TD à rendre sur l'agorithme EM

MECH Bealy

10/10/2021

Simulation

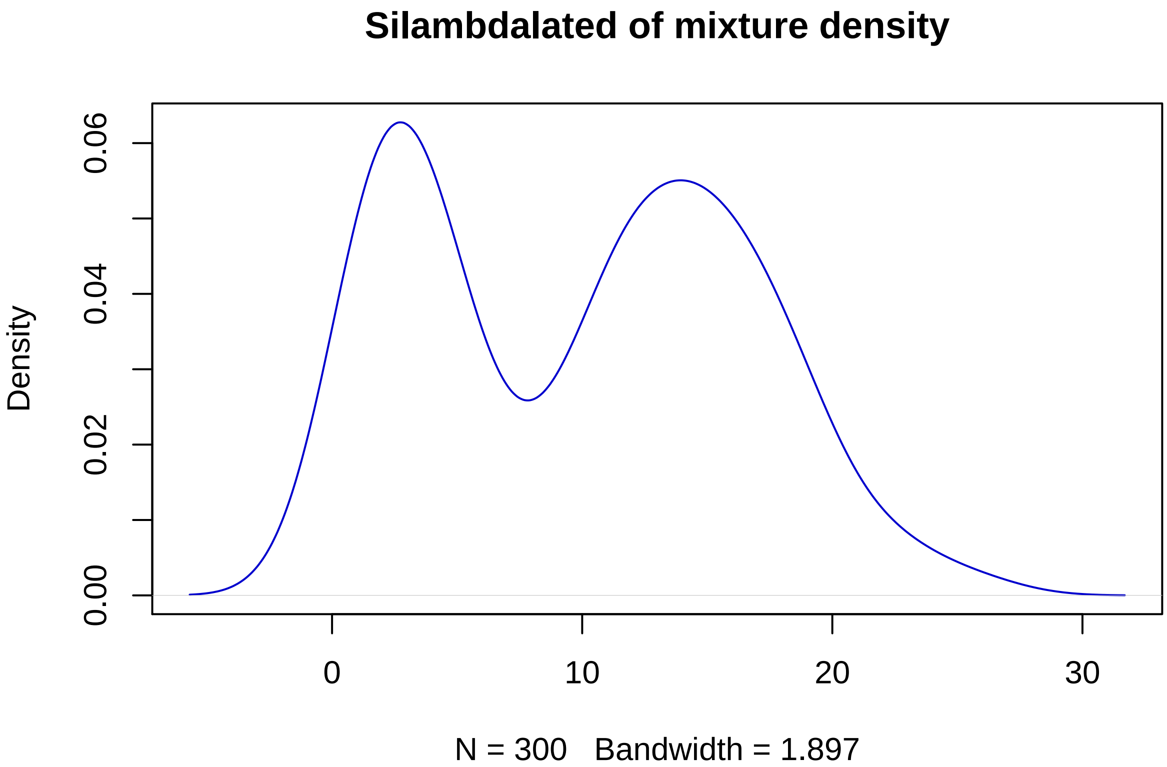
```
n1 = 100; n2 = 200;
pi1 = 0.4; pi2 = 1 - pi1;
lambda1 = 3 ; lambda2 = 15
x1 = rpois(n1, lambda1)
x2 = rpois(n2,lambda2)
```

3. Créer un vecteur de 300 valeurs entières (100 "1", suivi de 200 "2").

```
X = c(x1, x2)
```

4. Simuler un mélange de lois de Poisson à deux composantes:

```
Mix <- function(n, pi, lambda){
  K = length(pi)
  Z = t(rmultinom(n, size = 1, prob = pi))
  X = c()
  for (i in 1:n){
    id = which(Z[i,] == 1)
    X = c(X, rpois(1, lambda[id]))
  }
  return(X)
}
pi = c(0.4, 0.6)
lambda = c(3, 15)
x = Mix(300, pi, lambda)
plot(density(x), col = "blue3", main = "Silambdalated of mixture density", pch = 19)
```



