# Exercise 3

TMA4300

Erling Fause Steen, Christian Oppegård Moen

Spring 2022

## Contents

## Problem C: The EM-algorithm and bootstrapping

### 1.

Let $x_1, ...x_n$ and $y_1, ..., y_n$ be independet random variables, where

$$x_i \sim \text{Exp}(\lambda_0) \;\text{ and }\; y_i \sim \text{Exp}(\lambda_1)$$

We observe

$$z_i = \max(x_i, y_i) \;\text{ for }\; i = 1, ..., n$$

and

$$u_i = I(x_i \geq y_i) \;\text{ for }\; i = 1, ..., n.$$

The joint distribution of $(x_i, y_i), i = 1, ..n$ is given by

$$\begin{aligned} f(x, y|\lambda_0, \lambda_1) &= \prod_{i=1}^{n} f_x(x_i|\lambda_0) \cdot f_y(y_i|\lambda_1) \\ &= \prod_{i=1}^{n} \lambda_0 e^{-\lambda_0 x_i} \cdot \lambda_1 e^{-\lambda_1 y_i}. \end{aligned}$$

This means that the log likelihood is given by

$$\ln f(x, y|\lambda_0, \lambda_1) = \sum_{i=1}^{n} \ln \lambda_0 + \ln \lambda_1 - \lambda_0 x_i - \lambda_1 y_i = n(\ln \lambda_0 + \ln \lambda_1) - \lambda_0 \sum_{i=1}^{n} x_i - \lambda_1 \sum_{i=1}^{n} y_i \quad .$$

We want to find

$$E\left[\ln f(x, y|\lambda_0, \lambda_1)|z, u, \lambda_0^{(t)}, \lambda_1^{(t)}\right].$$

which is given by

$$\begin{aligned} Q(\lambda_0, \lambda_1|\lambda_0^{(t)}, \lambda_1^{(t)}) &= E\left[\sum_{i=1}^{n} n(\ln \lambda_0 + \ln \lambda_1) - \lambda_0 \sum_{i=1}^{n} x_i - \lambda_1 \sum_{i=1}^{n} y_i \;\mid\; z, u, \lambda_0^{(t)}, \lambda_1^{(t)}\right] \\ &= n(\ln \lambda_0 + \ln \lambda_1) - \lambda_0 \sum_{i=1}^{n} E(x_i \mid z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)}) - \lambda_1 \sum_{i=1}^{n} E(y_i \mid z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)}). \end{aligned}$$

Now, we want to find $E(x_i \mid z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)})$ and $E(y_i \mid z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)})$. We start by considering the first conditional expectation. This can found by first considering

$$f(x_i \mid z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)}) = \begin{cases} z_i & \text{for} \quad u_i = 1 \\ \frac{\lambda_0^{(t)} \exp(-\lambda_0^{(t)} x_i)}{1 - \exp(-\lambda_0^{(t)} z_i)} & \text{for} \quad u_i = 0 \end{cases} \quad .$$

The expectation is given by

$$E[x_i|z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)}] = u_i z_i + (1 - u_i) \int_0^{z_i} x_i \frac{\lambda_0^{(t)} \exp(-\lambda_0^{(t)} x_i)}{1 - \exp(-\lambda_0^{(t)} z_i)} dx_i$$

where

$$\int_0^{z_i} x_i \frac{\lambda_0^{(t)} \exp(-\lambda_0^{(t)} x_i)}{1-\exp(-\lambda_0^{(t)} z_i)} dx_i = \frac{-z_i \lambda_0^{(t)} \exp(-\lambda_0^{(t)} z_i) - \exp(-\lambda_0^{(t)} z_i) + 1}{\lambda_0^t(-\exp(-\lambda_0^{(t)} z_i) + 1)}$$
$$= \frac{1}{\lambda_0^{(t)}} - \frac{z_i}{\lambda_0^{(t)}(1-\exp(-\lambda_0^{(t)}))}$$

We also need to find $E[y_i|z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)}]$. We first consider the pdf

$$f(y_i \mid z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)}) = \begin{cases} z_i & \text{for} \quad u_i = 1 \\ \frac{\lambda_1^{(t)} \exp(-\lambda_1^{(t)} y_i)}{1-\exp(-\lambda_1^{(t)} z_i)} & \text{for} \quad u_i = 0 \end{cases}.$$

