Local poisson regression

Group 1: Pariente Antonio, Bosch Guillem, Ebner Lena

21/11/2023

Local Poisson Regression

---

1. Bandwidth choice for the local Poisson regression

Modify the functions `h.cv.sm.binomial` and `loglik.CV` to obtain a bandwidth 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

$$\ell_{CV}(h) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\widehat{\Pr}^{(-i)}(Y = y_i|X = x_i)\right),$$

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

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

and

$$\lambda_i = \mathbb{E}(Y|X = x_i),$$

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

```
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", family="poisson"),
              h.select(x,y,method="aicc", family="poisson"))
      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
loglik.CV <- function(x,y,h){
  n <- length(x)
  loglik_values <- numeric(n)
  for (i in 1:n) {
    pred <- sm.poisson(x = x[-i], y = y[-i], h = h, eval.points = x[i], display = "none")$estimate
    loglik_values[i] <- log(dpois(y[i], lambda = pred))
  }
  return(-sum(loglik_values) / n)
}

# ORIGINAL method
# method loglik.CV: leave-one-out log-likelihood
# loglik.CV <- function(x,y,h){
#   n <- length(x)
#   pred <- sapply(1:n,
#       function(i,x,y,h){
#          sm.binomial(x=x[-i],y=y[-i],h=h,eval.points=x[i],display="none")$estimate
#       },   x,y,h)
#   return(-sum( y*log(pred/(1-pred)) + log(1-pred) )/n)
# }
```

