

*Aleksandra Baszczyńska**

SOME NONPARAMETRIC ESTIMATORS OF REGRESSION FUNCTION

Abstract. In the paper some nonparametric estimators of regression function are studied: Nadaraya–Watson estimator and k -nearest neighbour one. Properties of these estimators and possibilities of using them in practice are taken into consideration. A comparative study of the two estimators is presented. Different techniques of choosing method's parameters (kernel function, smoothing parameter h and parameter k) are used in this study to choose the optimal ones. Some practical rules are proposed and they are used in this study.

Key words: regression function, kernel function, smoothing parameter, k -nearest neighbour method, kernel method.

I. INTRODUCTION

A regression curve describes the relationship between a predictor variable X and a response variable Y . Knowledge of this relation is one of the basic problems in statistical practice. Quite often the regression curve itself is not the focus of interest, but its monotonicity, unimodality, location of zeros or the derivatives of it.

For n independent observations $\{(X_i, Y_i)\}_{i=1}^n$ regression relationship can be modelled as

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

* Ph.D., Chair of Statistical Methods, University of Łódź.

where:

$$m(x) = E[Y/X = x] = \frac{\int yf(x, y)dy}{f(x)}$$

is regression function, $f(x, y)$ is the joint density of X and Y , $f(x)$ is the marginal density of X ,

ε_i are independent observation errors with the same distribution with $E\varepsilon_i = 0$ and $D^2\varepsilon_i = \sigma^2 < \infty$.

In many cases a scatter plot of X_i versus Y_i is not sufficient to establish the regression relationship. Then the nonparametric estimation of regression function m can be used to approximate the mean response curve m . The term "nonparametric" refers to the situation when the mean curve does not have some prespecified functional form.

Nonparametric regression estimator has the following general form (compare: *Hardle 1991; Wand, Jones 1995*):

$$\hat{m}(x) = n^{-1} \sum_{i=1}^n W_{ni}(x) Y_i \quad (2)$$

where $\{W_{ni}(x)\}_{i=1}^n$ denotes a sequence of weights.

Every nonparametric regression estimator can be regarded as weighting averages of the response variables Y_i and the weights $\{W_{ni}(x)\}_{i=1}^n$, which may depend on the whole vector $\{X_i\}_{i=1}^n$ and control the amount of averaging. In the paper we present two common choices for the weights $\{W_{ni}(x)\}_{i=1}^n$, what leads to two nonparametric regression estimators: kernel and k -nearest neighbour ones.

2. NADARAYA-WATSON KERNEL ESTIMATOR

The sequence of weights in Nadaraya-Watson regression estimator is the following:

$$W_{ni}(x) = \frac{h^{-1} K\left(\frac{x - X_i}{h}\right)}{n^{-1} \sum_{i=1}^n h^{-1} K\left(\frac{x - X_i}{h}\right)} \quad (3)$$

Then the estimator has the form (compare: *Pagan, Ullah 1999; Huang, Brill 2001*):

$$\hat{m}(x) = \frac{n^{-1} \sum_{i=1}^n h^{-1} K\left(\frac{x - X_i}{h}\right) Y_i}{n^{-1} \sum_{i=1}^n h^{-1} K\left(\frac{x - X_i}{h}\right)} \quad (4)$$

where K is the kernel function, h is the bandwidth (smoothing parameter).

Kernel function is a symmetric function satisfying:

$$\int_{-\infty}^{+\infty} K(u) du = 1, \quad \int_{-\infty}^{+\infty} u K(u) du = 0, \quad \int_{-\infty}^{+\infty} u^2 K(u) du = k_2 \neq 0.$$

A variety of kernel functions is possible in practice. Some of them used most often are presented in Cz. Domański, K. Pruska, W. Wagner (1998) (compare: Rosenblatt 1956; Baszczyńska 2005, 2006). The property of smoothness of the kernel is inherited by the corresponding estimator (4), so the proper choice of kernel function is the basic problem in the procedure of estimation of regression function.

The smoothing parameter $h(n) > 0$ satisfies

$$\lim_{n \rightarrow \infty} h(n) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} h(n) = \infty.$$

We can note that increasing the bandwidth implies increasing the amount of smoothing in the estimation, and decreasing the bandwidth leads to a less smooth estimate.

3. k -NEAREST NEIGHBOUR ESTIMATOR

The weights in k -nearest neighbor (k -nn) estimator introduced by D. Loftsgaarden and C. Quesenberry (1965) in the field of density estimation are defined as:

$$W_{ki}(x) = \begin{cases} \frac{n}{k}, & \text{if } i \in J_x \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where $J_x = \{i | X_i \text{ is one of the } k \text{ nearest observations to } x\}$. It means that J_x is the set of indices of the k -nearest neighbours of x .

In the regarded estimators, the parameter k regulates the amount of smoothing. k -nn estimator is a weighted average in a varying neighbourhood, since the kernel estimator is a weighted average of the response variables in a varying neighbourhood. This neighbourhood is defined through those X -variables which are among the k -nearest neighbours of a point x .

