ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS
SBORNÍK MENDELOVY ZEMĚDĚLSKÉ A LESNICKÉ UNIVERZITY V BRNĚ

Ročník LIV 9 Číslo 3, 2006

NONPARAMETRIC ESTIMATES REMARKS

J. Poměnková

Received: January 10, 2006

Abstract

POMĚNKOVÁ, J.: *Nonparametric estimate remarks*. Acta univ. agric. et silvic. Mendel. Brun., 2005. Acta univ. agric. et silvic. Mendel. Brun., 2006, LIV, No. 3, pp. 93–100

Kernel smoothers belong to the most popular nonparametric functional estimates. They provide a simple way of finding structure in data. The idea of the kernel smoothing can be applied to a simple fixed design regression model. This article is focused on kernel smoothing for fixed design regresion model with three types of estimators, the Gasser-Müller estimator, the Nadaraya-Watson estimator and the local linear estimator. At the end of this article figures for ilustration of desribed estimators on simulated and real data sets are shown.

kernel, Gasser-Müller estimator, Nadaraya-Watson estimator, Local linear estimator

INTRODUCTION

Let values x be a fix design points and values Y be a simulated or real values depending on values x. Let deal with idea how to describe data structure (x, Y) using nonparametric estimates. Attention is focused on three well-known estimation types, namely Nadaraya-Watson, Local linear and Gasser - Müller estimate. First two types have been called local polynomial types. Both are very suitable for describing structures of chosen data files (simulated or real values), but their using for estimate of derivation is more complicated. From this point of view Gasser-Müller estimate, which has been called convolution type, is more suitable.

It will be shown how to construct above-mentioned estimators and also practical demonstration of discussed estimators on simulated and real data files on charts. At first on simulated functions $f(x) = 11 - (\mathrm{tg}(5x + 6))/(3\sin(12x))$ and $f(x) = \sin(\pi x)\cdot\cos(3\pi x^5)$, at the second on the June average temperatures measured in Prague during the period 1771–1890 and in Warsaw measured during the period 1755–1874.

THEORETICAL BACKGROUND

Consider a standard regression model of the form

$$Y_i = m(x_i) + \varepsilon_i,\ i = 1, ..., n,$$

where m is an unknown regression function and ε_i are independent random variables for which

$$E(\varepsilon_i) = 0,\ D(\varepsilon_i) = \sigma^2 > 0,\ i = 1, ..., n.$$

The aim of kernel smoothing is to find suitable approximation $\hat{m}$ of an unknown function m. We will assume the design points x_i are equidistantly distributed over the interval [0, 1], that is $x_i = \frac{i-1}{n}, i = 1,...,n$. In next $Lip[a, b]$ denotes the class of continuous functions satisfying the inequality $|g(x) - g(y)| \le L|x - y|$, $\forall x, y \in [a, b]$, $L > 0$ is a constant.

Definition: Let κ be a nonnegative even integer and assume $\kappa \ge 2$. The function $K \in Lip\,[-1, 1]$, *support* $(K) = [-1, 1]$, satisfying the following conditions

$$i)\quad K(-1) = K(1) = 0$$

$$ii)\quad \int_{-1}^{1} x^j K(x)dx = \begin{cases} 0, & 0 < j < \kappa, \\ 1, & j = 0, \\ \beta_k \ne 0 & j = \kappa. \end{cases}$$

is called a *kernel* of order κ and a class of all these kernels is marked $S_{0\kappa}$. These kernels are used for an estimation of the regression function (see Müller, H. J. 1988).

Kernel estimates can be generally expressed in the form $\hat{m} = \sum_{i=1}^{n} W_i(x;\, h) Y_i$, where $W_i(x, h)$ denotes weight function, $h = h_n$ is a positive constant called bandwidth, $h < x < 1 - h$, K is a kernel. Let $K \in S_{0\kappa}$, set $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$, $h \in (0, 1)$. Denote that bandwidth, kernel and order of the kernel have big influence on quality of an estimate.