2. Local Poisson regression for Country Development Data

Consider the country development dataset (file `HDI.2017.subset.csv`) containing information on development indicators measured in 179 countries (Source: Human Development Data (1990-2017), The Human Development Report Office, United Nations). 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)
```

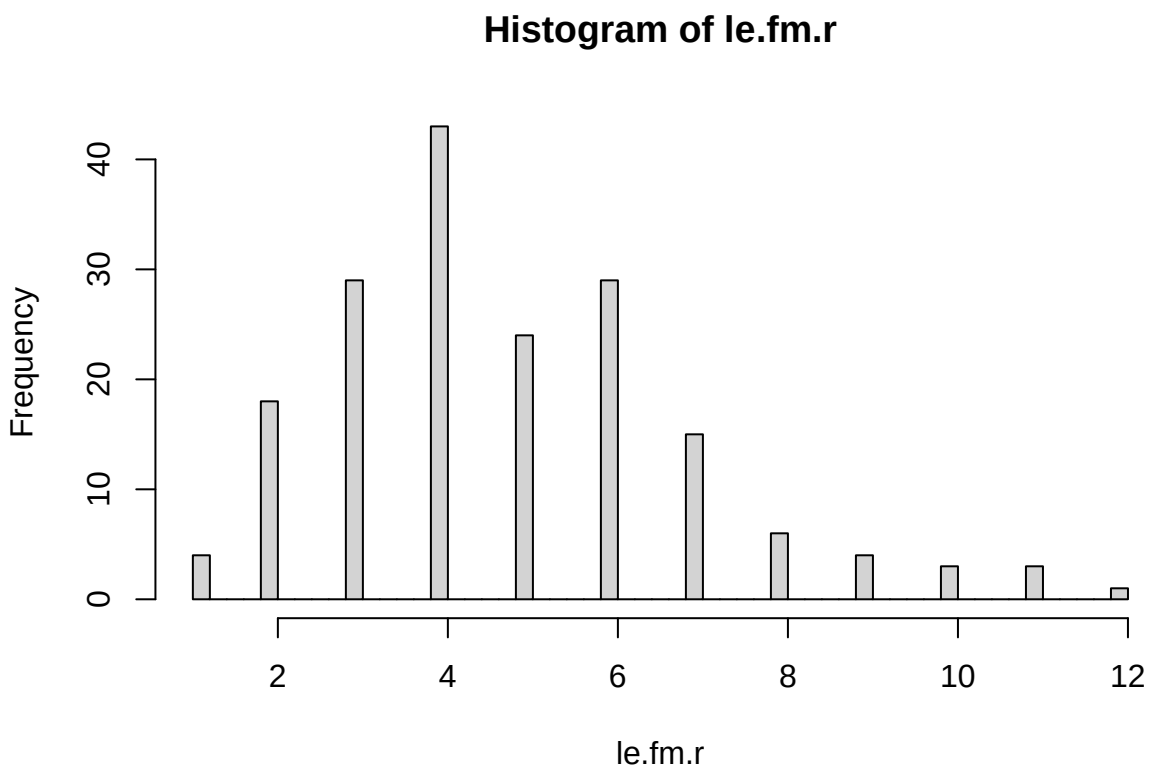
```
## Package 'sm', version 2.2-5.7: type help(sm) for summary information
```

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

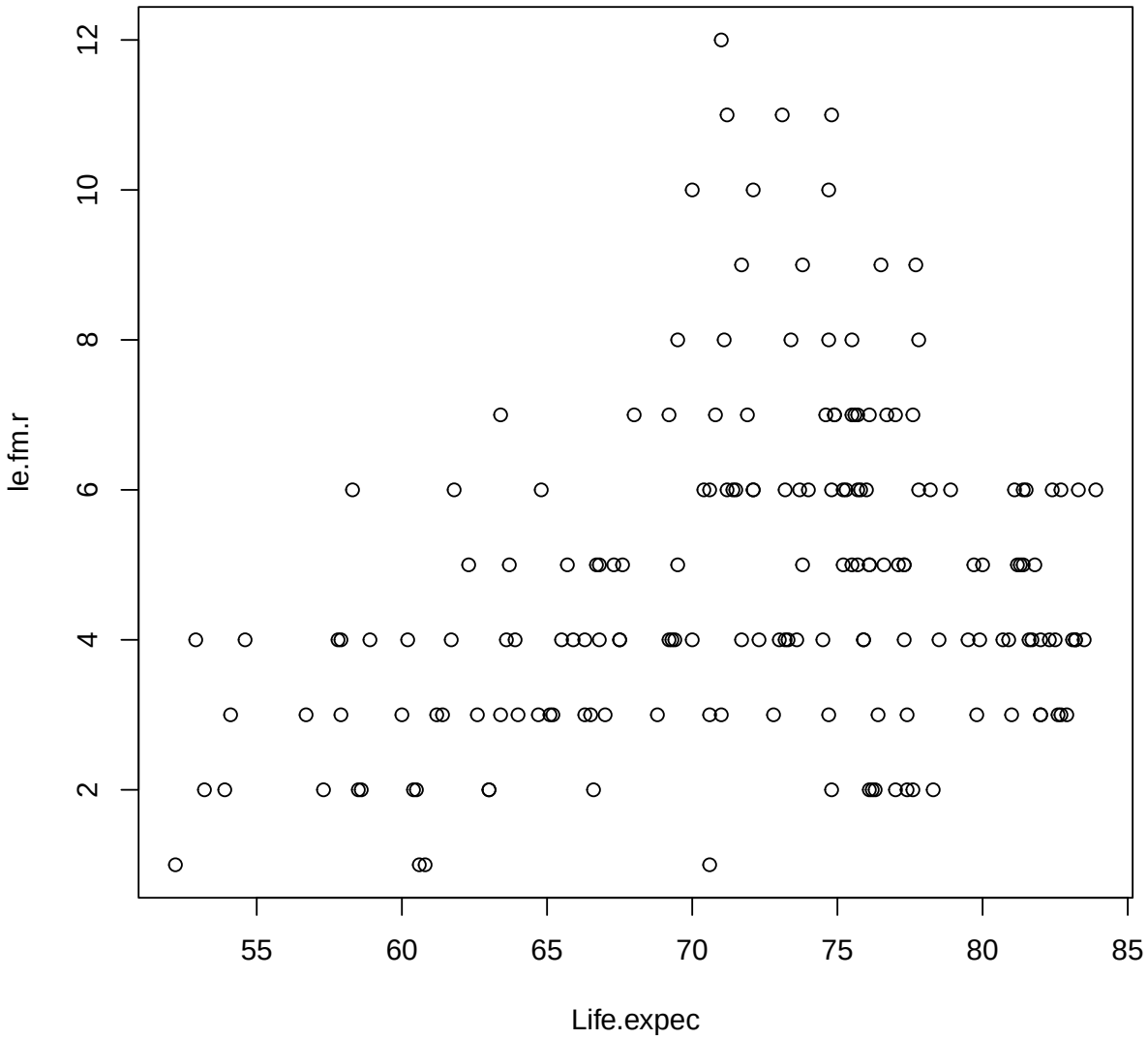
```
##     country_name Life.expec Life.expec.f Life.expec.m le.fm Inf.Mort.rat
## AFG  Afghanistan       64.0         65.4         62.8   2.6         53.2
## ALB      Albania       78.5         80.6         76.5   4.1         12.0
## DZA      Algeria       76.3         77.6         75.1   2.5         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
```

