Local Poisson Regression
*

Andreu Meca, Jan Leyva, Victor Duque, Geraldo Gariza

3/22/2021

1. Bandwidth choice for the local Poisson regression

Modify the functions h.cv.sm.binomial and loglik.CV to obtain a band- width choice method for the local Poisson regression based on the leave-one-out cross-validation (loo-CV) estimation of the expected likelihood of an independent observation. Remember that the loo-CV estimation of the expected log-likelihood of an independent observation, when using h as bandwidth, is

$$l_{cv}(h) = \frac{1}{n}\sum_{i=1}^{n} log(\hat{Pr}_h^{-i}(Y = y_i|X = x_i))$$

Where $\hat{Pr}_h^{-i}(Y = y_i|X = x_i)$ is an estimation of

$$Pr(Y = y_i|X = x_i) = e^{-\lambda_i}\frac{\lambda_i^{y_i}}{y_i!},$$

should be estimated by maximum local likelihood using h as bandwidth (for instance, using the function sm.poisson from the R package sm).

```
# Changing the loglik function with sm.poisson

data <- read.csv("HDI.2017.subset.csv", sep = ";", dec = ',')


loglik.CV_poisson <- function(x,y,h){
  n <- length(x)-1
  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((sum(log(exp(-pred)*(pred^y)/factorial(y))))/n)
}



# Changing the h.cv with method= CV_poisson


h.cv.sm_poisson <- function(x,y,rg.h=NULL,l.h=10,
                            method=loglik.CV_poisson){
```

```
  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.max(cv.h)]))
}
```

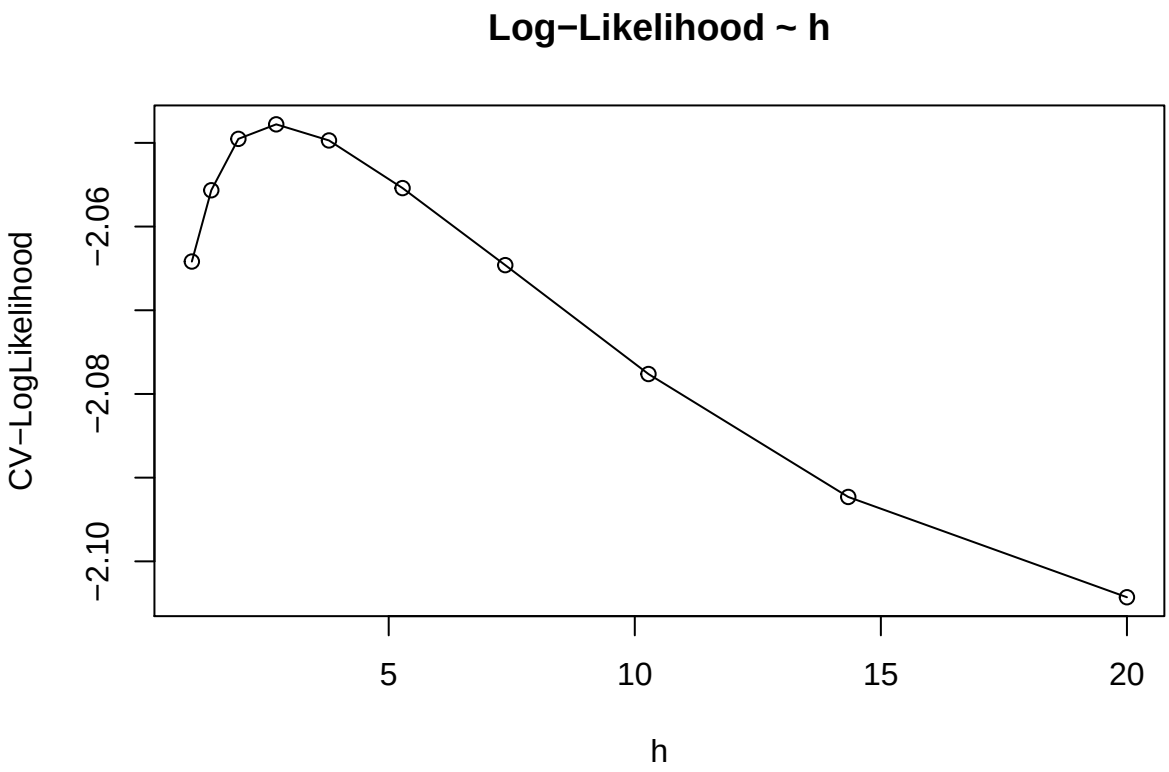
2. Local Poisson regression for Country Development Data