METHODOLOGY

Recall to the standard regression model in the section 1 we are searching for unknown regression function m. We will use weighted least square method for an estimate of the function m. A vector $\beta = (\beta_0, ..., \beta_p)'$ such, that value of the following sum

$$S = \sum_{i=1}^{n} \left(Y_i - \underbrace{\left(\beta_0 + \beta_1 (x_1 - x) + ... + \beta_p (x_1 - x)^p\right)}_{\beta} \right)^2 K_h(x_i - x) \to \min,$$

K_h is nonnegative kernel, is searched. This vector can be found using polynomial of p-th degree in points x_i and is denoted as $\hat{\beta}$. Suitable method for finding corresponding vector $\hat{\beta}$ is weighted least square method, it's principle can be expressed by the relation $\frac{\partial S}{\partial \beta} = 0$. Assume that

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & (x_1 - x)^1 & \cdots & (x_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (x_n - x)^1 & \cdots & (x_n - x)^p \end{pmatrix},$$

$$W = \begin{pmatrix} K_h(x_1 - x) & 0 & \cdots & 0 \\ 0 & K_h(x_2 - x) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K_h(x_n - x) \end{pmatrix}.$$

Then $\frac{\partial S}{\partial \beta} = 0 \Leftrightarrow X^T W Y = X^T W X \beta$. If inverse matrix $(X^T W X)^{-1}$ exists, vector $\hat{\beta}$ takes the form

$$\hat{\beta} = (X^T W X)^{-1} \cdot X^T W Y$$

and depends on the degree p of the polynomial, values Y, x and also matrices X and W (matrix of weight functions). Resultant regression function is equivalent to the estimate $\hat{\beta} = \hat{m}(x, p, h)$.

Let's target the polynomial degree $p = 0$. Then, an estimate has been called Nadaraya-Watson estimate. Assuming that $\sum_{i=1}^{n} K_h(x_i - x) \neq 0$, inverse matrix $(X^T W X)^{-1} = \left(\sum_{i=1}^{n} K_h(x_i - x) \right)^{-1}$ exists. Thus an estimate can be expressed in the following form

$$\hat{m}_{NW}(x; 0, h) = \frac{\sum_{i=1}^{n} K_h(x_i - x) Y_i}{\sum_{i=1}^{n} K_h(x_i - x)}.$$

It holds $\hat{m}_{NW}(x; 0, h) = \sum_{i=1}^{n} V_i(x) Y_i$, where

$$V_i(x) = \frac{K_h(x_i - x)}{\sum_{i=1}^{n} K_h(x_i - x)} \text{ and } \sum_{i=1}^{n} V_i(x) = 1.$$

Let's target the polynomial degree $p = 1$. Then, an estimate has been called Local linear estimate and hold

$$X = \begin{pmatrix} 1 & (x_1 - x) \\ \vdots & \vdots \\ 1 & (x_n - x) \end{pmatrix}, W = \begin{pmatrix} K_h(x_1 - x) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_h(x_n - x) \end{pmatrix}.$$

Denote

$$\hat{s}_r(x, h) = \frac{1}{n} \sum_{i=1}^{n} (x_i - x)^r K_h(x_i - x),$$

$$X^T W X = \begin{pmatrix} \sum_{i=1}^{n} K_h(x_i - x) & \sum_{i=1}^{n} (x_i - x) K_h(x_i - x) \\ \sum_{i=1}^{n} (x_i - x) K_h(x_i - x) & \sum_{i=1}^{n} (x_i - x)^2 K_h(x_i - x) \end{pmatrix}.$$

If $x_j \neq x_i$ for $j \neq i$, K_h nonnegative kernel, then $X^T W X$ is positively definite and following relations hold

$$X^T W X = \begin{pmatrix} n\hat{s}_0(x,h) & n\hat{s}_1(x,h) \\ n\hat{s}_1(x,h) & n\hat{s}_2(x,h) \end{pmatrix} = n \begin{pmatrix} \hat{s}_0(x,h) & \hat{s}_1(x,h) \\ \hat{s}_1(x,h) & \hat{s}_2(x,h) \end{pmatrix},$$

$$\det\left(X^T W X\right) = \hat{s}_0(x,h)\hat{s}_2(x,h) - n\hat{s}_1(x,h)^2 =$$

