 ## number of iterations
M = 20
new_theta = n/sum(y)
diff = 1
results = numeric(N)
for(i in 1:N){
```

```r
  results[i] = new_theta
  old_theta = new_theta
  #print(old_theta)
  Y = matrix(data = rep(x = y, M), nrow = length(y), ncol = M)
  Z = t + rexp(n = M * m, rate = old_theta)
  Z = matrix(data = Z, nrow = m, ncol = M)
  X = rbind(Y,Z)
  new_theta = n/mean(apply(X,2, sum))
  #new_theta = n/mean(colSums(X))
  M = M + 1
  #print(new_theta)
}


plot(results, main = "EM Estimates of theta",
     type = "l", col = "blue", ylim = c(5, 20),
     xlab = "samples", ylab = "theta estimate")
abline(h = truetheta, col = "green")
abline(h = n/sum(y), col = "red") ## MLE estimate of observed data
```




```r
#thetaMLE= 1/(mean(y[y < t]) + (n-r)*t/r)
```

MCEM implementation

```r
M = 20
y_observed = y[y < t]
new_est = mean(y_observed)
results = numeric(200)
for(i in 1:200){
  old_est = new_est
```

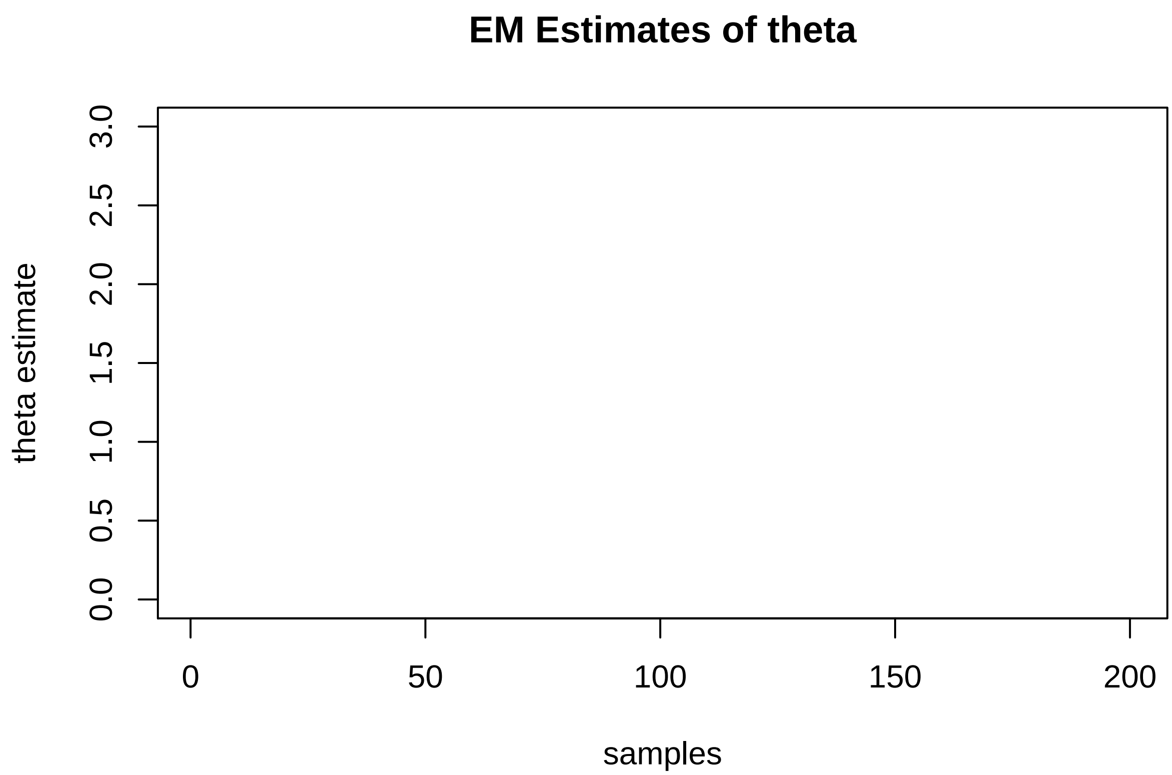
```
  samples = rexp(n = m * M, rate = t + old_est)
  Y = matrix(data = rep(y_observed, M), nrow = length(y_observed), ncol = M)
  Z = matrix(data = samples, nrow = m, ncol = M)
  complete = rbind(Y,Z)
  new_est = 1/mean(colMeans(complete))
  #print(new_est)
  results[i] = new_est
  M = M + 5
}

print("tail sample:")
```

```
## [1] "tail sample:"
```

```
print(tail(results))
```

```
## [1] 13.09676 13.09683 13.10563 13.10617 13.09906 13.09419
```

```
plot(results, main = "EM Estimates of theta", type = "l",
     xlab = "samples", ylab = "theta estimate",
     ylim = c(0, 3))
abline(h = theta, col = "red")
```




