Task 4 : Bandwidth choice in the local Poisson model

Martin Guy and Hannes Leskela

5 décembre 2016

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Introduction

The goal of this exercise is to implement bandwidth choice functions for the local Poisson regression, and fit that model working with the file `countries.txt`. This file contains information on development indicators measured in 132 countries (Source: World Bank,1992). We will then compare our model with a standard nonparametric regression fit (with `sm.regression`) and a parametric fitting of a Poisson Generalized Linear Model (using `glm`).

Analysing the data

First look

We will be working with the file countries.txt containing information on development indicators measured in 132 countries. We will focus on the following variables: - life.exp (Life expectancy at birth) - inf.mort (Infant mortality rate) - le.fm (Difference Life expectancy at birth for females minus Life expectancy at birth for males)

The variable `le.fm` always takes non-negative values, except for one country, so we get rid of it and now consider `le.fm.0`. Here is a histogram of `le.fm.0`.

Histogram of le.fm.0
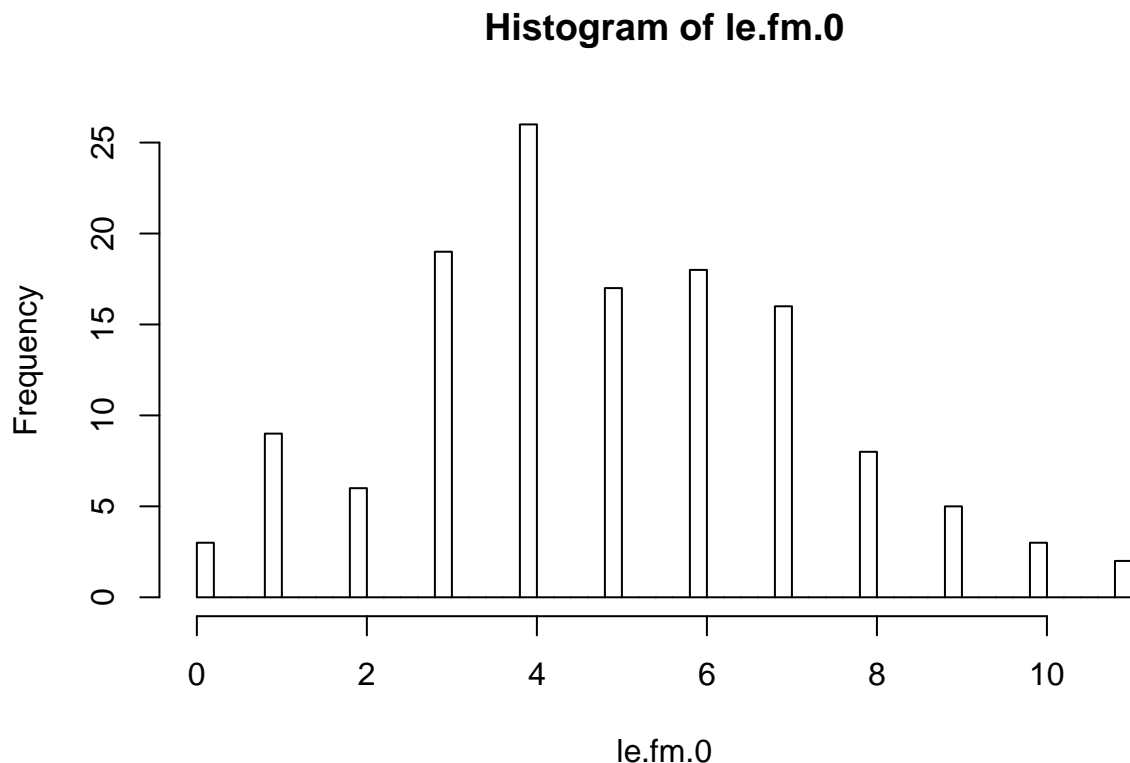
We can observe than in every country (except one) the life expectancy at birth for females is higher than for males.