$$= \sum K_h(x_j - x) K_h(x_i - x)(x_j - x)^2 > 0.$$

Corresponding formula for the Local linear estimate is

$$\hat{m}_{LL}(x;1,h)=\underbrace{\frac{1}{n}\frac{1}{\hat{s}_0(x,h)\hat{s}_2(x,h)-\hat{s}_1(x,h)^2}\begin{pmatrix}\hat{s}_0(x,h) & -\hat{s}_1(x,h)\\ -\hat{s}_1(x,h) & \hat{s}_2(x,h)\end{pmatrix}}_{(X^TWX)^{-1}}X^TWX=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\frac{\left[\hat{s}_2(x,h)-\hat{s}_1(x,h)\right]K_h(x_i-x)}{\hat{s}_0(x,h)\hat{s}_2(x,h)-\hat{s}_1(x,h)^2}Y_i.$$

A different type of nonparametric estimate of the function m is the Gasser-Müller estimate defined as

$$\hat{m}_{GM}(x)=\sum_{i=1}^{n}Y_i\frac{1}{h}\int_{s_{i-1}}^{s_i}K\left(\frac{x-u}{h}\right)du,$$

where x_i, $i = 1, ..., n$, $x_i \in [0, 1]$ are points of plan ordered according to their size, $s_0 = 0$, $s_i=\frac{x_{i+1}+x_i}{2}, i=1,..n-1$ a $s_n = 1$. For the members of the smoothing matrix $s_{ij} = V_i(x_j, h)$, $i, j = 1, ..., n$ in point of plan x_j with bandwidth h for the Gasser-Müller estimate holds

$$V_i(x_j,h)=\frac{1}{h}\int_{s_{i-1}}^{s_i}K\left(\frac{x_j-u}{h}\right)du.$$

The quality of an estimate has been influenced by the value of bandwidth. If the bandwidth is small, the estimate is undersmoothed. If the bandwidth is large, the estimate is oversmoothed. One of the suitable instrument for an evaluation of the estimation quality is average mean square error $AMSE(\hat{m}(x), h)$.

The kernel estimation can be viewed as linear smoother. Let's have a smoothing matrix $S = (s_{ij})$, $s_{ij} = V_i(x_j, h)$, $i, j = 1, ..., n$ defined according to the type of an estimate. Then the estimate $\hat{m}$ can be expressed by the following formula $\hat{m} = SY$. For these estimators, the condition $\sum_{i=1}^{n} s_{ij} = 1$, $j = 1, ..., n$ holds. Denote the vector of estimators by $\hat{m} = (\hat{m}(x_1), ..., \hat{m}(x_n))^T$ and the vector of bias by $b = m - Sm$, $b = (b_1, ..., b_n)$. Then for the average mean square error $AMSE$ the following formula holds

$$AMSE(\hat{m},h)=\frac{1}{n}\sum_{i=1}^{n}\operatorname{var}(\hat{m}(x_i))+\frac{1}{n}\sum_{i=1}^{n}b_i^2=\frac{tr(SS^T)}{n}\sigma^2+\frac{b^Tb}{n}.$$

If we want to find the optimum value of the bandwidth $h_{opt,\,AMSE}$ using $AMSE(\hat{m}, h)$, we enumerate the value of this function using the formula above for $h \in H_n = [1/n; 2]$, n is a measuring range, and we search for a minimum of this function. For an optimum value $h_{opt,\,AMSE}$ the formula $h_{opt,\,AMSE} = \arg\min_{h\in H_n} AMSE(\hat{m}, h)$ holds (see Poměnková, 2005).

These results are only theoretical since formula defined above contains an unknown function m. One possible solution is using „Cross-validation method".