Then we find the expectation

$$E[y_i|z_i, u_i, \lambda_0^{(t)}, \lambda_1^{(t)}] = (1-u_i)z_i + u_i \left( \frac{1}{\lambda_1^{(t)}} - \frac{z_i}{\exp(\lambda_1^{(t)} z_i) - 1} \right)$$

Thus, we end up with the expression

$$E[\ln f(\mathbf{x}, \mathbf{y}|\lambda_0, \lambda_1)|\mathbf{z}, \mathbf{u}, \lambda_0^{(t)}, \lambda_1^{(t)}]$$
$$= n(\ln\lambda_0 + \ln\lambda_1) - \lambda_0 \sum_{i=1}^n \left[ \frac{1}{\lambda_0^{(t)}} - \frac{z_i}{\lambda_0^{(t)}(1-\exp(-\lambda_0^{(t)}))} \right] - \lambda_1 \sum_{i=1}^n \left[ \frac{1}{\lambda_0^{(t)}} - \frac{z_i}{\lambda_0^{(t)}(1-\exp(-\lambda_0^{(t)}))} \right].$$

This is what we expected to find.

## 2.

In this problem we want to implement the EM-algorithm. We have found the conditional expectation $Q(\lambda_0, \lambda_1) = Q(\lambda_0, \lambda_1|\lambda_0^{(t)}, \lambda_1^{(t)})$. This corresponds to the E-step in the EM algorithm. The M-step of the algorithm is to determine

$$(\lambda_0^{(t+1)}, \lambda_1^{(t+1)}) = \text{argmax} \;\; Q(\lambda_0, \lambda_1).$$

This can be found by setting the partial derivates and $Q(\lambda_0, \lambda_1)$ equal to zero.

$$\frac{\partial}{\partial\lambda_0} Q(\lambda_0, \lambda_1) = \frac{n}{\lambda_0} - \sum_{i=1}^n \left( u_i z_i + (1-u_i) \left( \frac{1}{\lambda_0^{(t)}} - \frac{z_i}{e^{\lambda_0^{(t)} z_i} - 1} \right) \right) = 0$$

$$\frac{\partial}{\partial\lambda_1} Q(\lambda_0, \lambda_1) = \frac{n}{\lambda_1} - \sum_{i=1}^n \left( (1-u_i) z_i + u_i \left( \frac{1}{\lambda_1^{(t)}} - \frac{z_i}{e^{\lambda_1^{(t)} z_i} - 1} \right) \right) = 0$$

We solve these two equations for $\lambda_0$ and $\lambda_1$ respectively. This gives the M-step

$$\lambda_0^{(t+1)} = n / \sum_{i=1}^n \left( u_i z_i + (1-u_i) \left( \frac{1}{\lambda_0^{(t)}} - \frac{z_i}{e^{\lambda_0^{(t)} z_i} - 1} \right) \right)$$
$$\lambda_1^{(t+1)} = n / \sum_{i=1}^n \left( (1-u_i) z_i + u_i \left( \frac{1}{\lambda_1^{(t)}} - \frac{z_i}{e^{\lambda_1^{(t)} z_i} - 1} \right) \right)$$

Let $\lambda^{(t)} = (\lambda_0^{(t)}, \lambda_1^{(t)})$. We want to implement the EM-algorithm and we use the convergence criterion

$$d(x^{(t+1)}, x^t) = ||\lambda^{(t+1)} - \lambda^{(t)}||_2 < \epsilon.$$

The function below returns the conditional expectation, that is the E-step of the EM algorithm.

```
cond_expectation <- function(lambda0, lambda1, lambda0t, lambda1t, u, z) {
    n = length(u)
    exp = n * (log(lambda0) + log(lambda1)) - (lambda0 * sum(u * z + (1 - u) * (1/lambda0t -
        (z)/(exp(lambda0t * z) - 1)))) - (lambda1 * sum(u * z + (1 - u) * (1/lambda1t -
        (z)/(exp(lambda1t * z) - 1))))
    return(exp)
}
```

