Monte Carlo EM

Jonathan Navarrete

The EM Algorithm

Given a random sample of size n, with observed sample $\mathbf{X} = (X_1, \ldots, X_m)$ and *missing* random sample $\mathbf{Z} = Z_{m+1}, \ldots, Z_n$ we seek to compute

$$\hat{\theta} = \arg\max L(\theta|\mathbf{X}, \mathbf{Z})$$

Although $\mathbf{Z}$ is unobservable, we assume that $(\mathbf{X}, \mathbf{Z}) \sim \mathbf{f}(\mathbf{x}, \mathbf{z}|\theta)$.

We place a conditional distribion on $\mathbf{Z}$ given the observed data $\mathbf{x}$,

$$k(\mathbf{z}|\theta, \mathbf{x}) = f(\mathbf{x}, \mathbf{z}|\theta)/g(\mathbf{x}|\theta)$$

Here we assume that that $\mathbf{X} \sim g(\mathbf{x}|\theta)$, where

$$g(\mathbf{x}|\theta) = \int \mathbf{f}(\mathbf{x}, \mathbf{z}|\theta)\mathbf{dz}$$

The EM Algorithm

Denote the complete-data likelihood as $L^c(\theta|\mathbf{x},\mathbf{z})$ and the observed-data likelihood as $L(\theta|\mathbf{x})$. Then, for any value of θ, θ_i

$$logL(\theta|\mathbf{x}) = E[logL^c(\theta|\mathbf{x},\mathbf{z})] - E[logk(\mathbf{Z}|\boldsymbol{\theta_i},\mathbf{x})]$$

where the expectation is with respect to $k(\mathbf{z}|\theta_i,\mathbf{x})$. We can rewrite this as

$$E[logL^c(\theta|\mathbf{x},\mathbf{z})] = logL(\theta|\mathbf{x}) + E[logk(\mathbf{Z}|\theta_i,\mathbf{x})]$$

where our focus is concerned with maximizing $E[logL^c(\theta|\mathbf{x},\mathbf{z})]$.

The EM Algorithm

Denoting $E[logL^c(\theta|\mathbf{x},\mathbf{z})]$ = $Q(\theta|\theta_i,\mathbf{x})$, the EM algorith iterates through values of θ_i by maximizing $Q(\theta|\theta_i,\mathbf{x})$.

The EM Algorithm

`Pick a starting value` $\hat{\theta}_0$

`Then for i in 1:n do`

`1. Compute` (E-step)

$$Q(\theta|\theta_{i-1},\mathbf{x}) = E[logL^c(\theta|\mathbf{x},\mathbf{z})]$$

`where the expectation is with respect to` $k(\mathbf{Z}|\theta_\mathbf{i},\mathbf{x})$

`2. Maximize` $Q(\theta|\theta_{i-1},\mathbf{x})$ `in` θ `and take`

$$\hat{\theta}_i = \arg\max Q(\theta|\theta_{i-1},\mathbf{x})$$

`repeat until convergence criteria is met`

The First Exercise

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

Below is the setup for the first exercise.

The First Exercise

There are two light bulb survival experiments.

In the first, there are N bulbs, $y_1, \ldots, y_N$, whose exact lifetimes are recorded. The lifetimes have an exponential distribution, such that $y_i \sim Exp(\theta)$.

In the second experiment, there are M bulbs, $x_1, \ldots, x_M$. 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 $E_1, \ldots, E_M$ 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).

The First Exercise

Let $X_1, \ldots, X_M$ 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 $\mathcal{Y} = (Y_1, \ldots, Y_N, E_1, \ldots, E_M)$ and the unobserserved data is $\mathcal{X} = (X_1, \ldots, X_M)$.

The complete data log-likelihood is obtained by

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

7/31

The First Exercise

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,

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

The First Exercise

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

The First Exercise

$$E[X_i|\mathcal{Y}] = E[X_i|E_i] = \begin{cases} t+\theta & \text{if } E_i = 1 \\ \theta - t\frac{e^{-t/\theta}}{1-e^{-t/\theta}} & \text{if } E_i = 0 \end{cases} \tag{2}$$

The First Exercise

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

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

The First Exercise

For the denominator, we get

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

and for the numerator we obtain

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

The First Exercise

Altogether, we obtain

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

The First Exercise

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

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

The First Exercise

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

15/31

The First Exercise

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 $\theta^{(j-1)}$.