Cross-validation works by leaving points (x_i, Y_i) out one at a time and estimating the smooth at x_i based on the remaining $n - 1$ points. One the constructs cross-validation sum of squares

$$CV(h)=\frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\hat{m}_{-i}(x_i)\right)^2,$$

where $\hat{m}_{-i}(x_i)$ indicates the fit at x_i, computed by leaving out the ith data point. The optimum value of the bandwidth $h_{opt,\,CV}$ using $CV(h)$ is obtained on the set $H_n = [a_k n^{-1/(2k+1)}, b_k n^{-1/(2k+1)}]$ for some $0<a_k<b_k<\infty$ by minimizing cross validation function, i.e. $h_{opt,\,CV} = \arg\min_{h\in H_n} CV(\hat{m}, h)$.

This formula can be simplify to the form

$$GCV(h)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Y_i-\hat{m}(x_i)}{1-tr(V)/n}\right\}^2,$$

$$tr(V)=trace(V)=\sum_{i=1}^{n}V_i(x,h),$$

where W is weight matrix define above according to the type of an estimate (Nadaraya-Watson, Local linear and Gasser-Müller estimate) (see Poměnková, 2005). This method has been called generalized cross – validation method.

APPLICATION

The algorithmes described in the preceding part were used for finding an estimate of chosen data structures. The first two data sets consist of 120 simulated values of functions

$$f(x)=11-\frac{tg(5x+6)}{3\sin(12x)} \text{ and}$$

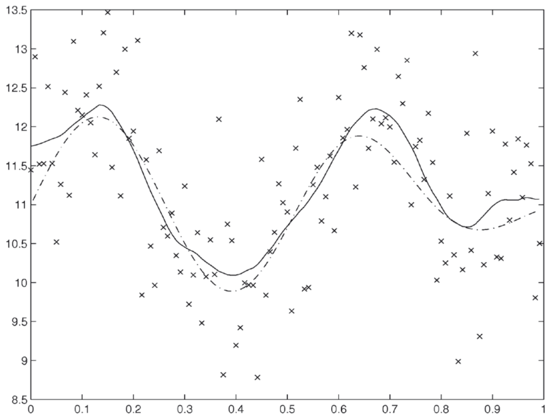
$$f(x)=\sin(\pi x)\cdot\cos(3\pi x^5),$$

$\sigma^2 = 0.7$ was used. The Epanechnikov kernel

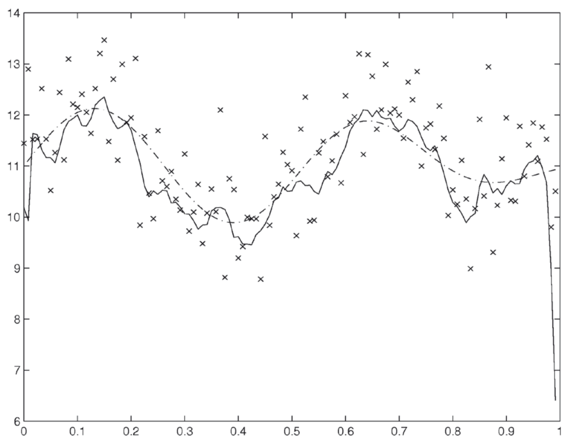
$$K(x)=\frac{3}{4}(1-x^2),\quad K\in S_{02}.$$

Value of the bandwidth h was found by using cross-validation method for all estimates.

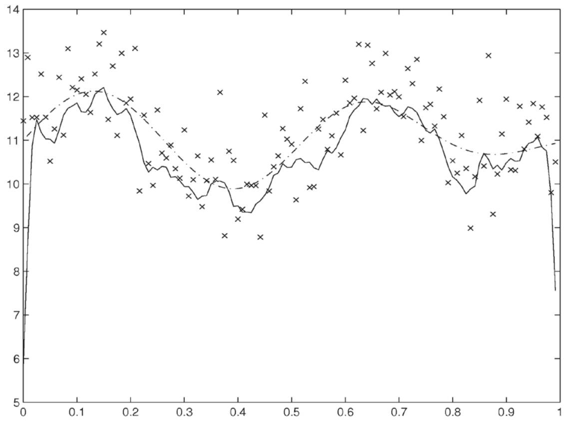
$$f(x) = 11 - \frac{tg(5x+6)}{3\sin(12x)}$$



