Université catholique de Louvain
Institut de Statistique

LSTAT 2150
Statistique non paramétrique : Méthodes de lissage

Project
Number 22

Project on variance estimation for non-parametric regression

1. Estimate the variance function in the following regression model by an approach which is guaranteed to produce a non-negative estimator.

   Simulate $\{(X_i, Y_i)\}_{i=1}^n$, $n = 100$, from the following regression model:

   $$Y_i = m(X_i) + 0.5\ \sigma(X_i)\ \varepsilon_i, \quad i = 1, \ldots, n\ ,$$

   where $X_i = i/n\ \in (0, 1]$, $\{\varepsilon_i\} \sim\ \mathcal{N}(0, 1)$, and

   $$m(x)\ =\ 3\ \sqrt[4]{x} \cdot \cos(1.2/(x + 0.05))$$

   and

   $$\sigma^2(x)\ =\ 2\ +\ \sin(2\ \pi\ x)$$

   Investigate bias, variance and IMSE by numerical results of a Monte Carlo simulation study (using different sample sizes, e.g. $n = 25, 50, 100, 500, 1000$).
   Do the results depend strongly on which estimator to use for the function $m(x)$? Compare with at least two different estimators for $m(x)$, a kernel estimator and a non-kernel estimator.

2. Hint: for the choice of the smoothing parameters use some default method proposed in R. Are your results very sensitive to which choice you take? Test a couple of combination of the two smoothing parameters, the one for estimating $m(x)$ and the one for estimating $\sigma^2(x)$.

Requirements: The front page should contain the name of the course, the title of the project, the name of the student, and the date. Be explicit in the description of your work. Explain all the steps. Do not write unneeded theory. Your work should be consistent. However, do not exceed 10 pages of main project description (additional figures, tables

and code can be put into an annex if necessary). Provide numerical and/or graphical results when needed. Write appropriate descriptions for the tables and figures and do not forget to use legend(). Provide your [R] code in the annex to the project with short comments.

Date of submission: please see on the Moodle site