Nonparametric admissible estimators

E. Matzner-Løber

P-A. Cornillon, A. Gribinsky, N. Hengartner, T. Kerdreux, N. Klutchnikoff

CIRM 2016

Regression

$(X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$ n pairs of observations

$$Y_i = m(X_i) + \varepsilon_i$$

$$Y = m + \varepsilon.$$

Aim : estimation of m the unknown function.

Smoothing

We estimate m np (smoothing) :

$$\hat{m} = S_\lambda Y$$

where

- S_λ is the smoothing matrix
- λ is the smoothing parameter (bandwidth, penalty coefficient, number of neighboors...).

Classical smoother

- moving average $S_{ij} = 1/$ number of X in the neighborhood
- binning $S_{ij} = 1/$ number of X in the bin
- kernel $S_{ij} = K_h(X_i - X_j)/\sum_l K_h(X_i - X_l)$
- regression spline $S = B(B'B)^{-1}B'$
- smoothing spline $S = N(N'N + \lambda\Omega_N)^{-1}N'$
- $k$ppv and mutual mkppv

Main idea

Assume λ big, so that the initial smoother is very smooth ; then

- estimate the bias
- correct the previous smoother

and iterate.

Procedure

- choose a smooth pilot S_1 and get $\hat{m}_1 = S_1 Y$.
- evaluate the conditionnal bias $B(\hat{m}_1) = (S_1 - I)m$
- estimate it by for example $(S_1 - I)S_1 Y$
- correct $\hat{m}_1$ and get

$$\hat{m}_2 \quad = \quad S_1 Y - (S_1 - I)S_1 Y = [I - (I - S)^2]Y$$

- Iterate

$$\widehat{m}_k \quad = \quad [I - (I - S)^k]Y.$$

Bias correction

The behaviour of $\widehat{m}_k$ is related to the spectrum of $I - S$.

Smoothing Spline, Duchon spline, Thin plate spline... OK

If the kernel is not positive definite as for example Epanechnikov, uniform, S admits negative eigen values.

What is a meaning of a smoother with negative eigen values ?

Admissibility

Consider linear estimator $\hat{m} = MY$ and quadratic risk

$$\begin{aligned} R(\hat{m}, m) &= \mathbb{E}(\hat{m} - m)'(\hat{m} - m) \\ &= \sigma^2 tr(M'M) + m'(I - M)'(I - M)m. \end{aligned}$$

Cohen (1966) showed that a linear estimator $\hat{m} = MY$ is admissible among all linear estimator iff

- M is symmetric
- the eigenvalues of M are in $[0, 1]$

Most of the smoother are not symmetric, in this talk we propose a procedure which suitably modify the initial smoother.

Symmetrisation

$$\sigma^2 tr(S'S) + m'(I - S)'(I - S)m.$$

1. arithmetic mean $S_a = (S + S')/2$
2. geometric mean $S_g = (S'S)^{1/2}$ which conserves the variance
3. conserving the bias (Cohen) $S_C = I - [(I - S)'(I - S)]^{1/2}$
4. ... with a step factor (Zhao) $S_Z = I - \rho[(I - S)'(I - S)]^{1/2}$
5. if S could be written as WA with W a diagonal matrix of positive weights and A symmetric, define $S_w = W^{1/2} A W^{1/2}$
6. ...

Averaging is not a good idea

If the initial smoother is row stochastic than the maximum eigen value of S_a or S_g is strictly bigger than 1.

Idea of the proof, using Rayleigh coefficient and find a vector u such that $u'Su > u'u$.

Negative eigen values

If the initial smoother has negative eigen values, S_c and S_w are not admissible. For S_Z it will depend on the choice of ρ.

Kernel smoother

Recall that $S_{ij} = K_h(X_i - X_j)/\sum_l K_h(X_i - X_l)$

If the initial smoother uses a positive definite kernel (Gaussian), S_c, S_Z and S_w are admissible.

However if one wants a smoother which could be evaluated at any points, we recommand to use S_W so $\hat{m} = W^{1/2}KW^{1/2}$

Interpretation and evaluation at any x

$$
\begin{aligned}
\hat{m}_{NW}(x) &= \frac{1}{\hat{n}hf_n(x)} \sum_i K_h(X_j - X_i) Y_i \\
\hat{m}_{MM}(x) &= \sum_i K_h(X_j - X_i) \frac{Y_i}{\hat{n}hf_n(X_i)}
\end{aligned}
$$

The new estimator is a mixture of the classical NW estimator and the intern estimator (Mack and Müller 1989) :

$$
\hat{m}_W(x) = \frac{1}{\sqrt{nh\hat{f}_n(x)}} \sum_{i=1}^{n} K_h(x - X_i) \frac{Y_i}{\sqrt{nh\hat{f}_n(X_i)}}.
$$

Theoretical properties in progress.