1a.: *A Nadaraya-Watson estimation of the function f(x),* h_{NW}*=0.085 (dash-dotted line is the function f(x), solid line is estimation of the function f(x)).*

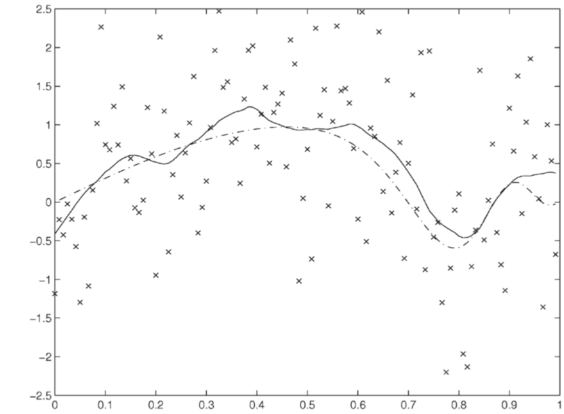


1b.: *A Local linear estimation of the function f(x),* h_{LL}*=0.026 (dash-dotted line is the function f(x), solid line is estimation of the function f(x)).*

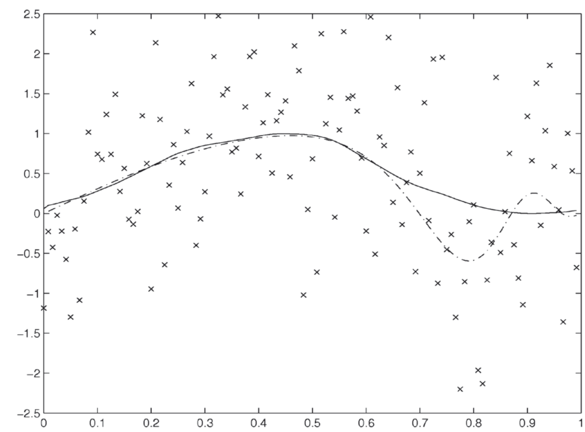


1c.: *A Gasser-Müller estimation of the function f(x),* h_{GM}*=0.026 (dash-dotted line is the function f(x), solid line is estimation of the function f(x)).*

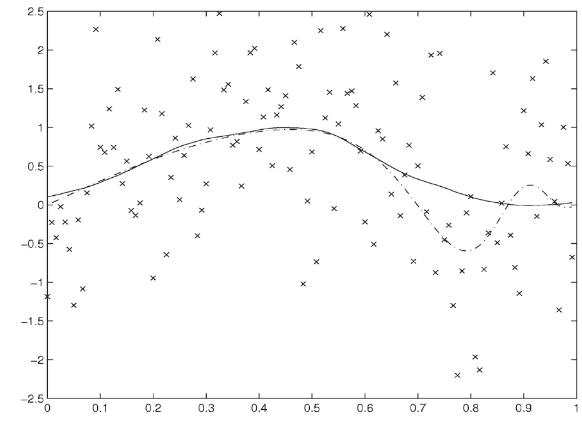
$$f(x) = \sin(\pi x) \cdot \cos(3\pi x^5)$$



2a.: *A Nadaraya-Watson estimation of the function f(x),* h_{NW}*=0.1 (dash-dotted line is the function f(x), solid line is estimation of the function f(x)).*

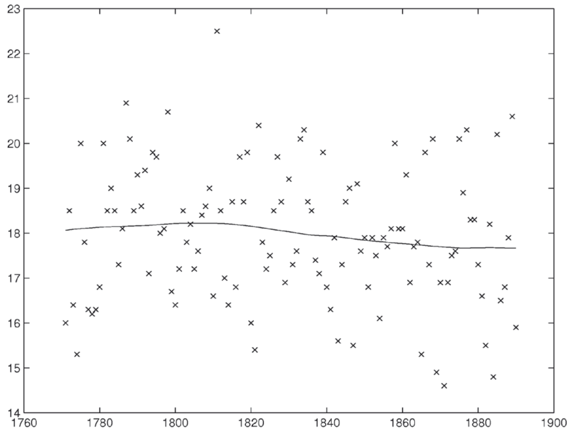


2b.: *A Local linear estimation of the function f(x),* h_{LL}*=0.255 (dash-dotted line is the function f(x), solid line is estimation of the function f(x)).*