The bias and the variance of the two studied estimators of the regression function are the following:

	Kernel estimator	k -nn estimator
Bias	$h^2 \frac{(m''f + 2m'f')(x)}{2f(x)} \int u^2 K(u) du$	$\left(\frac{k}{n}\right)^2 \frac{(m''f + 2m'f')(x)}{8f^3(x)} \int u^2 K(u) du$
Variance	$\frac{\sigma^2(x)}{nhf(x)} \int K^2(u) du$	$\frac{\sigma^2(x)}{k} \int K^2(u) du$

It can be noticed that the variance of the k -nn regression estimator does not depend on the density of X (k -nn estimator averages over exactly k observations independently of the distribution of the X variables). For $k = 2nhf(x)$ (or equivalently $h = \frac{k}{2nf(x)}$), the asymptotic mean squared error (sum of variance and the squared of the bias) at x is the same for kernel and k -nearest neighbour estimators.

4. THE STUDY

The study was conducted to indicate how parameters of methods of estimation influence on fitting the estimators to the true regression curve.

In the study n ($n = 10, 30, 100, 300$) points $\{(X_i, Y_i)\}$ were considered, where $Y_i = \sin^3(2\pi X_i^3) + \varepsilon_i$, X_i have uniform distribution on the interval $[0, 1]$, and ε_i have $N(0, 0.02)$, $N(0, 0.1)$ or $N(0, 0.7)$.

Different values of n and different parameters of normal distribution of error can be treated as variants of study. In this study true regression function is known, so it is possible to evaluate how far from this true function the estimator is constructed. In order to find the optimal parameters of estimation methods the measure of fitting BSK as follows:

$$BSK = \frac{1}{n} \sum_{i=1}^n [m(x_i) - \hat{m}(x_i)]^2$$

should be the minimum.

For the kernel estimator seven kernel functions were used. For all these estimators smoothing parameter h minimised BSK is regarded as optimal. In that way it is possible to compare the values of optimal parameters among estimators with different kernel functions. The results for $n = 100$ are presented in Tab. 1.

Table 1

Values of smoothing parameters h for nonparametric kernel estimators minimizing BSK

Kernel function	$\varepsilon_i \sim N(0, 0.02)$		$\varepsilon_i \sim N(0, 0.1)$		$\varepsilon_i \sim N(0, 0.7)$	
	parameter h	BSK	parameter h	BSK	parameter h	BSK
Epanechnikov	0.013	0.004 214	0.028	0.008 955	0.042	0.020 204
Gaussian	0.016	0.003 853	0.027	0.008 363	0.05	0.019 343
Quartic	0.04	0.004 091	0.07	0.008 606	0.12	0.020 007
Triangle	0.04	0.004 115	0.07	0.008 269	0.11	0.020 161
Uniform	0.02	0.004 207	0.06	0.010 811	0.08	0.023 072
Triweight	0.05	0.004 043	0.08	0.008 482	0.14	0.019 841
Cosinus	0.03	0.004 219	0.06	0.008 896	0.10	0.020 332

Source: author's calculations.

It is easy to notice that estimators with two of regarded kernels Gaussian and Epanechnikov need smaller parameters h . It means that these kernels have smoothing properties inside and that is why the estimator h does not need such a big smoothing parameter as in the case of other kernels. This property can also be seen in estimation of density function (compare: Baszczyńska, 2006).

The results for all variants of sample size, but only for Gaussian and uniform kernel are presented in Tab. 2.

For the two studied kernel functions (uniform and Gaussian) measure of fitting BSK is the smallest when variance is equal 0.02. The bigger variance, the bigger BSK . For small variance of ε_i , the smoothing parameter is small.

The same variants of the study were investigated in the estimation using k -nn estimators of regression function. The optimal parameter k is found in the same way as in the case of kernel estimator – it is such a parameter for which measures of fitting (BSK) are the smallest.

Table 2

Values of optimal parameter h in kernel estimator (values of BSK)

Sample size	Variance ε_i	Uniform kernel		Gaussian kernel	
		parameter h (BSK)		parameter h (BSK)	
$n = 10$	0.02	0.069	(0.002 749)	0.035	(0.002 744)
	0.1	0.069	(0.012 880)	0.043	(0.012 585)
	0.7	0.069	(0.088 854)	0.100	(0.080 664)
$n = 30$	0.02	0.017	(0.008 104)	0.012	(0.008 191)
	0.1	0.070	(0.028 931)	0.063	(0.012 624)
	0.7	0.119	(0.051 142)	0.082	(0.257 722)
$n = 100$	0.02	0.020	(0.004 207)	0.016	(0.003 853)
	0.1	0.060	(0.010 811)	0.027	(0.008 363)
	0.7	0.074	(0.021 719)	0.050	(0.019 343)
$n = 300$	0.02	0.028	(0.002 318)	0.017	(0.002 172)
	0.1	0.043	(0.005 326)	0.024	(0.005 246)
	0.7	0.051	(0.019 605)	0.034	(0.016 618)

Source: author's calculations.