Under is a function that implement M-step.

```
M_step <- function(lam0, lam1, u, z) {
    n = 200
    lambda0next = n/sum(u * z + (1 - u) * (1/lam0 - z/(exp(lam0 * z) - 1)))
    lambda1next = n/sum((1 - u) * z + u * (1/lam1 - z/(exp(lam1 * z) - 1)))

    return(c(lambda0next, lambda1next))
}
```

Under the EM algorithm is implemented.

```
EM_algorithm <- function(lambda, u, z, epsilon = 1e-14) {
    lambda0 = lambda[1]
    lambda1 = lambda[2]
    lambda = c(lambda0, lambda1)
    list0 <- c()
    list1 <- c()
    for (i in 1:300) {
        lambda0t = M_step(lambda0, lambda1, u, z)[1]
        lambda1t = M_step(lambda0, lambda1, u, z)[2]
        lambdat = c(lambda0t, lambda1t)
        list0 <- c(list0, lambda0t)
        list1 <- c(list1, lambda1t)
        norm = norm(lambdat - lambda, type = "2")
        lambda0 = lambda0t
        lambda1 = lambda1t
        lambda = c(lambda0t, lambda1t)
        if (norm < epsilon) {
            break
        }
    }
    return(list(lambdas0 = list0, lambdas1 = list1))
}

# The estimated MLEs of lambda0 and lambda1
lambdas <- EM_algorithm(c(2.5, 5), u, z)
lambdas0 = lambdas$lambdas0
lambdas1 = lambdas$lambdas1
```

```r
MLE_lambda0 = lambdas0[length(lambdas0)]
MLE_lambda1 = lambdas1[length(lambdas1)]
```

```r
MLE_lambda0
```

```
## [1] 3.463089
```

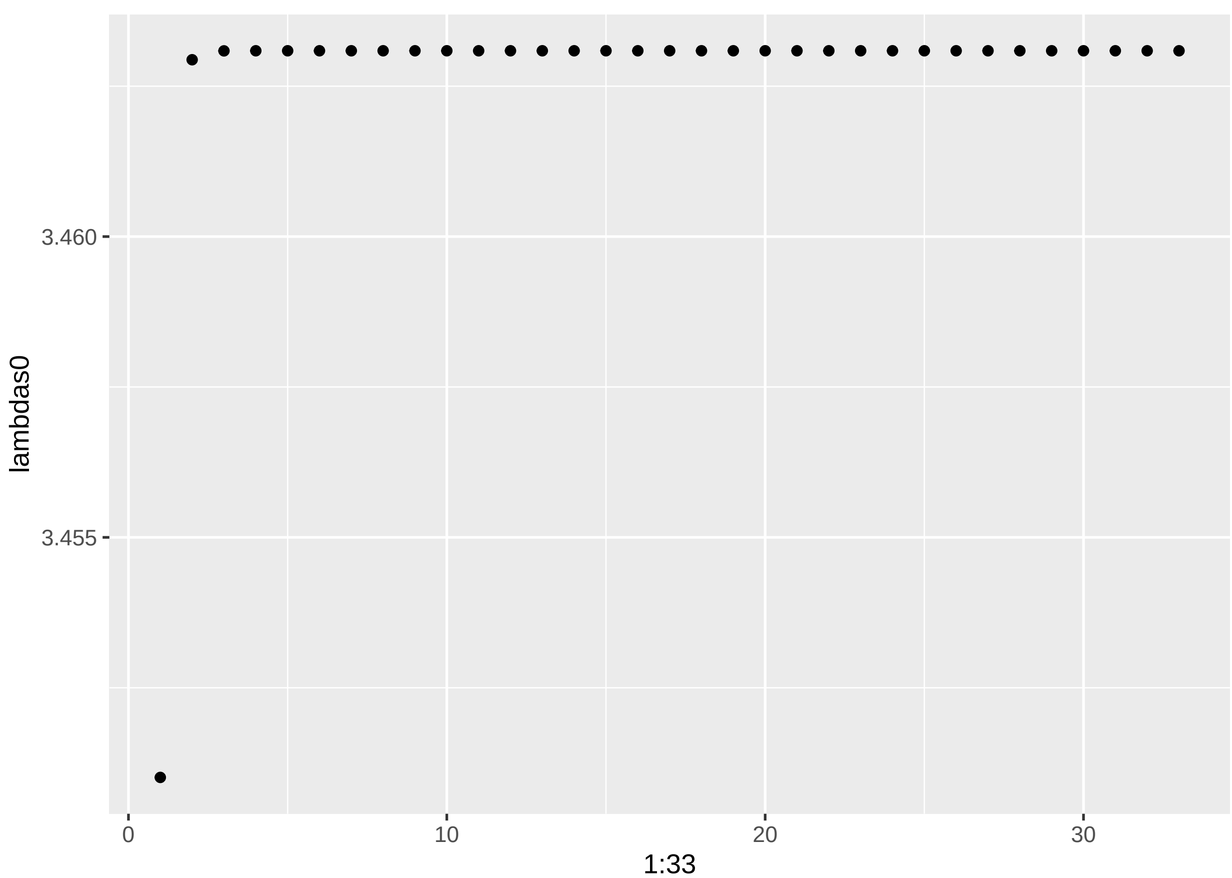
```r
MLE_lambda1
```

```
## [1] 9.334877
```

