

Exemplo com o algoritmo EM

```
## Algoritmo EM
## Exemplo adaptado de Hastie, Tibshirani & Friedman (2008),
##   The Elements of Statistical Learning, 2nd ed., Sec. 8.5.1

## Função densidade
fdens <- function(x, alfa, mu1, sig12, mu2, sig22) {
  (1 - alfa) * dnorm(x, mu1, sqrt(sig12)) +
  alfa * dnorm(x, mu2, sqrt(sig22))
}

## f.d.a.
fda <- function(x, alfa, mu1, sig12, mu2, sig22) {
  (1 - alfa) * pnorm(x, mu1, sqrt(sig12)) +
  alfa * pnorm(x, mu2, sqrt(sig22))
}

## Diferença relativa máxima
eps <- 1e-4

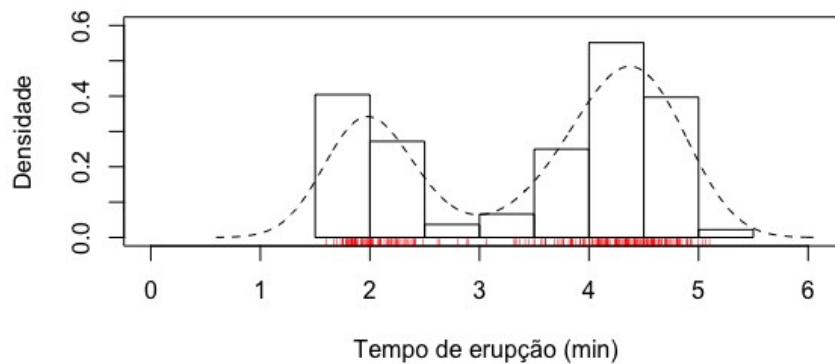
# Dados
?faithful

mydata <- faithful$eruptions

n <- length(mydata)
cat("\n Tamanho da amostra:", n)

  Tamanho da amostra: 272

hist(mydata, main = "", freq = FALSE, xlim = c(0, 1.2 * max(mydata)),
  xlab = "Tempo de erupção (min)", ylab = "Densidade", ylim = c(0, 0.6))
rug(mydata, col = "red")
lines(density(mydata), lty = 2)
```



```

# Iniciando ...
set.seed(62017)
alfa0 <- 0.5
mu10 <- sample(mydata, 1)
mu20 <- sample(mydata, 1)
sig120 <- sig220 <- var(mydata)
iter <- 0
dif <- 1

cat("\n Algoritmo EM \n Tolerância:", eps)

  Algoritmo EM
  Tolerância: 1e-04

cat("\n Estimativas iniciais (alfa, mu1, sig12, mu2, sig22): \n",
    c(alfa0, mu10, sig120, mu20, sig220))

  Estimativas iniciais (alfa, mu1, sig12, mu2, sig22):
  0.5 2.1 1.302728 4.15 1.302728

# Passos E e M
while (dif > eps) {
  iter <- iter + 1

  # Passo E
  gama <- alfa0 * dnorm(mydata, mu20, sqrt(sig220)) /
    fdens(mydata, alfa0, mu10, sig120, mu20, sig220)

  # Passo M
  mu1 <- weighted.mean(mydata, w = 1 - gama)
  sig12 <- weighted.mean((mydata - mu1)^2, w = 1 - gama)
  mu2 <- weighted.mean(mydata, w = gama)
  sig22 <- weighted.mean((mydata - mu2)^2, w = gama)
  alfa <- mean(gama)

  # Critério de parada
  dif <- max(abs((alfa - alfa0) / alfa0), abs((mu1 - mu10) / mu10),
    abs((sig12 - sig120) / sig120), abs((mu2 - mu20) / mu20),
    abs((sig22 - sig220) / sig220))

  alfa0 <- alfa
  mu10 <- mu1
  sig120 <- sig12
  mu20 <- mu2
  sig220 <- sig22
}

cat("\n Estimativas após", iter, "iterações")
cat("\n alfa, mu1, sig12, mu2, sig22: \n", c(alfa, mu1, sig12, mu2, sig22))
cat("\n Critério de parada:", dif)

```

```

Estimativas após 22 iterações
  alfa, mu1, sig12, mu2, sig22:
0.6515924 2.018615 0.05552279 4.27335 0.1910156

```

```

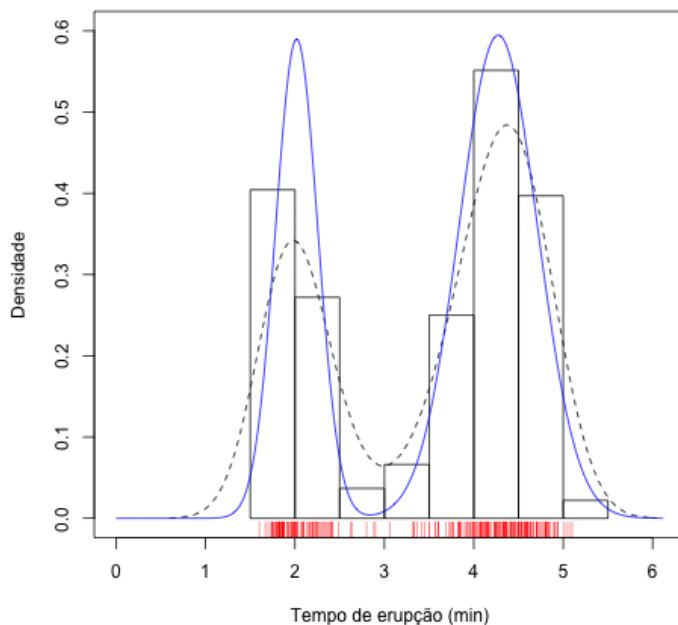
Critério de parada: 6.502095e-05

```

```

# Distribuição ajustada
curve(fdens(x, alfa, mu1, sig12, mu2, sig22), add = TRUE, col = "blue",
      n = 501)

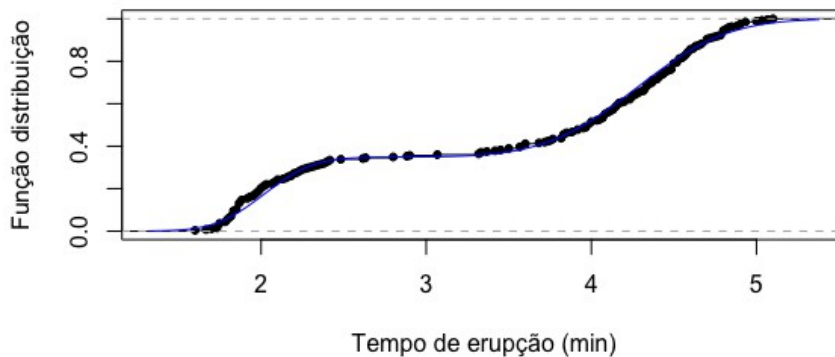
```



```

plot(ecdf(mydata), pch = 20, main = "", ylab = "Função distribuição",
     xlab = "Tempo de erupção (min)")
curve(fda(x, alfa, mu1, sig12, mu2, sig22), add = TRUE, col = "blue",
      n = 501)

```



Nota 1. Execute o algoritmo com diferentes estimativas iniciais para os parâmetros.

Nota 2. Modifique o critério de parada de modo que o algoritmo seja encerrado após um certo número de iterações ter sido executado, mesmo que a diferença máxima ($diff$) ainda seja maior do ϵ (eps).

Nota 3. Realize um teste de bondade do ajuste da mistura de duas distribuições normais.