K-nn, 1950

$$\mathcal{S}_n = \{X_1, \cdots, X_n\} \quad \mathcal{N}_i = \{X_j \in \mathcal{S}_n | X_j \text{ is } K\text{-nn of } X_i\}$$

$$S_{ij}^{Knn} = \begin{cases} \frac{1}{k} & \text{if } X_j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases} .$$

The adjacency matrix associated to the K-nn direct graph is M.

$\sum_{j=1}^{n} M_{ij} = k$.

$H_j = \sum_{i=1}^{n} M_{ij}$ which counts the number of neighboor X_j belongs to, is a random variable, if H_j is bigger than k, X_j is called a hub.

Mutual k-nn, 1978

$$\mathcal{M}_i = \{X_j \in \mathcal{S}_n | X_j \in \mathcal{N}_i \text{ and } X_i \in \mathcal{N}_j\}$$

The adjacency matrix associated is

$$\widetilde{M}_{ij} = \begin{cases} 1 & \text{if } X_j \in \mathcal{M}_i \\ 0 & \text{otherwise} \end{cases}$$

The number of mk-nn of X_i, $K_i = \sum_{j=1}^n \widetilde{M}_{ij}$ is bounded by k.

$$S^{Mnn} = W\widetilde{M}$$

$$W = \text{diag}(\dots \frac{1}{\sum_{j=1}^n \widetilde{M}_{ij}} \dots)$$

Mutual

Since we could write S^{Mnn} as $W\widetilde{M}$, we could symmetrise by

$$S_W = W^{1/2}\widetilde{M}W^{1/2}$$

Let us consider the quadratic form associated to $\widetilde{M}$.
Consider $\widetilde{M}$ with $k \geq 3$. If among $\mathcal{S}_n$ there is 3 points (A, B, C) st (A,B) and (B,C) are mutual and (A,C) are not than

$$\lambda_{min}(\widetilde{M}) < 0$$

Proof : take the vector u which is 0 everywhere except at position A (resp B and C) with values -1 (resp 2 et -1) than the quadratic form

$$u^t\widetilde{M}u = -2 < 0$$

Connexion with graphs

Consider the graph with 3 nodes and 2 edges, the matrix is

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

and the eigen value is $1 - \sqrt{2}$ ($(1 + 2cos(k\pi/(n + 1))$.

The eigen values are not negative if all the subgraphs are complete.

For K-nn ?

Since M the adjacency matrix is not symmetric, one has to work with Cohen or Zhao estimator.

We cand proove that the max eigen value of $(I - S)(I - S)^t$ is strictly bigger than 1 so S_C is not admissible and for S_Z it depends again of the choice of ρ.

When smoother admits negative eigen values, Zhao proposed to replace the negative values by 0 or by its absolute values. Let us propose a general simple procedure.

General procedure

Start with the adjacency matrix A (or a smoother)

1. evaluate $N_1 = AA'$ or $N_2 = A'A$.
2. find W_i such that $W_i N_i$ is row stochastic
   $W = \text{diag}(\ldots \frac{1}{\sum_{j=1}^{n} N_{ij}} \ldots)$
3. evaluate $S_i = W_i^{1/2} N_i W_i^{1/2}$.

This estimator is admissible because
(1) it is symmetric
(2) the eigen values are in $[-1, 1]$ because it is row stochastic
(3) the eigen values are in $[0, 1]$ because it is positive definite.

Interpretation and evaluation at any point

$(AA')_{ij}$ # of points belonging to $\mathcal{N}_i$ and $\mathcal{N}_j$, and is bounded by k.

$(A'A)_{ij}$ # of points which have both X_i and X_j in their K-nn.

Denote by l_j the row sum of AA' or $A'A$

$$\hat{m}(X_j) = \frac{1}{\sqrt{l_j}} \sum_{i=1}^{n} c_{ji} \frac{Y_i}{\sqrt{l_i}}.$$

This estimator could be evaluated at any point :

$$\hat{m}(x) = \frac{1}{\sqrt{l_x}} \sum_{i=1}^{n} c_{xi} \frac{Y_i}{\sqrt{l_i}}.$$

Same spirit as the new kernel estimator.

Conclusions

We propose a new kernel estimator

$$\hat{m}_W(x) = \frac{nh}{\sqrt{\hat{f}_n(x)}} \sum_{i=1}^{n} K_h(x - X_i) \frac{Y_i}{\sqrt{\hat{f}_n(X_i)}}.$$

We propose a simple procedure which transform a smoother into an admissible one and for which it is possible to evaluate at any points as for K-nn type smoother

$$\hat{m}(x) = \frac{1}{\sqrt{I_x}} \sum_{i=1}^{n} c_{xi} \frac{Y_i}{\sqrt{I_i}}.$$

It is possible to iterate such new smoother (ibr) and it is computationally interesting via the eigen decomposition.