```r
le.fm.r <- round(le.fm)
hist(le.fm.r,br=40)
```

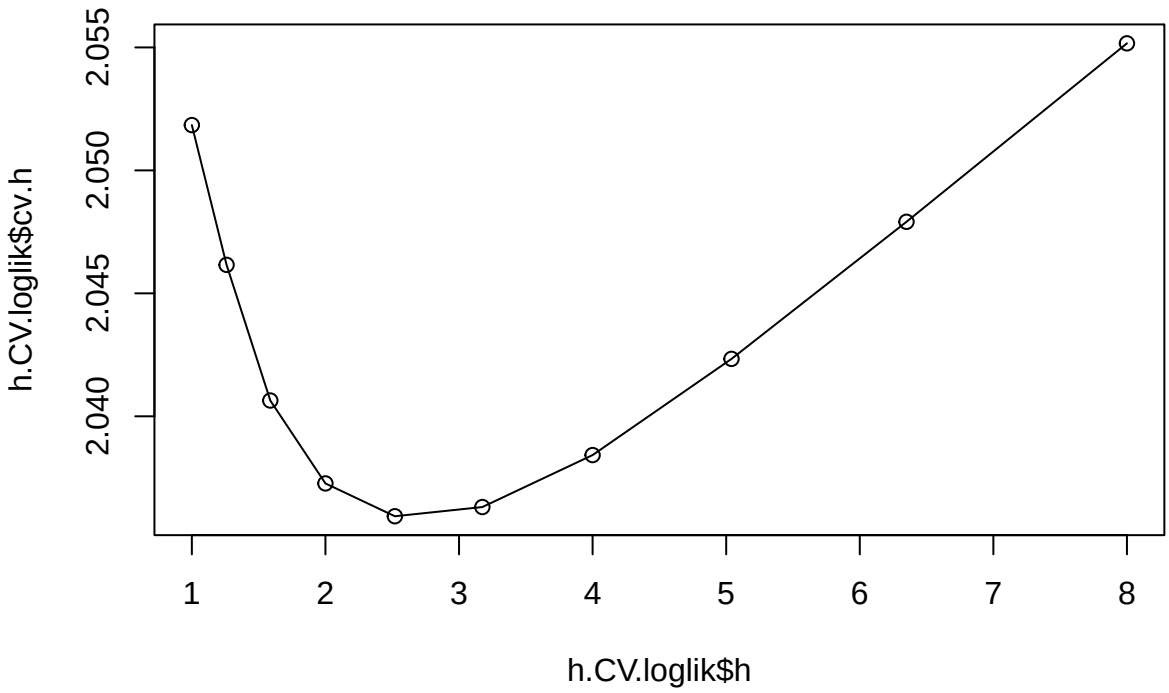



```r
plot(Life.expec,le.fm.r)
```



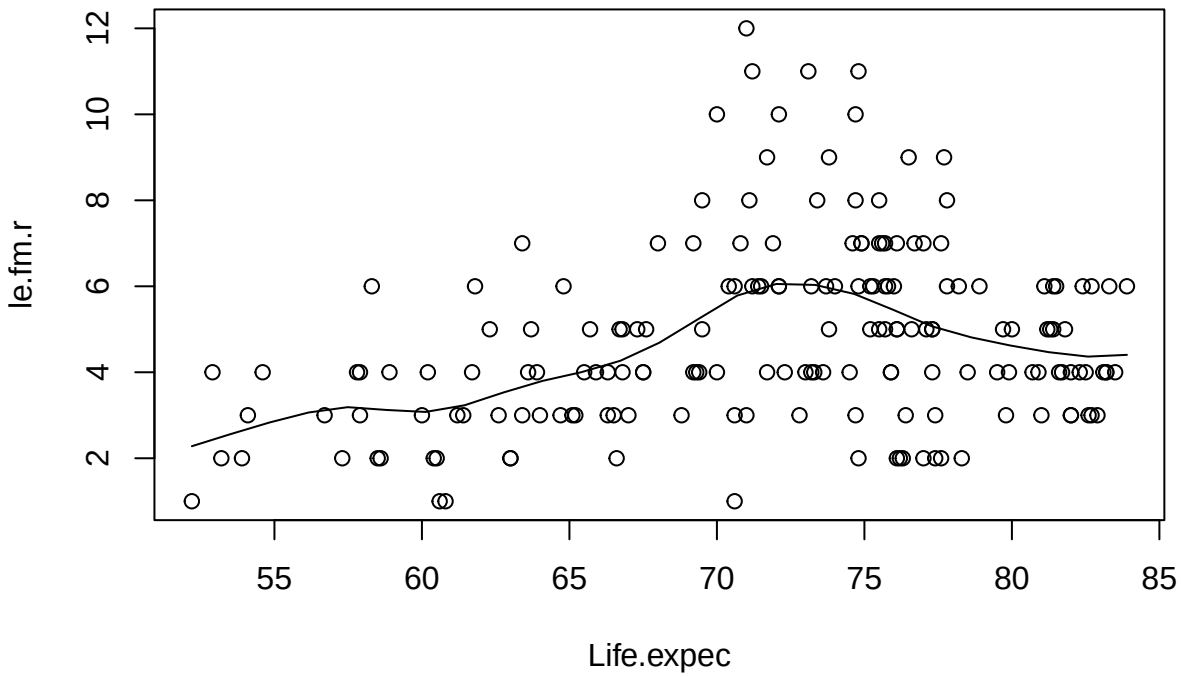
Best Bandwidth Choice

```
# using h.CV.loglik
h.CV.loglik <- h.cv.sm.poisson(Life.expec, le.fm.r, rg.h=c(1,8),method=loglik.CV)
plot(h.CV.loglik$h,h.CV.loglik$cv.h)
lines(h.CV.loglik$h,h.CV.loglik$cv.h)
```




```
library(sm)
sm.poisson(Life.expec,le.fm.r,h=h.CV.loglik$cv.h,col=1)
```

```
## Warning in W/h: longer object length is not a multiple of shorter object length
```


le.fm.r
2
4
6
8
10
12
55
60
65
70
75
80
85
Life.expec