```
set.seed(4567)
N = 50
theta = 2
c = 0.5
y = sort(rexp(n = N, rate = theta))
bigC = sum(y> c)
print(sum(y> c))
```

```
## [1] 19
```

```
print("MLE:")
```

```
## [1] "MLE:"
```

```
print(1/mean(y))
```

```
## [1] 1.862943
```

```
indices = which(y>c) ## which indices show where the data is to be censored?



y_MC = y
y_MC[indices] = NA
theta.tp1 = 0.1
MCEMout = numeric(20)
for(i in 1:2000){
  theta.t = theta.tp1 ## update the conditional parameter
  temp = rexp(n = bigC, rate = theta.t )
  y_MC[indices] = temp
  #theta.tp1 = 1/mean(y_MC, na.rm = TRUE) ## doesn't work!
  theta.tp1 = N / (sum(y_MC) + (bigC/theta.t))
  MCEMout[i] = theta.tp1
  #print(theta.tp1)
}

print(mean(MCEMout))
```

```
## [1] 1.888438
```

Another Exponential Distribution problem

Suppose $x_1, ..., x_n \sim Exp(\theta)$ where $x_1, ..., x_n$ are ordered (sorted). After time t, the data has become censored; m observations are censored. Only r are observed such that n - m = r. Let $y = (x_1, .., x_r)^T$ be the observed data and $z = (x_i, ..., x_m)^T$ be unobserved (censored) data.

Our likelihood function is then

$$
\begin{aligned}
L(\theta|Y, Z) &= \prod_{i=1}^{r} \frac{1}{\theta} e^{y_i/\theta} \times \prod_{i=1}^{m} \frac{1}{\theta} e^{z_i/\theta} \\
&= \theta^{-r} e^{-r\bar{y}/\theta} \times \theta^{-m} e^{-\sum_{i=1}^{m} z_i/\theta}
\end{aligned}
$$

The log-likelihood is

$$
\begin{aligned}
ln(L(\theta|Y, Z)) &= -r \times ln(\theta) - r\bar{y}/\theta \times m \times ln(\theta) - \sum_{i=1}^{m} z_i/\theta \\
&= Q(\theta)
\end{aligned}
$$

We find the conditional expectation of z_i given the observed data, knowing that $z_i \sim Exp(\theta - t)$, so by the memoryless property,

$$E[z_i|y] = E[x_i|x_i > t] = \theta + t$$

We substitute the conditional expectation into the log-likelihood and obtain,

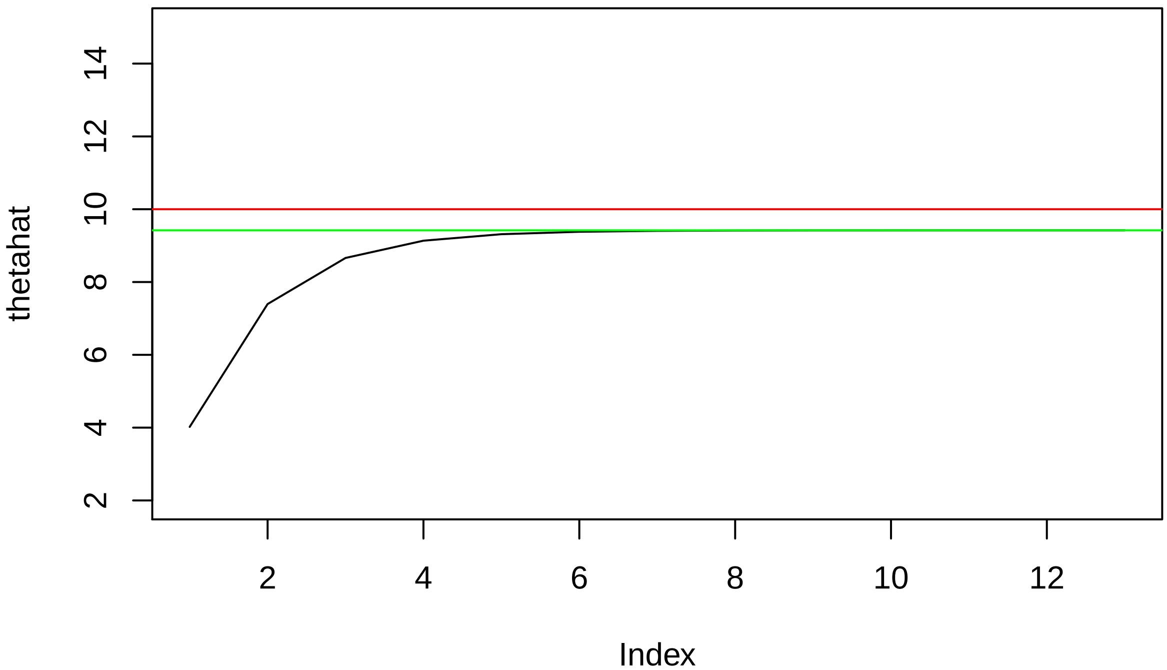
$$
\begin{aligned}
Q(\theta_{j+1}) &= -r \times ln(\theta) - r\bar{y}/\theta \times m \times ln(\theta) - \sum_{i=1}^{m} E[z_i|y]/\theta \\
&= -r \times ln(\theta) - r\bar{y}/\theta \times m \times ln(\theta) - \sum_{i=1}^{m} (\theta_j + t)/\theta
\end{aligned}
$$

We find the conditional MLE
$$\hat{\theta}_{j+1} = \frac{1}{n}(r\bar{y} - m(\theta_j + t))$$

Our simple algorithm to solve this is then,

