Change-Points in Nonparametric Regression

Jose Chorikavil Thomas
Central Plantation Crops Research Institute, Regional Station
Vittal-574 243
Karnataka, India
E-mail:ctjos@yahoo.com

Ismail Beary
Department of Statistics, Mangalore University
Mangalagangothri-574 199
Karnataka, India

1. Introduction.

Local linear regression smoothers are generally used in order to obtain a smooth fit of a regression function whenever there is no suitable parametric model available. Sometimes a generally smooth function might contain some isolated discontinuity or change points in the function or its derivative. In practice, we are often interested in the location and size of change points of the regression function. McDonald and Owen (1986) used split linear fit of the data to estimate the change point in the function. Muller(1992) and Loader(1996) proposed jump detection methods based on the difference between two one-sided kernel smoothers. In this paper, estimators of location and size of jumps or discontinuities in a regression function and/or its derivatives are proposed. The estimators are based on fitting local polynomial regression with dummy variables for the jumps. The proposed method does not require that the number and order of jumps to be known in advance as do most other existing methods. The comparison of sample mean squared errors shows that the proposed method performs better than that of Loader(1996). The proposed method is applied to the data of area under arecanut in India.

2. Proposed Method

Consider the nonparametric regression model $y_i=m(x_i)+\varepsilon_i$, with design points $0\le x_1<x_2<\ldots<x_n\le 1$ and ε_i are iid random errors having mean 0 and finite variance σ^2. Under the assumption that the regression function m(x) is smooth, the locally weighted polynomial regression estimate of m(x) is $\hat{\alpha}_0$, the solution for α_0 to the following problem:

$$\text{(1)} \qquad \text{Minimize} \sum_{j=1}^{n}\left\{y_j-\alpha_0-\alpha_1(x_j-x)-\alpha_2(x_j-x)^2/2!-....\alpha_p(x_j-x)^p/p!\right\}^2 K\left(\frac{x-x_j}{h}\right)$$

where p is the order of the local polynomial, K is a kernel function and h is a bandwidth. Let there exist a jump point for the regression function at $x_t \in [h,1-h]$ with jump size Δ_j for $m^{(j)}$, the j^{th} derivative of m, j=0,1,…,p, then the above minimization function becomes

$$\text{(2)} \qquad \text{Minimize} \sum_{j=1}^{n}\left\{y_j-\sum_{i=0}^{p}\frac{\alpha_i(x_j-x)^i}{i!}-\left[\Delta_0+\sum_{i=1}^{p}\frac{\Delta_i(x_j{}^i-x_t{}^i)}{i!}\right]I_{(x_t,1)}(x_j)\right\}^2 K\left(\frac{x-x_j}{h}\right)$$

where I is the indicator function. To estimate the unknown jump point x_t and jump sizes Δ_j, j=0,1,…p, solve the weighted least squares problem (2) with $x_t=x_k$ for all $x_k\in[h,1-h]$ and let s_k be the ratio of the mean regression sum of squares due to the estimates of $\Delta=[\Delta_0\ \Delta_1 \ldots \Delta_p]$ to the mean residual sum of squares. Then the estimate of the jump point is given by

$$x_{\hat{t}} = \arg \max_{x_k \in [h, 1-h]} (s_k)$$

and the corresponding estimates of the coefficient vector $\Delta=[\Delta_0\ \Delta_1\ \ldots \Delta_p]$ be the estimates of the jump sizes. The above procedure can easily be extended to the case of more than one jump points. The jump regression function can be estimated by fitting piece-wise local linear regression in between the estimated change points.

3. Numerical Examples

The jump regression model $y=\sin(6.3x_i)+1.0I_{[0.5,\ 1]}(x_i)+\varepsilon_i$, $x_i = i/256$, $i=1,\ldots,256$ with errors ε_i from $N(0,\sigma^2)$ and $\sigma=0.40$ is used to obtain 256 observations for the simulation study. The comparison of sample mean squared errors shows that the proposed method performs better than that of Loader(1996). One set of generated data (+) along with the estimated (dotted line) and the true (solid line) regression function are shown in Figure 1(a). The proposed method is applied to the data of area under arecanut in India from the year1967 to 1998. The data of area (+) and its estimated trend function m(t) are shown in Figure 1(b). It has been observed that the slope of the simple growth rate function of area has a jump at the year 1977. The estimated simple growth rate function m'(t) is given in Figure 1(c).

(a) (b) (c)

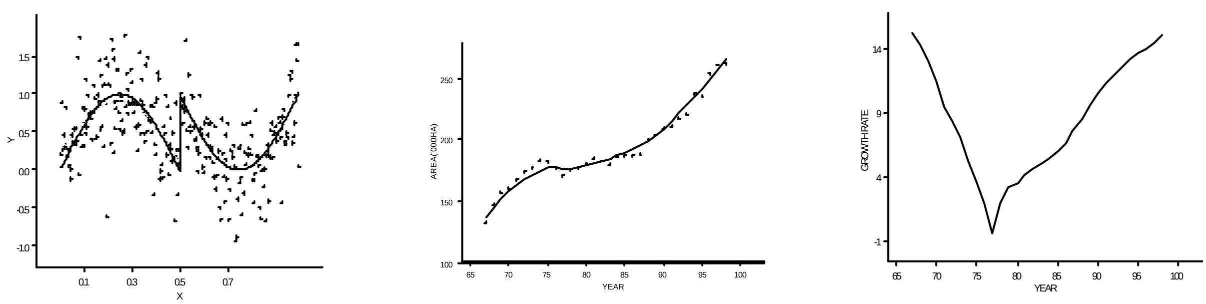

Figure 1

REFERENCES

Loader, C.R.(1996)change point estimation using nonparametric regression. Annals of Statistics 24, 1667-1678.

McDonald, J.A. and Owen, A.B.(1986). Smoothing with split linear fits. Technometrics 28, 195-208.

Muller, H.G.(1992). Change points in nonparametric regression analysis. Annals of Statistics 20, 737-761.

RESUME

Estimators of location and size of jumps or discontinuities in a regression function and/or its derivatives are proposed. The estimators are based on fitting local polynomial regression with dummy variables for the jumps. The proposed method does not require that the number and order of jumps to be known in advance and it will detect jumps of any order or simultaneous jumps in the function and/or its derivatives of any order. The performance of the proposed method is verified through simulation studies. We apply this method to the data of area under arecanut in India.