Consider the country development dataset (file HDI.2017.subset.csv) con- taining information on development indicators measured in 179 countries (Source: Human DevelopmentData(1990-2017). Variable le.fm always takes non-negative values. Define le.fm.r as the rounded value of le.fm: le.fm.r <- round(le.fm) Fit a local Poisson regression modeling le.fm.r as a function of Life.expec. Use sm.poisson from the R package sm with the bandwidth obtained by loo-CV.

```
library(sm)
```

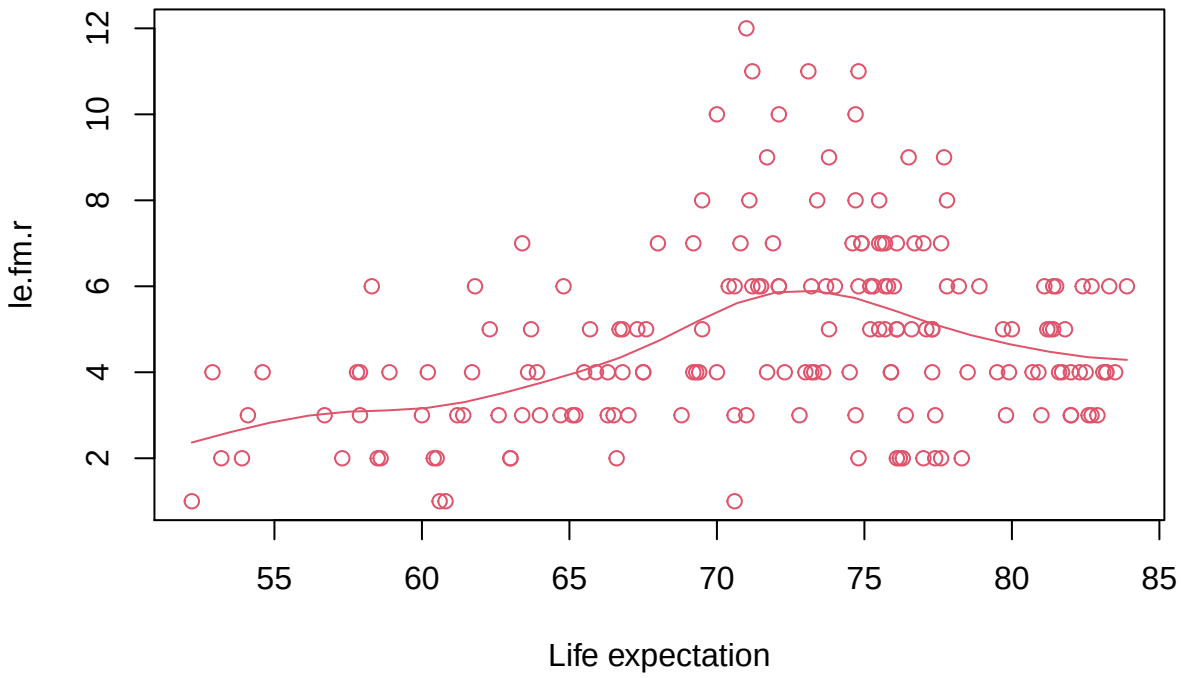
```
## Package 'sm', version 2.2-5.6: type help(sm) for summary information
```

```
data <- read.csv("HDI.2017.subset.csv", sep = ";", dec = ',')

le.fm.r <- round(data$le.fm)


## 2. Local Poisson regression for Country Development Data ------------------
h.CV.loglik <- h.cv.sm_poisson(y = le.fm.r,x = data$Life.expec,rg.h=c(1,20),
                               method=loglik.CV_poisson)

plot(h.CV.loglik$h,h.CV.loglik$cv.h, main="Log-Likelihood ~ h",
     xlab="h", ylab="CV-LogLikelihood")
lines(h.CV.loglik$h,h.CV.loglik$cv.h)
```



We have calculated the maximum likelihood of the h by leave one out cross-validation,in order to obtain the best value of h: 2.7144176.

```
sm.poisson(y = le.fm.r,x = data$Life.expec,h =h.CV.loglik$h.cv,xlab="Life expectation")
```



After that, We have used the $\hat{h}$ to represent the Local Poisson Regression with the function *sm.poisson* of the packcage *sm*. We can observe with values around of 65 and 80 **life expectation** We get the highest values of the response variable **le.fm.r**.