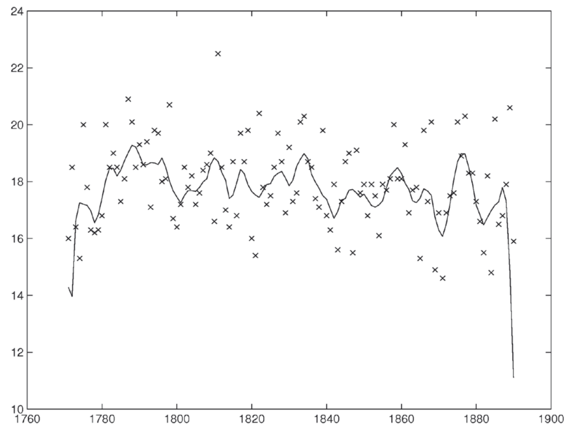


2c.: *A Gasser-Müller estimation of the function f(x),* h_{GM}*=0.255 (dash-dotted line is the function f(x), solid line is estimation of the function f(x)).*

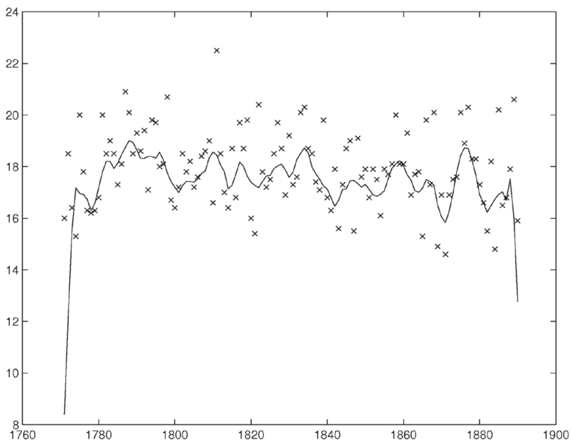
Prague (1771–1890)



3a.: *A Nadaraya-Watson estimation of the average June temperatures measured in Prague (1771–1890), h_{NW}=0.44.*

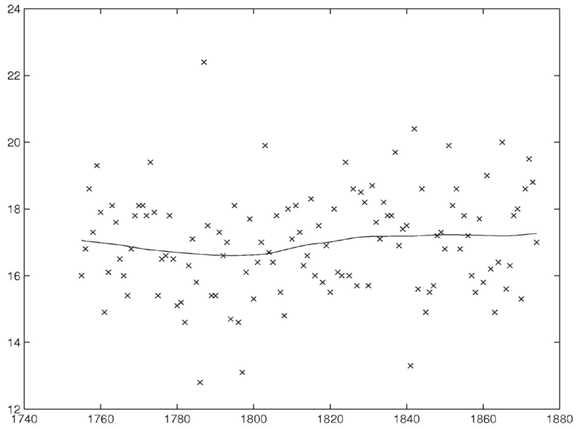


3b.: *A Local linear estimation of the average June temperatures measured in Prague (1771–1890), h_{LL}=0.027.*

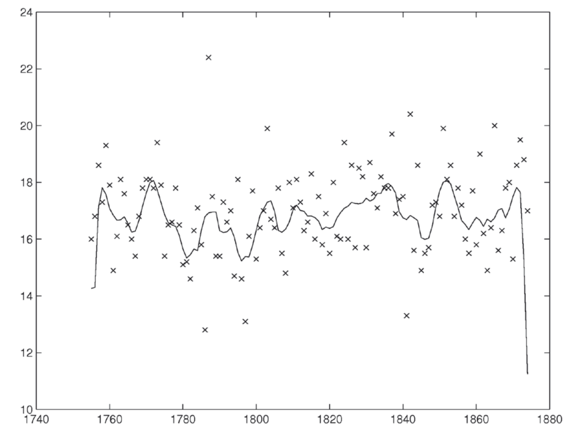


3c.: *A Gasser-Müller estimation of the average June temperatures measured in Prague (1771–1890), h_{GM}=0.036.*