```
## Traditional EM for censored exponential
## Generate Data
set.seed(456)
truetheta=10
t=9 ##censoring time
n=200
x=rexp(n,1/truetheta) #R uses rate

## Determine observed vs. missing data
y=replace(x = x, list = x>t, values = t) ## replace x > t, with t
y=sort(y)
r=sum(x<t) ## observed data
yc=y[1:r] ## observed data
ycbar=mean(yc) ## mean of the observed data
m = n-r


### EM
thetahat = ycbar ## MLE using observed data
cur = ycbar ## current estimate of theta for EM loop
diff=1 ## set difference variable for while-loop
while(diff>10^-4)
{
    i = length(thetahat)
    new_estimate = (r*ycbar+ m*(t + thetahat[i]))/n
    thetahat=c(thetahat, new_estimate) ## grow the vector of estimates
    diff = abs(thetahat[i+1]-thetahat[i])
}
```

```
plot(thetahat, type="l", ylim = c(2, 15))
thetaMLE=ycbar+(n-r)*t/r
abline(h=thetaMLE, col = "green") ## MLE estimate
abline(h = truetheta, col = "red") ## true theta
```



The true MLE estimate of θ is obtained by the following:

$$
\begin{aligned}
P(X > t) &= \int_t^{\infty} f(x)dx \\
&= \int_t^{\infty} \frac{1}{\theta} e^{x/\theta} dx \\
&= e^{-\infty} - (-e^{t/\theta}) \\
&= e^{t/\theta}
\end{aligned}
$$

Therefore, the complete data likelihood is

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{r} \frac{1}{\theta} e^{y_i/\theta} \times \prod_{i=1}^{m} \frac{1}{\theta} e^{t/\theta} \\
&= \theta^{-r} e^{-r\bar{y}/\theta} \times e^{-\sum_{i=1}^{m} t/\theta} \\
&= \theta^{-r} e^{-r\bar{y}/\theta} \times e^{-mt/\theta}
\end{aligned}
$$

And the log-likelihood is

$$
\begin{aligned}
ln(L(\theta)) &= -r \times ln(\theta) - r\bar{y}/ - mt/\theta \\
&= Q(\theta)
\end{aligned}
$$

Whose derivative is

$$
\begin{aligned}
\frac{d}{d\theta} ln(L(\theta)) &= \frac{-r}{\theta} + r\bar{y}/\theta^2 + mt/\theta^2 \\
&= 0 \\
&\rightarrow \hat{\theta}_{MLE} = \bar{y} + \frac{(n-r) \times t}{r}
\end{aligned}
$$

```r
## Monte Carlo EM for censored exponential
set.seed(456)
thetahat = ycbar
cur = ycbar
M = 20
diff=1
while(diff > 10^-4){
  i = length(thetahat)
  z = t + rexp(M*m,1/thetahat[i])
  Z = matrix(nrow=m, ncol=M, data = z)
  YC = matrix(nrow=r, ncol=M, data = rep(yc,M))
  simcomplete = rbind(Z, YC)
  new_estimate = mean(apply(simcomplete,2,mean))
  thetahat=c(thetahat, new_estimate)
  diff=abs(thetahat[i+1]-thetahat[i])
  M = M + 1
}



plot(thetahat, type="l",
     main = "MCEM estimates for theta",
     xlab = "iterations", ylab = "theta estimate",
     ylim = c(0, 11))
thetaMLE=ycbar+(n-r)*t/r
abline(h=thetaMLE, col = "green") ## MLE estimate
abline(h = truetheta, col = "red") ## true theta
```