Homework 3 Part1

Rafael Espinosa Castaneda

2023-02-12

Question a

```r
# load libraries
library(tidyverse)
library(gridExtra)
library(xts)
library(depmixS4)
library(gamlss.data)




#---------------------- Question a ------------------
data(polio)

months<-seq(as.Date("1970-01-01"), as.Date("1983-12-01"), by="months")


data_complete<-data.frame("Count"=polio,"Time"=months)

####################################################
# Fit a 2-state Poisson HMM to the earthquake data from
# "Hidden Markov Models for Time Series" by
# Zucchini, MacDonald, and Langrock (2016)


#eq_dat <- data.frame("count" = quakes,"year" = year)

# create the HMM (non-stationary)
polio_mod <- depmix(Count ~ 1,
                    data = data_complete,
                    family = poisson(),
                    nstates = 2,
                    ntimes = length(months))

# fit the model using an EM algorithm
polio_fit <- fit(polio_mod)
```

```
## converged at iteration 39 with logLik: -260.0328
```

```r
# get the posterior distribution of the states
data_complete$states <- posterior(polio_fit, type = "viterbi")$state
```

```
# plot the earthquake data with the most likely state

ggplot(data_complete,
       aes(x = Time, y = Count)) +
  geom_point(size = .7, color = "darkblue") +
  geom_line(color = "darkblue")  +
  geom_text(label=data_complete$states,nudge_y =0.45)+
  ggtitle("Polio cases over time")
```

Polio cases over time
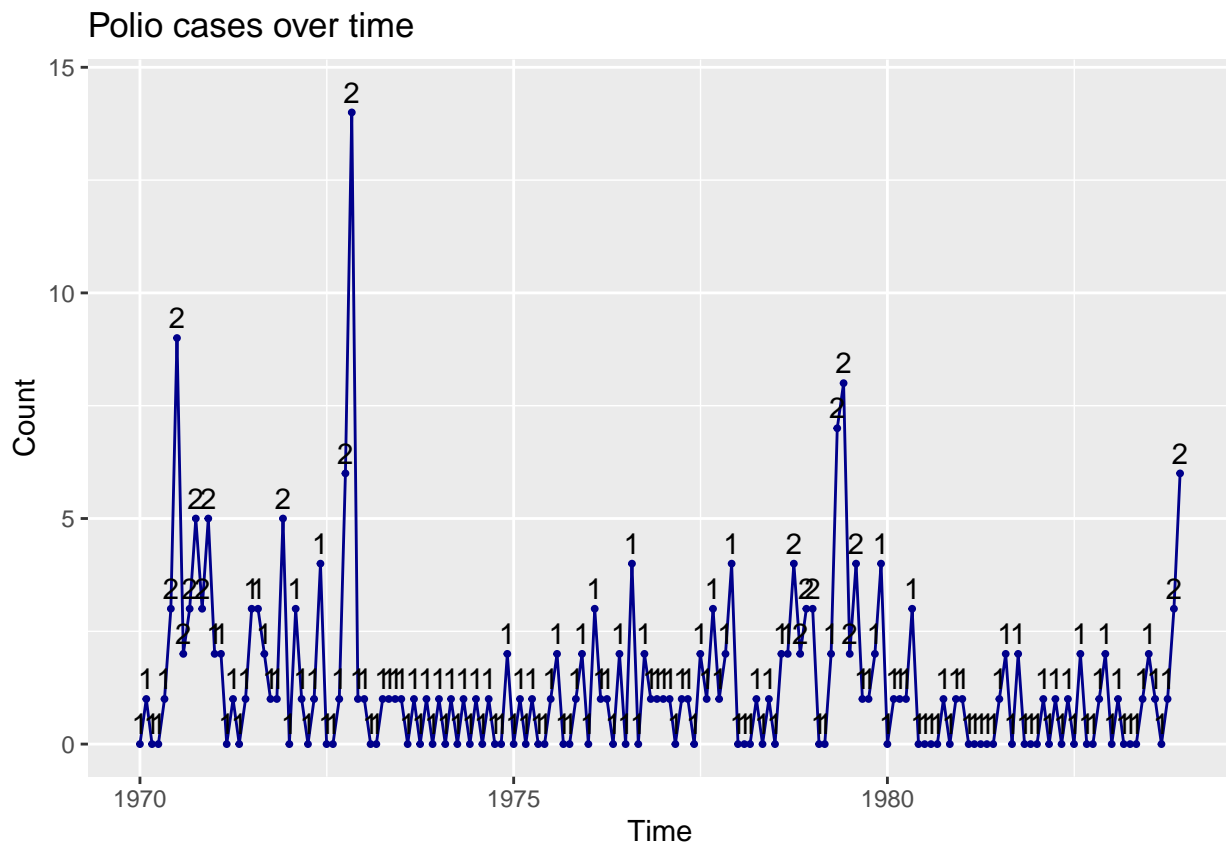
<div align="center">NUMERICAL VALUES-ANSWER</div>

```
## [1] "The AIC is"

## [1] 530.0655

## [1] "The estimated lambda values are"

## (Intercept) (Intercept)
##   0.7902001   4.1780950

## [1] "The transition matrix values are"

## [1] " From State 1 to State 1  0.932178581948805"

## [1] " From State 1 to State 2  0.0678214180511952"

## [1] " From State 2 to State 1  0.3304620304541"

## [1] " From State 2 to State 2  0.6695379695459"
```