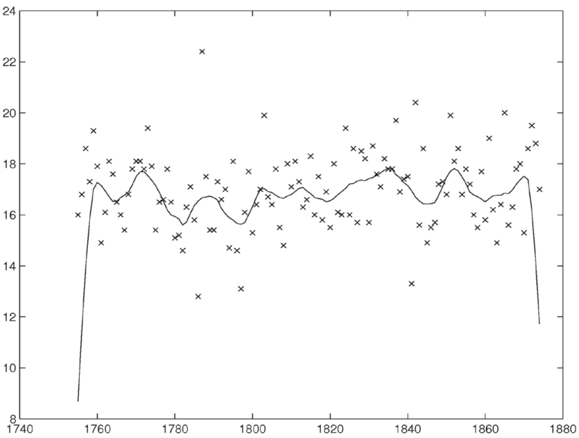
Warsaw (1755–1874)



4a.: *A Nadaraya-Watson estimation of the average June temperatures measured in Warsaw (1755–1874), h_{NW}=0.29.*



4b.: *A Local linear estimation of the average June temperatures measured in Warsaw (1755–1874), h_{LL}=0.026.*



4c.: *A Gasser-Müller estimation of the average June temperatures measured in Warsaw (1755–1874), h_{GM}=0.06.*

The other two data sets consist of 120 values of June average temperatures measured in Prague during the period 1771–1890 and in Warsaw during the period 1755–1874. Denote, that values t (time period 1771–1890 for Prague, 1755–1874 for Warsaw) is for calculation transformed to values $x \in [0; 1]$ according to the assumption at the start of the "Theoretical background". The Epanechnikov kernel

$$K(x) = \frac{3}{4}\left(1 - x^2\right), \quad K \in S_{02},$$

was used. Value of the bandwidth h was found by using cross-validation method for all estimates.

RESULTS

Presented paper demonstrate possibilities how to construct an estimate of the function using nonparametric estimation, especially the Nadaraya-Watson, the Local linear and the Gasser-Müller estimate. An application of mentioned type estimation is shown at two types of the data file, simulated and real.

At first demonstration on simulated data is shown. From graphical presentation of resultant estimate can be seen that in case of the function

$$f(x) = 11 - (\mathrm{tg}(5x + 6))/(3\sin(12x)),$$

the Nadaraya-Watson estimate appears to be more suitable (Fig. 1a). Remaining figures of estimates, for the Local linear (Fig. 1b) and for the Gasser – Müller (Fig. 1c)), have boundary effects, which should be suitable to remove using some special method (reflection method, cyclic method) (see Poměnková, 2004b; Koláček, 2005). And also they look undersmoothed. In case of the function $f(x) = \sin(\pi x)\cdot\cos(3\pi x^5)$, the Nadaraya-Watson estimate (Fig. 2a) has a small boundary effect. Figure for the Local linear estimate (Fig. 2b) similar to the Gasser-Müller (Fig. 2c) estimator gives good result in part of the interval [0;1], but from global point of view attend to a right boundary effect, i.e. near to the point 1, and to oversmoothing of an estimate.

In the case of real data June average temperatures measured in Prague during the period 1771–1890 and in Warsaw during the period 1755–1874 has been procesed. In both cases for the Local linear (Fig. 3b, Fig. 4b) and the Gasser-Müller estimate (Fig. 4b, Fig. 4c) we can state, that estimates have a boundary effect. From this point of view the Nadaraya-Watson estimate (Fig. 3a, Fig. 4a) for both data files appears to be more suitable.

As mentioned above a bandwidth and a kernel (the order of the kernel) have big influence on quality of an estimate. It could not be excluded that using different type of kernel function will bring better results (see Poměnková, 2004b; Poměnková, 2005; Koláček, 2005).