Choice of the bandwidth

We modified the `h.cv.sm.binomial.R` file to a `h.cv.sm.poisson.R` file which calculates the bandwidth for a local Poisson regression. Using it we find a value for the bandwidth for our local Poisson model:
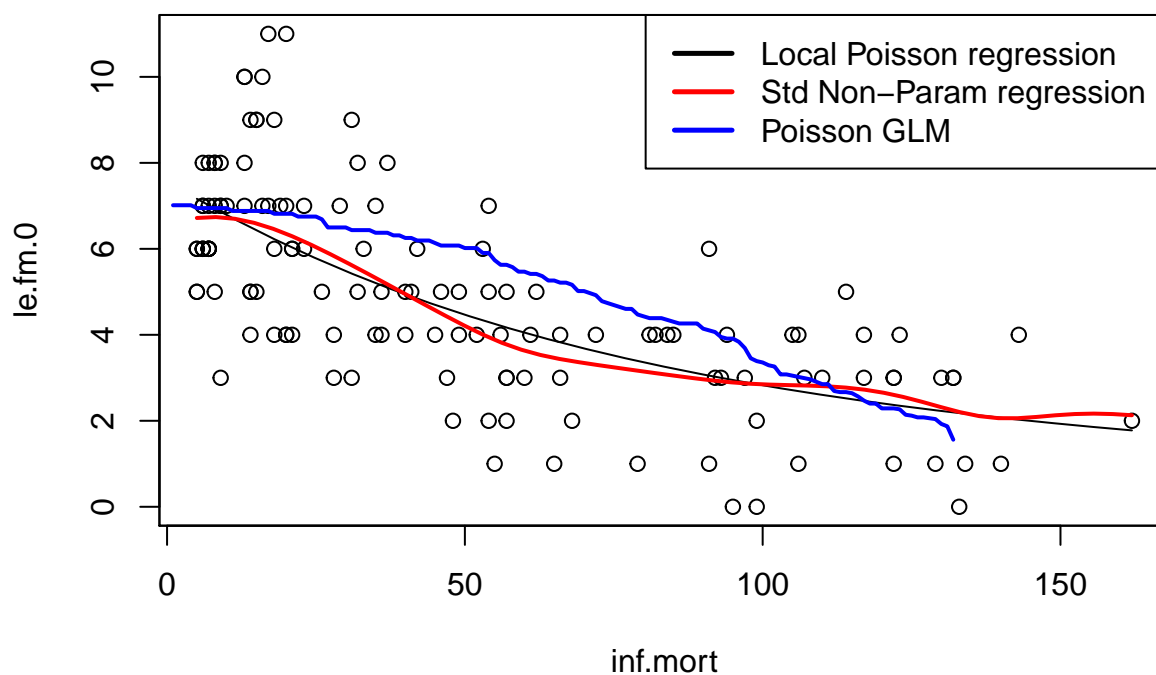
```
h1.CV.prob <- h.cv.sm.poisson(inf.mort, le.fm.0, method=prob.missclas.CV)
(h1 <- h1.CV.prob$h.cv)
```

```
## [1] 58.28751
```

First models: le.fm.0 as a function of inf.mort

First, we will modelize le.fm.0 as a function of infant mortality rate. In every case we will use local Poisson regression (black), a standard nonparametric regression fit (red), using a Poisson Generalized Linear Model (blue).

```
## [1] 58.28751
```


Second models: le.fm.0 as a function of life.exp

Then, we will modelize le.fm.0 as a function of life expectancy at birth.
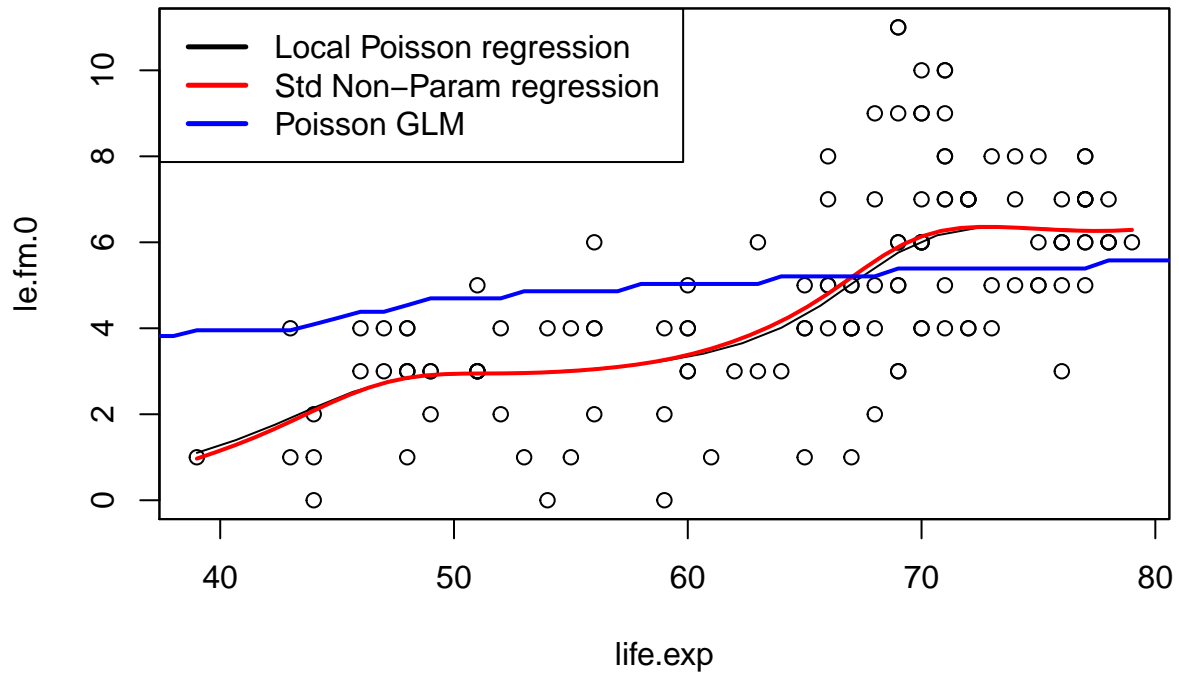
Conclusion

We observe in both models that non-parametric models seem to be better than the Poisson Generalized Linear Model. Indeed, the non-parametric models can fit better the data than parametric models, as we do not force the "shape" of predictors in the first case.