AMA Lab 5: Local Poisson Regression

Gerard Gómez, Rudio Fida Cyrille, Cecilia Pérez

2023-11-18

```
library(sm)
```

```
## Warning: package 'sm' was built under R version 4.2.3
```

```
## Package 'sm', version 2.2-5.7: type help(sm) for summary information
```

```
countries<-read.csv2(file="HDI.2017.subset.csv",row.names = 1)
attach(countries)
head(countries)
```

```
##     country_name Life.expec Life.expec.f Life.expec.m le.fm Inf.Mort.rat
## AFG  Afghanistan       64.0         65.4         62.8   2.6         53.2
## ALB      Albania       78.5         80.6         76.5   4.1         12.0
## DZA      Algeria       76.3         77.6         75.1   2.5         21.6
## AGO       Angola       61.8         64.7         59.0   5.7         54.6
## ARG    Argentina       76.7         80.4         73.0   7.4          9.9
## ARM      Armenia       74.8         77.8         71.4   6.4         11.9
##     Agric.employ..
## AFG           62.2
## ALB           40.3
## DZA           12.8
## AGO           50.6
## ARG            0.5
## ARM           34.4
```

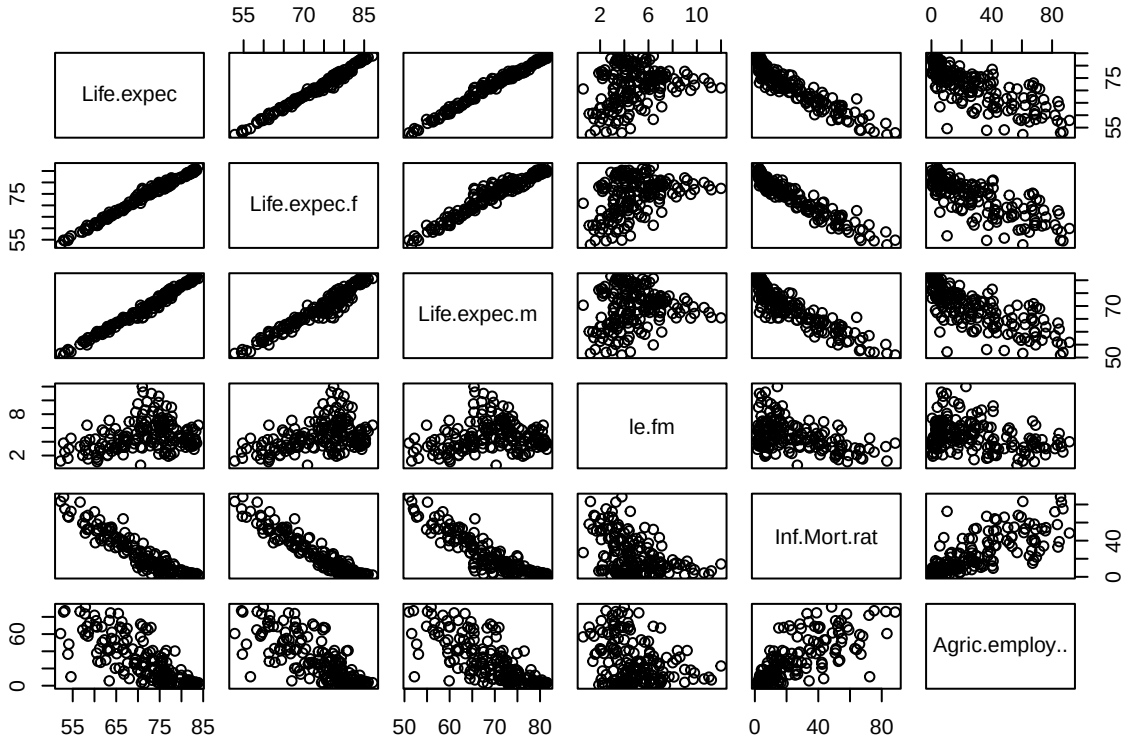
As a quick summary of the data, the mean life expectancy of the countries in this dataset is 71.85 being 73.40 as the median and for female 74.28 higher than the lif.expec of males, 69.45. The mean infant mortality rate is 23.26.

```
summary(countries)
```

```
##  country_name         Life.expec     Life.expec.f    Life.expec.m
##  Length:179         Min.   :52.20   Min.   :52.80   Min.   :51.00
##  Class :character   1st Qu.:66.55   1st Qu.:68.25   1st Qu.:64.55
##  Mode  :character   Median :73.40   Median :76.50   Median :70.40
##                     Mean   :71.85   Mean   :74.28   Mean   :69.45
##                     3rd Qu.:77.30   3rd Qu.:79.85   3rd Qu.:74.90
##                     Max.   :83.90   Max.   :87.10   Max.   :81.50
##      le.fm         Inf.Mort.rat   Agric.employ..
```

```
##  Min.   : 0.600   Min.   : 1.60   Min.   : 0.10
##  1st Qu.: 3.350   1st Qu.: 6.60   1st Qu.: 6.25
##  Median : 4.500   Median :15.10   Median :19.00
##  Mean   : 4.821   Mean   :23.26   Mean   :27.57
##  3rd Qu.: 6.050   3rd Qu.:37.40   3rd Qu.:41.95
##  Max.   :12.000   Max.   :88.50   Max.   :91.50
```

```
plot(countries[,2:7])
```