SUMMARY

Kernel estimations as a part of nonparametric functional estimates provide a simple tool for finding a structure in data. Special types of polynomial functions – kernels are used for an estimation of the resultant regression function. The aim of this paper is to present construction of three types of estimates – namely the Nadaraya – the Watson estimate, the Local linear estimate and the Gasser – Müller estimate. Methodology and description of discussed estimators is illustrated at the attached practical part – application on two types of data files. At first on simulated functions $f(x) = 11 - (\mathrm{tg}(5x + 6))/(3\sin(12x)$ and $f(x) = \sin(\pi x)\cdot\cos(3\pi x^5)$, at the second on the June average temperatures measured in Prague during the period 1771–1890 and in Warsaw measured during the period 1755–1874.

SOUHRN

Poznámky k neparametrickým odhadům

Jedny z nejjednodušších vyhlazovacích technik jsou jádrové odhady. Tyto odhady mají jednoduchou strukturu a asymptotický charakter. Mezi základní typy jádrových odhadů patří Nadaraya-Watsonův odhad, lokálně lineární odhad a Gasser-Müllerův odhad.

Nechť jsou hodnoty x pevně voleny experimentátorem a hodnoty Y, které mohou být reálné nebo simulované, jsou závislé na hodnotách x. Zabývejme se otázkou využití regresních křivek pro popis vztahu dvojice proměnných $(x_i, Y_i)^n_{i=1}$, $i = 1, ..., n$. Tento regresní vztah může být popsán rovnicí ve tvaru $Y_i = m(x) + \varepsilon_i$, $i = 1, ..., n$ s neznámou regresní funkcí m a chybou pozorování ε_i. Křivková aproximace neznámé funkce m je obvykle nazývána vyhlazováním.

Předkládaný příspěvek se zabývá využitím regresních křivek pro popis vztahu dvojice proměnných $(x_i, Y_i)^n_{i=1}$, $i = 1, ..., n$, přičemž pozornost je věnována neparametrickým odhadům s využitím tří speciálních typů odhadů, a to Nadaraya-Watsonova odhadu, lokálně lineárního odhadu a Gasser-Müllerova odhadu. Na závěr jsou uvedeny aplikace získaných a zpracovaných poznatků, a to na simulovaných i reálných datech.

jádro, Gasser-Müllerův odhad, Nadaraya-Watsonův odhad, lokální lineární odhad

ACKNOWLEDGMENTS

This work is the result of research project "The Czech Economy in the Process of Integration and Globalisation, and the Development of Agricultural Sector and the Sector of Service under the New Conditions of the Integrated European Market".

REFERENCES

GASSER, T., ENGEL, J.: The choice of weights in kernel regression estimation. Biometrika, 77, pp. 377–381, 1990.

HÄRDLE, W.: Applied Nonparametric Regression. Cambridge University Press, 349 p, 1990. ISBN 10: 0521429501

HOROVÁ, I., ZELINKA, J.: Základy a aplikace jádrových odhadů. Analýza dat 2000/II. Moderní statistické metody,Lázně Bohdaneč 21.–24. 11. 2000, pp. 141–167.

KOLÁČEK J.: Kernel estimation of the regression function (1st ed.). Brno: PhD-thesis, 2005.

MŰLLER, H. G.: Nonparametric Regression analysis of Longitudinal data.Lecture Notes in Statistics 46. Springer–Verlag, 1988.

POMĚNKOVÁ, J.: Gasser-Müller estimate. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2004a, vol. 3 LII, 3: pp. 167–176. ISSN 1211-8516.

POMĚNKOVÁ, J.: Edge effects of the Gasser-Muller estimator, Summer School Datastat03, Proceedings, Folia Fac. Sci. Nat. Univ. Masaryk. Brunensis, Mathematica 15, 2004b, pp. 307–314. ISBN 80-210-3564-1.

POMĚNKOVÁ, J.: Some aspects of regression function smoohting (1st ed.). Ostrava: PhD-thesis, 2005.

WAND, M. P., JONES, M. S.: Kernel Smoothing (1st ed.). Chapman & Hall, London, 1995.

Address

RNDr. Jitka Poměnková, Ph.D., Ústav statistiky a operačního výzkumu, Mendelova zemědělská a lesnická univerzita, Brno, Zemědělská 5, 613 00 Brno, Česká republika, e-mail: pomenka@mendelu.cz