$$\log(L(\theta)) = -(N+M)log(\theta) - \frac{1}{\theta}[N\bar{Y} + Z(t+\theta^{(j-1)}) + (M-Z)(\theta^{(j-1)} - tp^{(j-1)})] \tag{3}$$

where

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

The First Exercise

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

$$\frac{\mathrm{d}}{\mathrm{d}x} ln(L(\theta)) = 0$$

$$0 = -\frac{(N+M)}{\theta} + \frac{1}{\theta^2}\left[N\bar{Y} + Z(t + \theta^{(j-1)}) + (M-Z)\left(\theta^{(j-1)} - t \times \frac{e^{-t/\theta^{(j-1)}}}{1 - e^{-t/\theta^{(j-1)}}}\right)\right]$$

$$\frac{(N+M)}{\theta} = \frac{1}{\theta^2}\left[N\bar{Y} + Z(t + \theta^{(j-1)}) + (M-Z)\left(\theta^{(j-1)} - t \times \frac{e^{-t/\theta^{(j-1)}}}{1 - e^{-t/\theta^{(j-1)}}}\right)\right]$$

$$\theta = \left[N\bar{Y} + Z(t + \theta^{(j-1)}) + (M-Z)\left(\theta^{(j-1)} - t \times \frac{e^{-t/\theta^{(j-1)}}}{1 - e^{-t/\theta^{(j-1)}}}\right)\right] / (N+M)$$

17/31

The First Exercise

Thus, for each *j*th M-step, we will calculate

$$\theta^{(j)} = f(\theta^{(j-1)})$$

$$\theta = \left[N\bar{Y} + Z(t + \theta^{(j-1)}) + (M - Z)\left(\theta^{(j-1)} - t \times \frac{e^{-t/\theta^{(j-1)}}}{1 - e^{-t/\theta^{(j-1)}}}\right)\right] \,/\, (N + M)$$

The First Exercise

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

t = 5 ## time cut off
N = 100 ## sample size of ex 1
M = 50 ## sample size of ex 2
y = rexp(n = N, rate = rate)
x = rexp(n = M, rate = rate)
x = sort(x)
E = as.integer(x > t)  ## 0 & 1

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

The First Exercise

```
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
```

The First Exercise

```
## compare results
print(theta.jp1) ## EM theta estimate
```

```
## [1] 5.45417
```

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

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

```
mean(c(y, x)) ## compare against complete-data
```

```
## [1] 5.427108
```

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

EM Normal Example

Suppose $X = (x_1, \ldots, x_n)^T$ is a random sample from $N(\mu, 1)$. Let the observations be in order such that $x_1 < x_2 < \ldots < x_n$. Suppose that after time c, values are censored or missing, such that only $x_1, \ldots, x_m$ are observed, and $x_{m+1}, \ldots, x_n$ are unobserved. Then, $r = (n - m)$ would be the quantity missing. We will use the EM and MCEM algorithms to find approximations for μ. Let $Z = (x_{m+1}, \ldots, x_n)^T$.

First, construct the likelihood function.

$$
\begin{aligned}
L(\mu|x) &= \prod^{m} f(x_i|\mu, 1) \times \prod^{r} f(z_i|\mu, 1) \\
&= (2\pi)^{-n/2} exp(-\frac{1}{2}\sum_{i=1}^{m}(x_i - \mu)^2) \times exp(-\frac{1}{2}\sum_{i=1}^{m}(z_i - \mu)^2) \\
&\propto exp(-\frac{1}{2}\sum_{i=1}^{m}(x_i - \mu)^2) \times exp(-\frac{1}{2}\sum_{i=1}^{m}(z_i - \mu)^2)
\end{aligned}
$$

22/31

EM Normal Example

The log-likelihood is then

$$ln(L(\mu|X)) = -\frac{1}{2}\sum_{i=1}^{m}(x_i - \mu)^2) - \frac{1}{2}\sum_{i=1}^{m}(z_i - \mu)^2$$

We now find the conditional expectation $E[z_i|X]$

$$E[z_i|X] = E[z_i|x > c] = \int_c^\infty \frac{P(x_i > x|x_i > c)}{P(x_i > c)}$$
$$= \mu + \sigma\frac{\phi(c-\mu)}{1-\Phi(c-\mu)}$$

For notes on this derivation, see Truncated Normal Distribution

EM Normal Example

$$
\begin{aligned}
Q(\mu|\mu_t) &= -\frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^2) - \sum E[z_i|X] \\
&= -\frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^2) - \sum E[z|X] \\
&= -\frac{1}{2}\sum_{i=1}^{m}(x_i-\mu)^2) - (n-m)E[z|X]
\end{aligned}
$$

The MLE for μ is then,

$$
\begin{aligned}
\mu_{t+1} &= \frac{m\bar{x}}{n} + \frac{(n-m)E[z|X]}{n} \\
&= \frac{m\bar{x}}{n} + \frac{(n-m)(\mu_t)}{n} + \frac{(n-m)\phi(c-\mu_t)}{n\Phi(c-\mu_t)}
\end{aligned}
$$

EM Normal Example

```
set.seed(2345)
n = 100
mu = 4
sd = 1
x = rnorm(n, mu, sd) ## generate some data
c = 5 ## time cut off
w = x[x < c] ## obtain samples before time cut off
m = sum(x < c) ## number of observed samples
wbar = mean(w) ## observed mean
r = n - m ## difference in sample size
```