##1. Bandwidth choice for the local Poisson regression

```
loglik.CV.pois <- function(x,y,h){
  n <- length(x)
  pred <- sapply(1:n,
      function(i,x,y,h){
         sm.poisson(x=x[-i],y=y[-i],h=h,eval.points=x[i],display="none")$estimate
      },   x,y,h)
  return(1/n * sum(log(pred)))#I have my doubts for this line(I checked gerards' also but idk why i fee
}


h.cv.sm.pois <- function(x,y,rg.h=NULL,l.h=10,method=loglik.CV.pois){
   cv.h <- numeric(l.h)
   if (is.null(rg.h)){
      hh <- c(h.select(x,y,method="cv"),
              h.select(x,y,method="aicc"))#,hcv(x,y))
```

```
      rg.h <- range(hh)*c(1/1.1, 1.5)
   }
   i <- 0
   gr.h <- exp( seq(log(rg.h[1]), log(rg.h[2]), l=l.h))
   for (h in gr.h){
      i <- i+1
      cv.h[i] <- method(x,y,h)
   }
   return(list(h = gr.h,
               cv.h = cv.h,
               h.cv = gr.h[which.min(cv.h)]))
}
```

##2. Local Poisson regression for Country Development Data

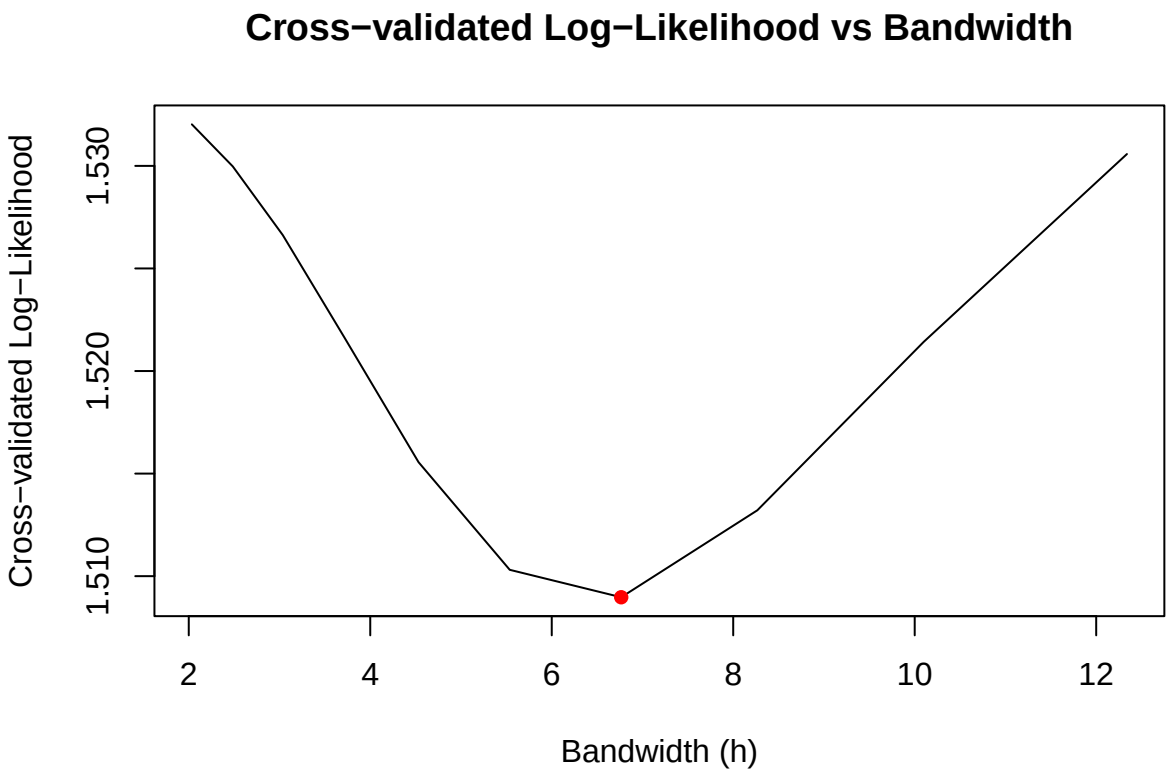
```
countries$le.fm.r <- round(countries$le.fm)
```

Here it is stated the code is performing cross-validation for Poisson kernel density estimation on a new dataset. Performs cross validation on the dataset provided with the values le.fm.r.

```
h.CV.pois.lefm <- h.cv.sm.pois(Life.expec, countries$le.fm.r, method = loglik.CV.pois)
```

Finally, the lineplot is plotted of cross-validated log-likelihood against bandwidth for a new dataset. In addition, a red point on the plot is added showing h for the minimum cross-validated log-likelihood.

```
plot(h.CV.pois.lefm$h, h.CV.pois.lefm$cv.h, type = "l",
     xlab = "Bandwidth (h)", ylab = "Cross-validated Log-Likelihood",
     main = "Cross-validated Log-Likelihood vs Bandwidth")
points(h.CV.pois.lefm$h.cv, min(h.CV.pois.lefm$cv.h), col = "red", pch = 16)
```


Cross−validated Log−Likelihood vs Bandwidth
Cross−validated Log−Likelihood
1.510
1.520
1.530
2
4
6
8
10
12
Bandwidth (h)