Algorithme EM pour une mélange de lois de Poisson à K composantes

1. Initail of algorihthme EM:

$$\theta^0 = (\pi_1^0, ..., \pi_{K-1}^0, \lambda_1^0, ..., \lambda_K^0)$$

2. Detail the computation of t_{ik}^q

$$\begin{aligned}
t_{ik}^q &= \mathbb{E}_{z_{ik}|x_i}[z_{ik}] \\
&= 1 \times \mathbb{P}(z_{ik} = 1|x_i) + 0 \times \mathbb{P}(z_{ik} = 0|x_i) \\
&= \frac{\mathbb{P}(z_{ik} = 1; x_i)}{\mathbb{P}(x_i)} \\
&= \frac{\mathbb{P}(x_i|z_{ik} = 1)\mathbb{P}(z_{ik} = 1)}{\sum_{l=1}^K \mathbb{P}(x_i|z_{il} = 1)\mathbb{P}(z_{il} = 1)} \\
&= \frac{\pi_k^q p(x_i; \theta^q)}{\sum_{l=1}^K \pi_l^q p(x_i; \theta^q)}
\end{aligned}$$

3. E step $\mathbb{Q}(\theta^q|\theta)$

$$\begin{aligned}
\mathbb{Q}(\theta^q|\theta) &= \mathbb{E}_{Z|X,\theta^q}[log(\mathbb{P}_\theta(X;Z)|X;\theta^q)] \\
&= \mathbb{E}_{Z|X,\theta^q}\left[\sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik} log(\pi_k^q \mathbb{P}_\theta(x_i|z_{ik}=1)\right] \\
&= \sum_{i=1}^{n}\sum_{k=1}^{K} \mathbb{E}_{Z|X,\theta^q}\left[z_{ik}=1;\theta^q\right] log\left(\pi_k^q \mathbb{P}(x_i|z_{ik}=1)\right) \\
&= \sum_{i=1}^{n}\sum_{k=1}^{K} t_{ik}^q log\left(\pi_k^q p(x_i;\theta^q)\right)
\end{aligned}$$

4. M step $\theta^{q+1} = argmax_\theta \mathbb{Q}(\theta^q|\theta)$

$$\forall k, \quad \frac{\partial \mathbb{Q}(\theta^q|\theta)}{\partial \lambda_k} = 0$$

Since we have,

$$\mathbb{Q}(\theta^q|\theta) = \sum_{i=1}^{n}\sum_{k=1}^{K} t_{ik}^q log\left(\pi_k p(x_i;\theta^q)\right)$$

Then,

$$\begin{aligned}
&\frac{\partial \mathbb{Q}(\theta^q|\theta)}{\partial \lambda_k} = \frac{\partial}{\partial \lambda_k}\sum_{i=1}^{n}\sum_{k=1}^{k} t_{ik}^q(-\lambda_k + x_i log\lambda_k - log x_i!) = 0 \\
&-\sum_{i=1}^{n} t_{ik}^q + \sum_{i=1}^{n}\frac{x_i t_{ik}^q}{\lambda_k} = 0 \\
&\Rightarrow \lambda_k^{q+1} = \frac{1}{\sum_{i=1}^{n} t_{ik}^q}\sum_{i=1}^{n} t_{ik}^q x_i \\
&\Rightarrow \pi_k^{q+1} = \frac{1}{n}\sum_{i=1}^{n} t_{ik}^q
\end{aligned}$$

2. Programmer l'étape E

```
Tik<- function(x , pi, lambda) {
  n = length(x)
  K = length(pi)
  tik = matrix(0, n, K)
  for(i in 1:n) {
    for(k in 1:K) {
      tik[i, k] = pi[k]*dpois(x[i], lambda[k])
    }
    tik[i, ] = tik[i, ] / sum(tik[i, ])
  }
  return(tik)
}
```

3. Programmer l'étape M

```
params <- function(x, gamma) {
# gamma is tik (the responsibilities)
  n = length(x)
  pi = colSums(gamma) / n
  K = length(pi)
  lambda = t(gamma)%*%x / colSums(gamma)
  return(list(pi=pi, lambda = lambda))
}
```

4. Tester algorithme EM

```
n1 = 100; n2 = 200;
lambda1 = 3 ; lambda2 = 15;
x1 = rpois(n1, lambda1)
x2 = rpois(n2,lambda2)
X = c(x1, x2)
pi = c(0.4, 0.6)
lambda = c(3, 15)
tik = Tik(X, pi, lambda)
gamma = tik
theta = params(X, gamma)
theta
```

```
## $pi
## [1] 0.3365714 0.6634286
##
## $lambda
##           [,1]
## [1,]  2.938143
## [2,] 14.994487
```