EM Normal Example

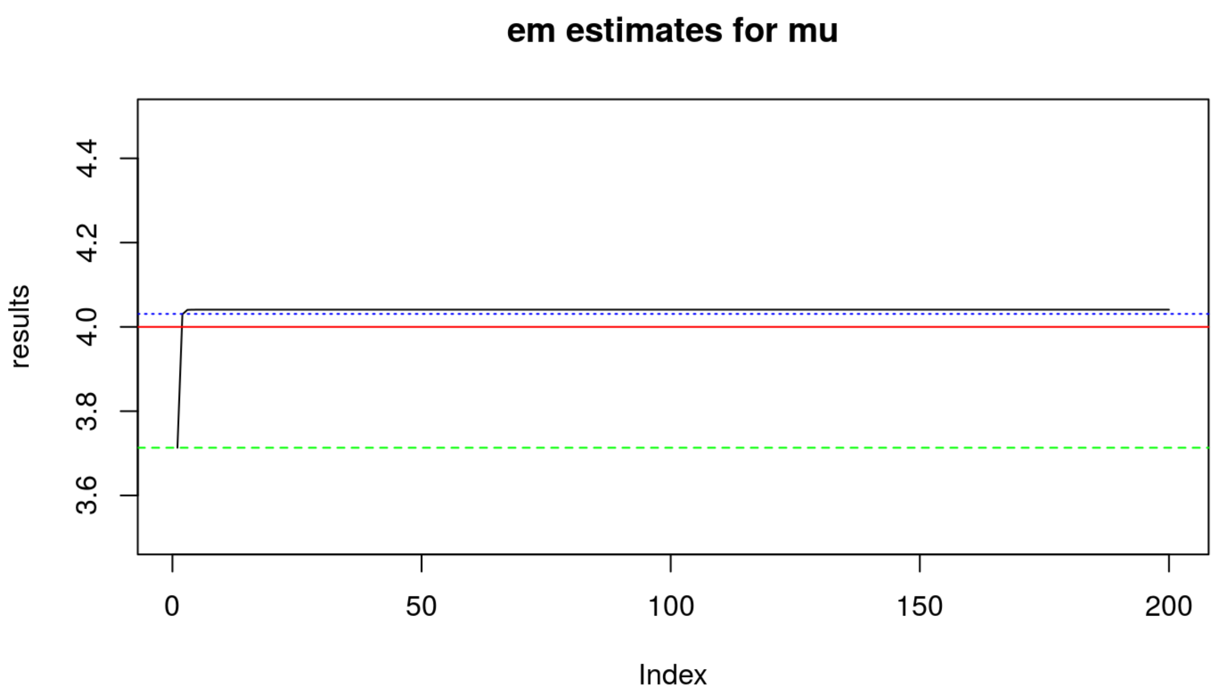
```
N = 200
mu_new = wbar
results = numeric(N)
for(i in 1:N){
    results[i] = mu_new
    mu_old = mu_new
    mu_new = m*wbar/n + (r*mu_old/n) +
      (r/n)*sd*(dnorm(c - mu_old))/(1 - pnorm(c - mu_old))  ## r/n instead of 1/n
    #print(mu_new)
}

print(tail(results))
```

```
## [1] 4.040821 4.040821 4.040821 4.040821 4.040821 4.040821
```

EM Normal Example

```
plot(results, type = "l", main = "em estimates for mu", ylim = c(3.5, 4.5))
abline(h = mu, col = "red")
abline(h = wbar, col = "green", lty = 2)
abline(h = mean(x), col = "blue", lty = 3)
```



Monte Carlo EM

A MC flavor of the EM algorithm

1. Draw missing data sets $\mathbf{Z_1, Z_2, \ldots, Z_m} \sim f_{Z|X}(z|x, \theta_i)$ where each $\mathbf{Z_i}$ is a vector of all missing values needed to complete the observed data set $(\mathbf{X, Z})$.
2. Calculate $\bar{Q}(\theta|\theta_{i-1}, X, \mathbf{Z_1, \ldots, Z_m}) = \frac{1}{m}\sum_{i=1}^{m} Q(\theta|\theta_{i-1}, X, \mathbf{Z_i})$

EM Normal Example

Monte Carlo EM

```
set.seed(2345)
n = 100
mu = 4
sd = 1
x = rnorm(n, mu, sd)
c = 5
w = x[x < c]
m = sum(x < c)
wbar = mean(w)
r = n - m
```

EM Normal Example

Monte Carlo EM

```
M = 10
N = 100
mu_new = wbar
results = numeric(N)
for(i in 1:N){
    results[i] = mu_new
    mu_old = mu_new
    ## abs(N(0,1)) + mu_old + (c - mu_old) to *approximate*
    ## the truncated samples we need
    Z = matrix(data = (c - mu_old) + (mu_old +  abs(rnorm(n = r*M, mean = 0, sd = 1))),
        nrow = r, ncol = M)
    mu_new = (m*wbar/n) + mean(colMeans(Z))*r/n
    M = M + 1
}
```

EM Normal Example

Monte Carlo EM

```
plot(results, type = "l", ylim = c(3.5, 4.5))
abline(h = mu, col = "red")
abline(h = wbar, col = "green", lty = 2)
abline(h = mean(x), col = "blue", lty = 3)
```