ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 194, 2005

*Aleksandra Baszczyńska**

SOME REMARKS ON THE CHOICE OF THE KERNEL FUNCTION IN DENSITY ESTIMATION

Abstract

The basic characteristic describing the behaviour of the random variable is its density function. Kernel density estimation is one of the most widely used nonparametric density estimations. In the process of constructing the estimator we have to choose two parameters of the method: the kernel function $K(u)$ and smoothing parameter h (bandwidth). In the paper, kernel method is discussed in detail, with particular emphasis on influence of the choice of the kernel function $K(u)$ on the quantity of smooothing. Monte Carlo study is presented, where seven kernel functions (Gaussian, Uniform, Triangle, Epanechnikov, Quartic, Triweight, Cosinus) are used in density estimation.

Key words: density estimation, kernel function, smoothing parameter.

I. INTRODUCTION

Density function is the basic characteristic describing the behaviour of the random variable. It is used in investigation of properties of a given set of data and provides a way of showing its structure.

The oldest density estimator for univariate case is the histogram. Popularity of the histogram is connected with its simplicity, but this estimator has some drawbacks (for example: influence of the placement of the bin edges on the estimator, estimating all densities by a step function). One of the most known and widely used methods of estimation of density function is the kernel method. The motivation for kernel density estimation is the average shifted histogram, which averages several histograms based on shifts of the bin edges. Kernel estimator does not have disadvantages of the histogram and provides simple and effective method of showing structure in a data set at the beginning of analysis.

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

Kernel estimator of a density function $f(x)$ is defined by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

where $K(u)$ is kernel function satisfying $\int_{-\infty}^{+\infty} K(u)du = 1$.

The idea of kernel estimator was introduced by Fix and Hodges in 1951 as nonparametric version of discriminant analysis. Rosenblatt (1956) considered kernel estimator with one special kernel function, and Parzen in 1962 introduced a general form of kernel estimator.

In practice, the kernel function $K(u)$ is a density function (for example normal function) and then estimator (1) is also density function. Some of the best known kernel functions are presented in Domański, Pruska, Wagner (1998).

Parameter h ($h > 0$) is a smoothing parameter, also called window width or bandwidth.

Expressions for $E(\hat{f}(x))$ and $D^2(\hat{f}(x))$ are the following:

$$E(\hat{f}(x)) = \int_{-\infty}^{+\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y)dy, \tag{2}$$

$$D^2(\hat{f}(x)) = \frac{1}{n}\left\{\int_{-\infty}^{+\infty} \frac{1}{h^2} K\left(\frac{x-y}{h}\right)^2 f(y)dy - \left[\int_{-\infty}^{+\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y)dy\right]^2\right\}. \tag{3}$$

Let the kernel function be a symmetric function satisfying:

$$\int_{-\infty}^{+\infty} K(u)du = 1$$

$$\int_{-\infty}^{+\infty} uK(u)du = 0$$

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

$$\text{for} \quad h(n) > 0$$

$$\lim_{n\to\infty} h(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} nh(n) = \infty,$$

the unknown density $f(x)$ has continuous derivatives of all orders required.

The bias and asymptotic mean integrated squared error of kernel estimator (1) is the following:

$$E(\hat{f}(x)) - f(x) = \int_{-\infty}^{+\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y)dy - f(x). \quad (4)$$

$$AMISE = \frac{1}{4} h^4 k_2^2 \int_{-\infty}^{+\infty} f''(x)^2 dx + \frac{1}{nh} \int_{-\infty}^{+\infty} K(u)^2 du. \quad (5)$$

II. MONTE CARLO STUDY

Monte Carlo study was conducted to indicate the influence of the choice of the kernel function on the quantity of smoothing in kernel density function. Analysis of properties of estimator was done in three basic variants, depending on distribution, from which the data were chosen. These variants are as follows:

– variant I: normal distribution N(0,0.2),

– variant II: mixture of normal distributions: $f(x) = 0.25f_1(x) + 0.75f_2(x)$, where $f_1(x)$ is density function N(0, 0.2), $f_2(x)$ is density function N(3, 0.5). Variant II presents two-modal distributions,

– variant III: mixture of normal distributions: $f(x) = 0.5f_1(x) + 0.25f_2(x) + 0.25f_3(x)$, where $f_1(x)$ is density function N(0, 0.2), $f_2(x)$ is density function N(3, 0.5), $f_3(x)$ is density function N(7, 0.5). Variant III presents three-modal distributions.

In the experiment we used some measures:

– mean squared error

$$BSK = \frac{1}{n} \sum_{i=1}^{n} [f(x_i) - \hat{f}(x_i)]^2 \quad (6)$$

– maximum value

$$MR = \max_i |f(x_i) - \hat{f}(x_i)| \quad (7)$$

– P is a number of cases, where the value of estimator is greater than value of density function in this point (over smoothing)

– L is a number of cases, where the value of estimator is less than value of density function in this point (under smoothing).

In the experiment, true density functions (described as variant I, II and III) were compared, using measures mentioned above, with the estimators of density function. Kernel estimation was done, based on 128 random observations chosen from populations (described as variant I, II and III), using one of the seven kernel functions (Gaussian, Uniform, Triangle, Epanechnikov, Quartic, Triweight, Cosinus) and using smoothing parameter, which minimizes mean squared error *BSK* (6). Minimalization of *BSK* causes that kernel density estimator can be treated as optimal for this value of parametr *h*. The analysis of values of smoothing parameter, minimizing *BSK*, for particular kernel function allow us to compare the properties of kernel function used in the estimation. The results of this part of study are presented in tables 1, 2, 3.