In addition, three practical rules of choosing smoothing parameter were proposed. These practical rules are modified for estimation of regression, in comparison with the practical rules used in density estimation (compare: Baszczyńska 2006) using uniform kernel and $k = 2nh$. They are as follows:

Practical rule I

$k_1 = [2.2n^{4/5}\sigma]$, where σ is estimated on the base of the sample, and $[z]$ denotes the largest integer less than or equal to z .

Practical rule II

$k_2 = [1.58Rn^{4/5}]$, where R is interquartile range.

Practical rule III

$k_3 = [1.8An^{4/5}]$, where $A = \min\left(\sigma, \frac{R}{1.34}\right)$.

The results, including optimal values of parameter k for all regarded variants of the study and values computed on the base of the practical rules with minimum value of measure of fitting BSK (in brackets), are presented in Tab. 3.

Table 3

Values of optimal parameter k of k -nearest neighbour estimation and parameters k_1, k_2, k_3 from practical rules with values of BSK

Sample size	Variance ε_i	Optimal parameter k (BSK)	Parameter k_1 (BSK)	Parameter k_2 (BSK)	Parameter k_3 (BSK)
$n = 10$	0.02	1 (0.007 009)	5 (0.066 824)	4 (0.058 288)	3 (0.062 566)
	0.1	2 (0.022 906)	5 (0.068 509)	6 (0.076 365)	5 (0.068 509)
	0.7	4 (0.074 299)	8 (0.117 450)	10 (0.105 743)	7 (0.116 076)
$n = 30$	0.02	3 (0.011 472)	16 (0.159 864)	9 (0.052 404)	8 (0.037 279)
	0.1	5 (0.019 503)	19 (0.197 889)	15 (0.166 068)	13 (0.130 075)
	0.7	10 (0.047 964)	30 (0.186 386)	30 (0.186 386)	27 (0.186 307)
$n = 100$	0.02	4 (0.004 158)	42 (0.100 920)	30 (0.038 486)	26 (0.029 761)
	0.1	7 (0.010 976)	50 (0.153 276)	42 (0.098 227)	36 (0.062 088)
	0.7	13 (0.034 071)	88 (0.215 922)	83 (0.232 837)	70 (0.232 585)
$n = 300$	0.02	13 (0.002 353)	100 (0.073 650)	74 (0.039 429)	63 (0.031 881)
	0.1	15 (0.007 396)	115 (0.104 589)	108 (0.009 0023)	92 (0.062 136)
	0.7	44 (0.024 464)	195 (0.233 358)	189 (0.232 594)	161 (0.192 285)

Source: author's calculations.

Optimal parameter minimizing BSK is bigger for large value of sample size. It means that for large n the estimator needs more observation to get the minimum measure of fitting. This measure is small for all cases where variance of ε_i is small. The conclusion based on comparing optimal parameter with parameter from practical rules is as follows: values of parameter k computed from practical rules differ from the optimal one. It means that, when practical rules are used widely in estimation of density function, there are some contraindications for using them in estimation of regression function.

REFERENCES

- Baszczyńska A. (2005), *Some Remarks on the Choice of the Kernel Function in Density Estimation*, "Acta Universitatis Lodzianis", Folia Oeconomica, **194**, 143–149.
- Baszczyńska A. (2006), *Choice of the Smoothing Parameter in Kernel Density Estimation*, "Acta Universitatis Lodzianis", Folia Oeconomica, **196**, 57–63.
- Domański Cz., Pruska K., Wagner W. (1998), *Wnioskowanie statystyczne przy nieklasycznych założeniach*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Hardle W. (1991), *Smoothing Techniques. With Implementation in S*, Springer-Verlag, New York.
- Huang M., Brill P. (2001), *A Nonparametric Regression Method*, "Nonlinear Analysis", **47**, 1467–1475.
- Loftsgaarden D., Quesenberry C. (1965), *A Nonparametric Estimate of a Multivariate Density Function*, Ann. Math. Statist. **36**, 1049–1051.
- Pagan A., Ullah A. (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
- Rosenblatt M. (1956), *Remarks on Some Nonparametric Estimation of a Density Function*, Ann. Math. Statist. **27**, 832–837.
- Wand M., Jones M. (1995), *Kernel Smoothing*, Chapman and Hall, London.

Aleksandra Baszczyńska

WYBRANE NIEPARAMETRYCZNE ESTYMATORY FUNKCJI REGRESJI

W pracy przedstawiono wybrane dwa nieparametryczne estymatory funkcji regresji: estymator jądrowy Nadaraya–Watsona oraz estymator k -najbliższego sąsiada. Podano ich własności, możliwości wykorzystania oraz dokonano porównania tych estymatorów. Przedstawiono również przykład zastosowania estymatora jądrowego regresji z uwzględnieniem właściwego doboru parametrów metody (funkcji jądra i parametru wygładzania h) oraz estymatora k -najbliższego sąsiada z uwzględnieniem właściwego doboru parametru k . Zaproponowano również praktyczne zasady wyboru parametrów estymacji funkcji regresji i wykorzystano je w przykładzie.