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(4567)
N = 500
theta = 2
c = 0.5
y = sort(rexp(n = N, rate = theta))
bigC = sum(y> c)


y[y> c] = c



theta.tp1 = 0.1
for(i in 1:2000){
  theta.t = theta.tp1
  theta.tp1 = N / (sum(y) + (bigC/theta.t))
  #print(theta.tp1)
}

print(theta.tp1)
```

```
## [1] 2.068357
```

```
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)
```

```r
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
```