

# Local Poisson regression

Biel Caballero Vergés, Svenja Menzenbach and Kleber Enrique Reyes Illescas

2023-11-22

## 1. Bandwidth choice for the local Poisson regression

In this section, we have modify the *h.cv.sm.binomial* (now *h.cv.sm.poisson*) and the *loglik.CV* functions in order to be able to use it for the local Poisson regression. We have simplified the code to have less parameters.

```
h.cv.sm.poisson <- function(x, y, rg.h = NULL, l.h = 10) {  
  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)  
  }  
  
  gr.h <- exp(seq(log(rg.h[1]), log(rg.h[2]), length.out = l.h))  
  
  for (i in 1:length(gr.h)) {  
    cv.h[i] <- loglik.CV(x, y, gr.h[i])  
  }  
  
  return(list(h = gr.h, cv.h = cv.h, h.cv = gr.h[which.max(cv.h)]))  
}
```

```
loglik.CV <- function(x, y, h) {  
  n <- length(x)  
  
  lambda <- 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(-lambda + y*log(lambda) -log(factorial(y)) )/n )  
}
```

## 2. Local Poisson regression for Country Development Data

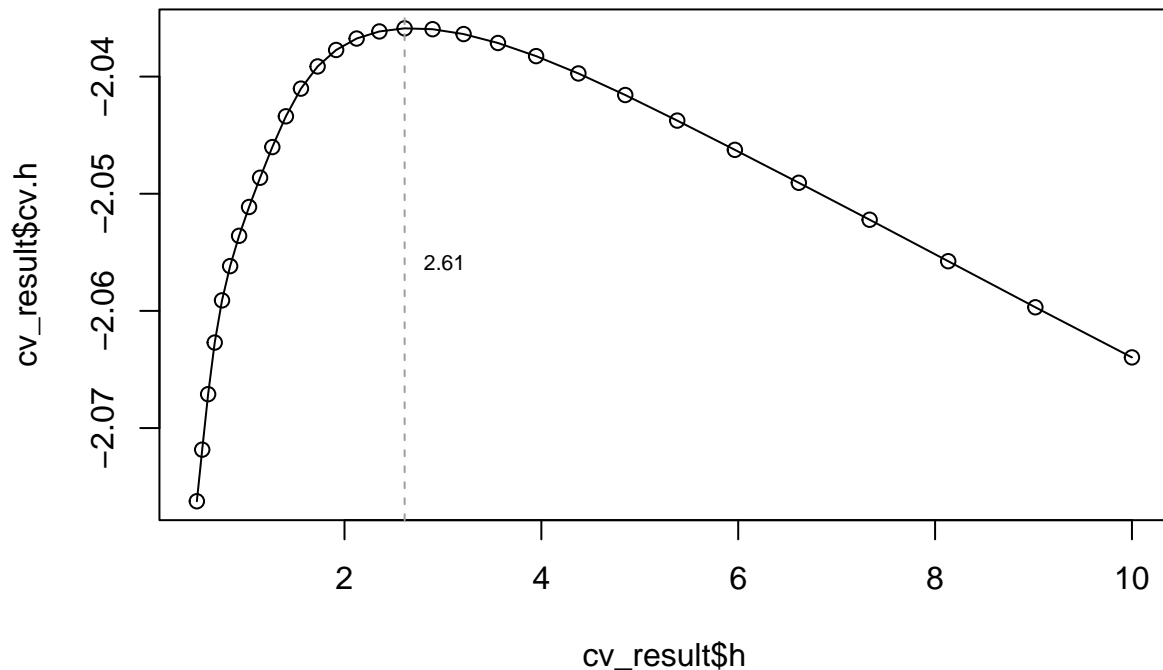
With our new function, we can acquire an optimal bandwidth for fitting our model.

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

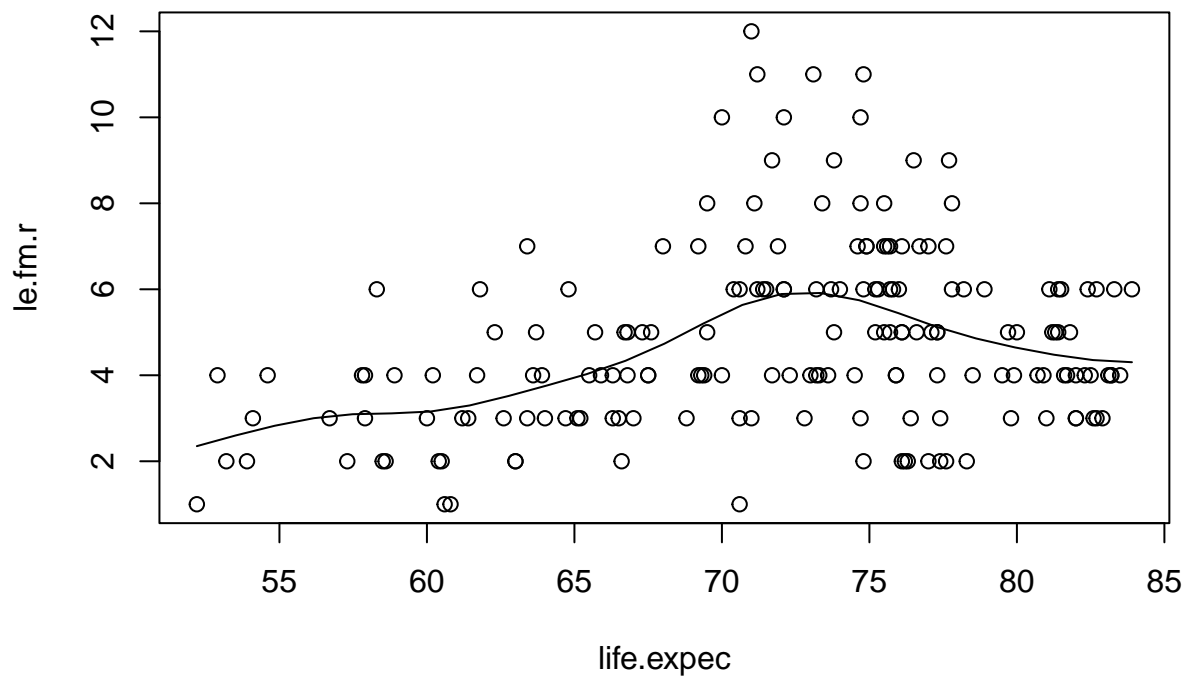
life.expec <- countries$Life.expec
le.fm.r <- round(countries$le.fm)

cv_result <- h.cv.sm.poisson(life.expec, le.fm.r, rg.h=c(0.5,10), l.h = 30)

plot(cv_result$h, cv_result$cv.h)
selected.bandwidth <- cv_result$h.cv
abline(v = selected.bandwidth, col="8", lty=2)
mid.point <- (max(cv_result$cv.h)+min(cv_result$cv.h))/2
text(selected.bandwidth, mid.point,
      round(selected.bandwidth*100)/100,cex=0.65, pos=4,col=1)
lines(cv_result$h, cv_result$cv.h)
```



```
model <- sm.poisson(x = life.expec, y = le.fm.r, h = selected.bandwidth, col=1)
```



The leave-one-out cross-validation determines 2.61 as optimal bandwidth. The fit of the local Poisson regression looks smooth and reasonable. It follows the course of high density of points while generalizing to all data. We do not observe overfitting.