

EM Algorithm

Shannon Houser

10/11/2021

```
#install.packages("palmerpenguins")  
library(palmerpenguins)
```

```
## Warning: package 'palmerpenguins' was built under R version 4.1.1
```

```
data(penguins)  
  
dat <- penguins[penguins$species != "Chinstrap",]  
dat <- dat[!is.na(dat$flipper_length_mm),]
```

Exercise 1

```
mean_a = 180  
mean_g = 200  
var_a = 20  
var_g = 20  
pi_a = .50  
pi_g = .50
```

```
p_z_a_vec = c()  
  
for (i in 1:nrow(dat)) {  
  p_z_a = pi_a * pnorm(dat$flipper_length_mm, mean_a, sqrt(var_a))/(pi_a * pnorm(dat$flipper_length_mm, mean_a, sqrt(var_a)) + pi_g * pnorm(dat$flipper_length_mm, mean_g, sqrt(var_g)))  
  append(p_z_a_vec, p_z_a)  
}  
  
pi_a_new = (1/nrow(dat))* sum(p_z_a)  
pi_g_new = 1 - pi_a_new  
  
pi_a = pi_a_new  
pi_g = pi_g_new
```

Exercise 2

As well, the second part of the lab will ask you to use observed blood type data from 300 individuals: Suppose in a group of 300 individuals we observe 135 with blood type A, 39 with blood type B, 108 with blood type O, and 18 with blood type AB.

```

n_a = 135
n_b = 108
n_o = 108
n_ab = 18
n = 300
p_a = .25
p_b = .25
p_o = .5

```

```

for (i in 1:10000) {
  n_aa = (n_a * (p_a)^2)/((p_a)^2 + 2*p_a*p_o)
  n_bb = (n_b * (p_b)^2)/((p_b)^2 + 2*p_b*p_o)
  n_ao = (2 * n_a * p_a * p_o)/((p_a)^2 + 2*p_a*p_o)
  n_bo = (2 * n_b * p_b * p_o)/((p_b)^2 + 2*p_b*p_o)
  n_oo = 300 - n_aa - n_bb - n_ao - n_bo

  p_a_new = (2*n_aa + n_ao + n_ab)/(2*n)
  p_b_new = (2*n_bb + n_bo + n_ab)/(2*n)
  p_o_new = (2*n_oo + n_ao + n_bo)/(2*n)

  # if (p_a_new == p_a) {
  #   print(p_a_new, p_b_new, p_o_new)
  #   break
  # }

  p_a = p_a_new
  p_b = p_b_new
  p_o_new = p_o_new
}

```