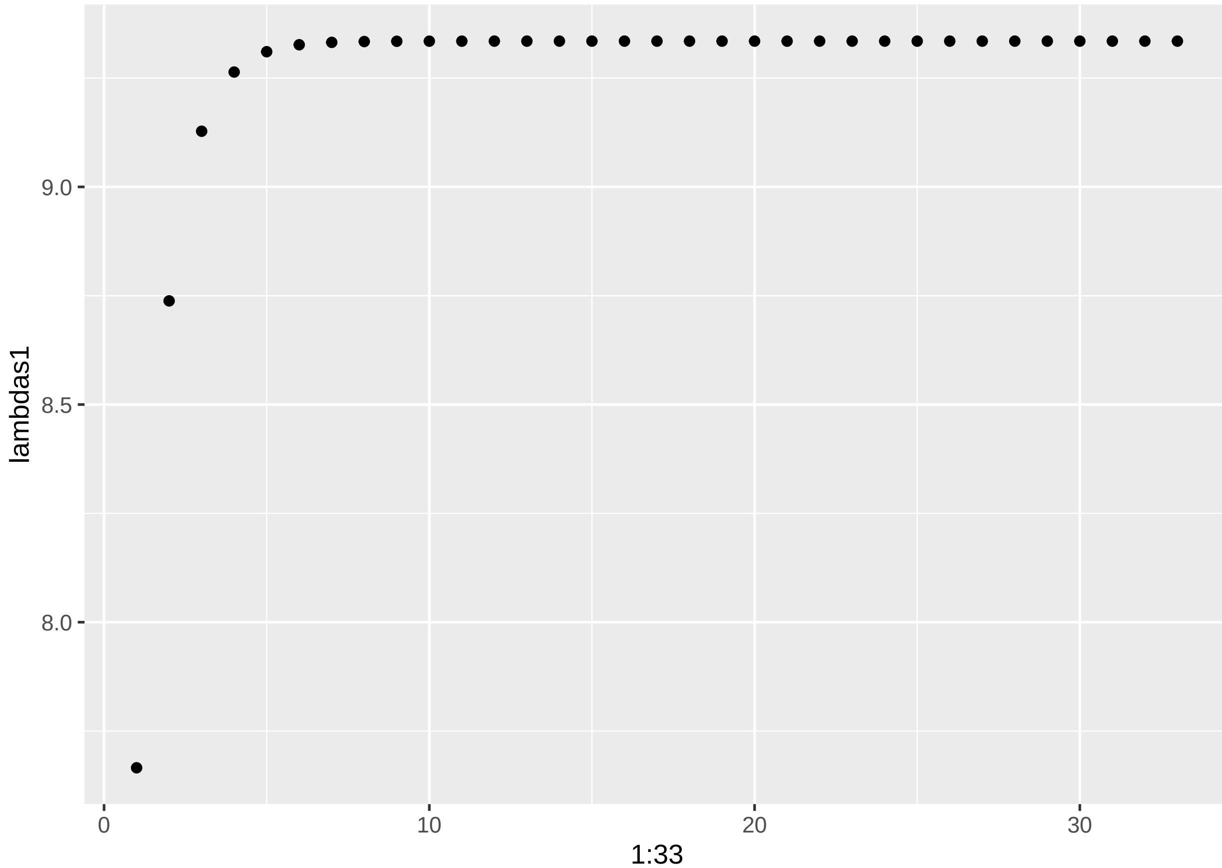
The maaximum likelihood estimates for $\lambda_0$ is 3.463089 and 9.3348769 for $\lambda_1$.