Part b

```
#--------------------- Question b ------------------

count_1<-0
count_2<-0

count_g4<-0

for (i in 1:length(data_complete$Count))
{

  if(data_complete$Count[i]>=4)
  {
    count_g4<-count_g4+1
  }


  if (data_complete$states[i]==1 && data_complete$Count[i]>=4)
  {
    count_1<-count_1+1
  }

  if (data_complete$states[i]==2 &&data_complete$Count[i]>=4)
  {
    count_2<-count_2+1
  }
}

print(paste("P(C_t=1|X_t>=4)=",count_1/(count_g4)))
```

```
## [1] "P(C_t=1|X_t>=4)= 0.266666666666667"
```

```
print(paste("P(C_t=2|X_t>=4)=",count_2/(count_g4)))
```

```
## [1] "P(C_t=2|X_t>=4)= 0.733333333333333"
```

```
two_cases_prob<-1-ppois(2, lambda = exp(getpars(polio_fit)[8]))
zero_cases_prob<-dpois(2, lambda = exp(getpars(polio_fit)[8]))

print(paste("Two cases probability at time t+1",two_cases_prob))
```

```
## [1] "Two cases probability at time t+1 0.786848181252518"
```

```
print(paste("Zero cases probability at time t+1",zero_cases_prob))
```

```
## [1] "Zero cases probability at time t+1 0.133783642461472"
```

ANSWER

1. The $\lambda's$ represent the mean number of cases of people with polio per month for each state.
2. The steps to solve this question was to calculate the probability to be in state 1 or 2 given that the observed people with disease was greater or equal to 4. Then, once we know the state at time t, we use the transition matrix to know to what state is more probable to be at time t+1. Finally, once we know the state at time t+1, we calculate the probabilities to observe 0 patients with polio and to observe at least two patients with polio. What I obtained was that It is more probable to be in state 2 at time t given that number of observed patients with polio is greater or equal to 4. From this, I used the transition matrix. From the transition matrix I got that is much more probable to stay in state 2 than going to state 1 for time t+1. Hence, $C_{t+1} = 2$. Finally, I calculated the probabilities o f zero patients at time t+1 and at least 2 patients at tiume t+1 given that $C_{t+1} = 2$.

The conclusion obtained was that : It is more likely to observe 2 or more cases at time t+1 given that at time t we observed at least 4

3. The model is consistent with basic knowledge of diasease dynamics. If this month we observed more than four infected people, it is likely that those four will spread the disease to other people or keep infected at least for another month. Hence, it is more probable that we observe at least two cases with the desease than no observing it

---

Part c

```
#---------------------- Question c -------------------


months<-seq(as.Date("1970-01-01"), as.Date("1983-12-01"), by="months")


data_complete<-data.frame("Count"=polio,"Time"=months)
#####################################################
# Fit a 3-state Poisson HMM to the earthquake data from
# "Hidden Markov Models for Time Series" by
# Zucchini, MacDonald, and Langrock (2016)



# create the HMM (non-stationary)
polio_mod2 <- depmix(Count ~ 1,
                    data = data_complete,
                    family = poisson(),
                    nstates = 3,
                    ntimes = length(months))

# fit the model using an EM algorithm
aic_pol3<-600
for (i in 1:100000)
{

  if(aic_pol3>531)
  {polio_fit2 <- fit(polio_mod2)
    aic_pol3<-AIC(polio_fit2)
```

```
  }
  else

  {
    break
  }
}
```

```
## converged at iteration 340 with logLik: -254.8587
## converged at iteration 247 with logLik: -253.9778
```

```
# get the posterior distribution of the states
data_complete$states2 <- posterior(polio_fit2, type = "viterbi")$state


# plot the polio data with the most likely state

ggplot(data_complete,
       aes(x = Time, y = Count)) +
  geom_point(size = .7, color = "darkblue") +
  geom_line(color = "darkblue")  +
  geom_text(label=data_complete$states2,nudge_y =0.45)+
  ggtitle("Polio cases over time")
```

Polio cases over time

```
## [1] "The AIC is"
```

```
## [1] 529.9555
## [1] "The parameters lambda values are"
## (Intercept) (Intercept) (Intercept)
##     8.257020    2.297927    0.648656
## [1] "The transition matrix values are"
## [1] " From State 1 to State 1  0.28022946581211"
## [1] " From State 1 to State 2  0.614119307198108"
## [1] " From State 1 to State 3  0.105651226989782"
## [1] " From State 2 to State 1  0.0913000932173533"
## [1] " From State 2 to State 2  0.695523770576647"
## [1] " From State 2 to State 3  0.213176136206"
## [1] " From State 3 to State 1  2.06288019242838e-40"
## [1] " From State 3 to State 2  0.0926845985714025"
## [1] " From State 3 to State 3  0.907315401428597"
```