

Local_poisson_regression

Gregoire Gasparini, Aurora Hofman, Sarah Musiol, Beatriu Tort

22 de marzo de 2020

Bandwidth choice for the local Poisson regression

The functions from ATENEA are modified to fit a local Poisson regression.

```
#The first function is not modified
h.cv.sm.poisson <- function(x,y,rg.h=NULL,l.h=10,method=loglik.CV){
  cv.h <- numeric(l.h)
  if (is.null(rg.h)){
    hh <- c(h.select(x,y,method="cv"),
            h.select(x,y,method="aicc"))
    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)]))
}

# method loglik.CV: leave-one-out log-likelihood

#Modifications: sm.poisson and the return function is the loglikelihood
#of the density of the poisson distribution.

loglik.CV <- 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)
  #faculty <- sapply(1:n, function(i, y){prod(1:y[i])},
  # if one wants the scalingfactor uncomment the line above
  return(-sum(y*log(pred)-pred)) #-log(faculty) only a scale
}
```

Local Poisson regression for Country Development Data

Consider the country development dataset containing information on development indicators measured in 179 countries.

```
countries<-read.csv2(file="HDI.2017.subset.csv",row.names = 1)
le.fm.r <- round(countries$le.fm)
```

Fit a local Poisson regression modeling `le.fm.0` as a function of `Life.expec`. Use `sm.poisson` from R package.

```
library(sm)
```

```
## Warning in fun(libname, pkgname): couldn't connect to display ":0"
```

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

```
x<- countries$Life.expec
y<- le.fm.r
```

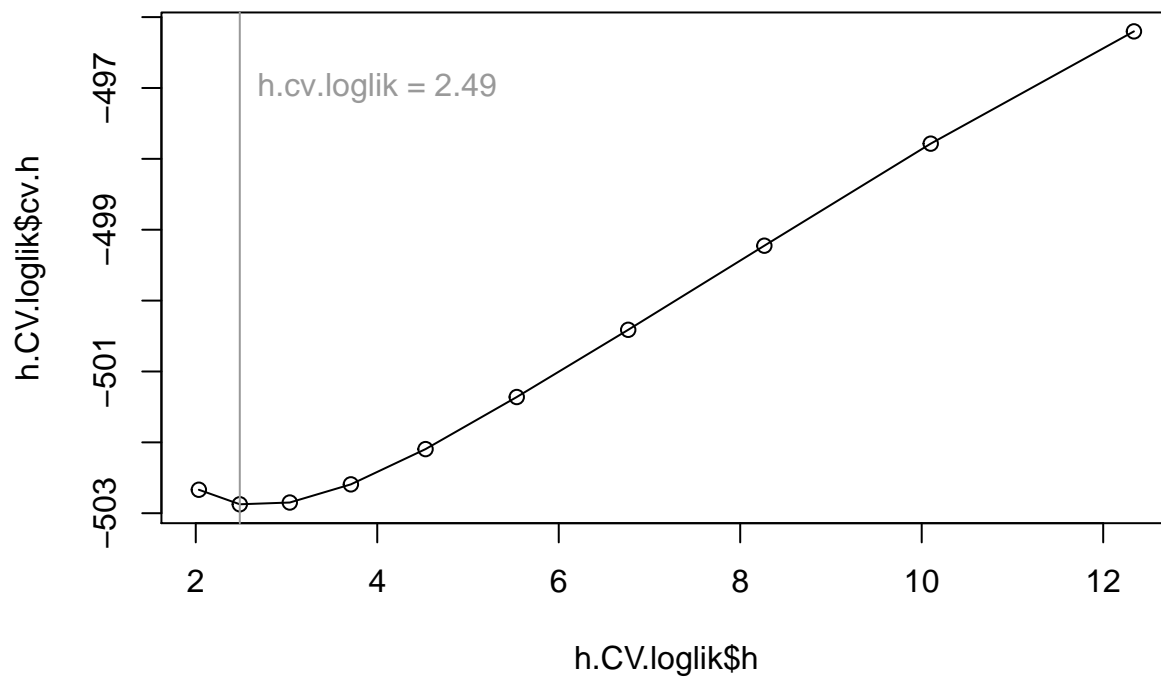
```
h.cv.sm.poisson(countries$Life.expec, le.fm.r)$h.cv
```

```
## [1] 2.485047
```

Plot of the loglikelihood as a function of Bandwidth.

```
h.CV.loglik <- h.cv.sm.poisson(countries$Life.expec,le.fm.r, l.h = 10)
plot(h.CV.loglik$h,h.CV.loglik$cv.h)
lines(h.CV.loglik$h,h.CV.loglik$cv.h)
abline(v=h.CV.loglik$h.cv, col = "gray60")
text(4, -497, paste("h.cv.loglik = ",round(h.CV.loglik$h.c,2),sep=""), col = "gray60")
title("Bandwidth cross - validation plot")
```

Bandwidth cross – validation plot



In this plot it can be seen that the optimal bandwidth is around 2.5, which is where the -loglikelihood is at

its minimum.

Now we make a local poisson regression where le.fm.r is plotted against life.expectency.

```
aux <- sm.poisson(countries$Life.expec,le.fm.r,h=h.CV.loglik$h.cv)
title(main=paste("h.cv.loglik = ",round(h.CV.loglik$h.cv,2),sep=""))
```