We also want to visualize the covergence.

```r
library(ggplot2)
lambdas = data.frame(lambdas)
ggplot(data = lambdas) + geom_point(mapping = aes(x = 1:33, y = lambdas0))
```




```r
ggplot(data = lambdas) + geom_point(mapping = aes(x = 1:33, y = lambdas1))
```




### 3.

In this task, the standard deviations and biases of $\hat{\lambda}_0$ and $\hat{\lambda}_1$ in addition to corr$[\hat{\lambda}_0, \hat{\lambda}_1]$ are estimated by using bootstrap. The pseudocode for the bootstrap algorithm is presented below.

**Algorithm 1** algorithm

1: **for** $b = 1, ..., B$ **do**
2: Bootstrap sample $(z_b^*, u_b^*)$ from $(\boldsymbol{z}, \boldsymbol{u})$ with replacement.
3: Estimate $(\hat{\lambda}_0, \hat{\lambda}_1)$ by EM-algorithm using $(z_b^*, u_b^*)$.
4: **end for**

The algorithm is then implemented

```
B = 200  # Seldom needed more samples
n = length(u$V1)  # length of lambdas0 from EM
t = length(lambdas$lambdas0)

set.seed(420)
lambda.T = matrix(nrow = B, ncol = 2)
for (i in 1:B) {
    # browser()
    Bs = sample(n, n, replace = TRUE)
    lt = EM_algorithm(c(2.5, 5), u$V1[Bs], z$V1[Bs])
```

```
    lambda.T[i, ] = cbind(lt$lambdas0[length(lt$lambdas0)], lt$lambdas1[length(lt$lambdas1)])
}

mu.boot = apply(lambda.T, 2, mean)
l.var.boot = var(lambda.T)
var.boot = sqrt(diag(l.var.boot))
bias.boot = mu.boot - c(lambdas$lambdas0[length(lambdas$lambdas0)], lambdas$lambdas1[length(lambdas$lamb
var.boot
```

```
## [1] 0.2566881 0.8805286
```

```
bias.boot
```

```
## [1] 0.01868494 0.19034030
```

```
sd(lambda.T)
```

```
## [1] 3.094066
```

```
cor(lambda.T)
```

```
##            [,1]       [,2]
## [1,]  1.0000000 -0.1135658
## [2,] -0.1135658  1.0000000
```

```
var.boot
```

```
## [1] 0.2566881 0.8805286
```

## 4.

We want to find an analytical formula of $f_{Z_i,U_i}(z_i, u_i|\lambda_0, \lambda_1)$. We start by looking at the case where $u_i = 0$, and thus $z_i = y_i$. The cdf is given by

$$F_{Z_i}(z_i|u_i = 0) = P(Y_i \leq z_i|X_i \leq y_i) = \int_0^{z_i}\int_0^{y_i} f_{Y_i}(y_i|\lambda_1)f_{X_i}(x_i|\lambda_0)dx_i dy_i$$

$$= \int_0^{z_i}\int_0^{y_i} \lambda_1 \exp(-\lambda_1 y_i)\lambda_0 \exp(-\lambda_0 x_i) = \int_0^{z_i} \lambda_1 \exp(-\lambda_1 y_i)(1 - \exp(-\lambda_0 y_i))dy_i$$

$$= -\lambda_1 \cdot \frac{\exp(-\lambda_1 z_i - \lambda_0 z_i) - 1}{-\lambda_1 - \lambda_0} - \exp(-\lambda_1 z_i) + 1$$

$$\implies f(z_i|u_i = 0) = \frac{\mathrm{d}F_{Z_i}(z_i|u_i = 0)}{\mathrm{d}z_i} = \exp(-\lambda_1 z_i)\lambda_1(1 - \exp(-\lambda_0 z_i))$$

For $u_i = 1$, we have

$$F_{Z_i}(z_i|u_i = 1) = P(X_i \leq z_i, Y_i \leq x_i) = \int_0^{z_i}\int_0^{x_i} f_{X_i}(x_i|\lambda_0)f_{Y_i}(y_i|\lambda_1)\mathrm{dy}_i\mathrm{dx}_i$$

$$= -\lambda_0 \frac{exp(-z_i\lambda_0 - z_i\lambda_1) - 1}{-\lambda_0 - \lambda_1} + 1$$

$$\implies f(z_i|u_i=1) = \frac{\mathrm{d}F_{Z_i}(z_i|u_i=1)}{\mathrm{d}z_i} = \exp(-\lambda_0 z_i)\lambda_0(1-\exp(-\lambda_1 z_i))$$

The likelihood is given by

$$L(\lambda_0, \lambda_1|\mathbf{z}, \mathbf{u}) = \prod_{i=0}^{n} f_{Z_i,U_i}(z_i, u_i|\lambda_0, \lambda_1)$$

where

$$f_{Z_i,U_i}(z_i, u_i|\lambda_0, \lambda_1) = \begin{cases} \lambda_1 e^{-\lambda_1 z_i}(1-e^{-\lambda_0 z_i}), & u_i = 0 \\ \lambda_0 e^{-\lambda_0 z_i}(1-e^{-\lambda_1 z_i}), & u_i = 1. \end{cases}$$

The log likelihood is therefore given by

$$l(\lambda_0, \lambda_1|\mathbf{z}, \mathbf{u}) = \sum_{i:u_i=0} \left(\ln(\lambda_1) - \lambda_1 z_i + \ln(1-e^{-\lambda_0 z_i})\right) + \sum_{i:u_i=1} \left(\ln(\lambda_0) - \lambda_0 z_i + \ln(1-e^{-\lambda_1 z_i})\right)$$

The maximum likelihood estimators can be found by solving

$$\frac{\partial l(\lambda_0, \lambda_1|\mathbf{z}, \mathbf{u})}{\partial \lambda_0} = 0$$

and

$$\frac{\partial l(\lambda_0, \lambda_1|\mathbf{z}, \mathbf{u})}{\partial \lambda_1} = 0.$$

The equations become

$$\frac{\partial l(\lambda_0, \lambda_1|\mathbf{z}, \mathbf{u})}{\partial \lambda_0} = \sum_{i:u_i=0}^{n} \frac{z_i \exp(\lambda_0 z_i)}{\exp(\lambda_0 z_i) - 1} + \sum_{i:u_i=1}^{n} \frac{1}{\lambda_0} - z_i = 0$$

and

$$\frac{\partial l(\lambda_0, \lambda_1|\mathbf{z}, \mathbf{u})}{\partial \lambda_1} = \sum_{i:u_i=1}^{n} \frac{z_i \exp(\lambda_0 z_i)}{\exp(\lambda_0 z_i) - 1} + \sum_{i:u_i=0}^{n} \frac{1}{\lambda_0} - z_i = 0$$

We solve this numerically. To check whether the soutions are maximas, we consider the Hessian.

$$

$$

```r
# Find the z-values where u=0 and where u=1

z_0 = z[u == 0]
z_1 = z[u == 1]

l <- function(lambdas) {
    ll <- sum(log(lambdas[2] * exp(-lambdas[2] * z_0) * (1 - exp(-lambdas[1] * z_0)))) +
        sum(log(lambdas[1] * exp(-lambdas[1] * z_1) * (1 - exp(-lambdas[2] * z_1))))
    return(-ll)
}

lambdas = optim(par = c(1, 1), fn = l)$par
```

The maximum likelihood estimator of $\lambda_0$ is 3.46589 and the maximum likelihood estimator of $\lambda_1$ is 9.3511034.

The difference from the values obtained for the EM algorithm is very small. An advantage of using this approach compared is that it is less computationally expensive. The EM algorithm can be slow.