Table 1. Values of smoothing parameter *h* minimizing BSK for variant I

Kernel function	Value of parametr *h*	*BSK*	*MR*	*P*	*N*
Epanechnikov	**0.0800**	0.028609	0.326695	40	88
Gaussian	**0.0800**	0.030831	0.418698	43	85
Quartic	**0.2100**	0.028370	0.355300	42	86
Triangle	**0.2000**	0.028185	0.359394	41	87
Uniform	**0.1300**	0.031556	0.453241	52	76
Triweight	**0.2400**	0.028592	0.368761	42	86
Cosinus	**0.1800**	0.028325	0.343490	41	87

Table 2. Values of smoothing parameter *h* minimizing *BSK* for variant II

Kernel function	Value of parametr *h*	*BSK*	*MR*	*P*	*N*
Epanechnikov	**0.1600**	0.004511	0.172078	45	83
Gaussian	**0.1800**	0.004702	0.161273	42	86
Quartic	**0.4200**	0.004522	0.168167	52	76
Triangle	**0.4000**	0.004471	0.163618	56	72
Uniform	**0.3000**	0.004993	0.169591	44	84
Triweight	**0.4800**	0.004542	0.168551	53	75
Cosinus	**0.3600**	0.004509	0.174501	49	79

Table 3. Values of smoothing parameter *h* minimizing *BSK* for variant III

Kernel function	Value of parametr *h*	*BSK*	*MR*	*P*	*N*
Epanechnikov	**0.0900**	0.003292	0.116821	82	46
Gaussian	**0.1100**	0.003353	0.108255	74	54
Quartic	**0.2600**	0.003134	0.114355	79	49
Triangle	**0.2500**	0.003139	0.109422	78	50
Uniform	**0.1400**	0.004459	0.145988	84	44
Triweight	**0.3000**	0.003137	0.112873	77	51
Cosinus	**0.2200**	0.003192	0.117113	82	46

The study was also expanded by calculating smoothing parameter h in the estimation the following the density functions: normal with parameters $\mu = 0$ and $\sigma = 1.3$, normal with parameters $\mu = 5$ and $\sigma = 0,2$, uniform on interval $< -1, 1 >$, uniform on interval $< -3, 4 >$, triangle on interval $< 1, 5 >$, gamma with parameters $\lambda = 2$ and $\alpha = 0.5$ (χ^2 with 2 degrees of freedom), gamma with parameters $\lambda = 0.5$ and $\alpha = 1$, gamma with parameters $\lambda = 0.5$ and $\alpha = 5$, gamma with parameters $\lambda = 2$ and $\alpha = 5$ (χ^2 with 10 degrees of freedom). The results concerning values of smoothing parameter h minimizing BSK and BSK (in brackets) are presented in Table 4.

On the basis of the results in presented Tables 1–4 we can divide regarded kernel functions into two groups. Gaussian, Epanechnikov and Uniform kernel functions are kernels that need smaller values of smoothing parameter and the second group: Quartic, Triangle, Triweight and Cosinus kernels need greater values of smoothing parameter in estimation. The first group of kernels is characterised by higher degree of smoothing. This division occurs not only for variant I, but also for two-modal and tree-modal distributions (variant II and variant III).

It allows to estimate value of smoothing parameter for paricular estimator with particular kernel function.

Moreover, values of mean squared error (BSK) in the presented study do not differ significantly.

The results in Table 1 and 4 indicate that:

1. Change of location parameter in normal distribution does not cause change of smoothing parameter h, which minimises BSK.

2. Change of scale parameter in normal distribution and in unimodal gamma distributions ($\alpha > 1$) causes constructing estimator with higher value of smoothing parameter.

3. Change of shape parameter for χ^2 distribution causes more value of smoothing parameter.

4. Exponential distribution ($\alpha = 1$) needs small value of smoothing parameter in comparison with distributions with the same scale parametr.

On the basis of the above analysis we can formulate a statement that values of smoothing parameters, which minimise BSK are different for different kernel functions. It can be explained by different smoothing properties of kernels in the estimation of density function.

Table 4. Values of smoothing parameter *h* minimizing *BSK* (expanded study)

Kernel function	Distributions								
	normal N(0, 1.3)	normal N(5, 0.2)	uniform [−1, 1]	uniform [−3, 4]	triangle [1, 5]	gamma $\lambda=0.5,\ \alpha=5$	gamma $\lambda=2,\ \alpha=5$	gamma $\lambda=2,\ \alpha=0.5$	gamma $\lambda=0.5,\ \alpha=1$
Epanechnikov	0.50 (0.000670)	0.08 (0.028609)	0.12 (0.011055)	0.40 (0.000901)	1.19 (0.009582)	0.46 (0.001128)	1.23 (0.000018)	0.011 (1.697906)	0.029 (0.061773)
Gaussian	0.55 (0.000718)	0.08 (0.030831)	0.16 (0.010688)	0.56 (0.000873)	1.30 (0.010023)	0.51 (0.001201)	1.31 (0.000018)	0.014 (1.703034)	0.032 (0.05989)
Quartic	1.37 (0.000672)	0.21 (0.028370)	0.38 (0.010978)	1.34 (0.000896)	3.21 (0.009737)	0.27 (0.001122)	3.28 (0.000017)	0.031 (1.707840)	0.083 (0.060569)
Triangle	1.30 (0.000667)	0.20 (0.028185)	0.40 (0.010814)	1.39 (0.000883)	3.00 (0.009868)	1.19 (0.001105)	3.05 (0.000016)	0.033 (1.709881)	0.079 (0.058834)
Uniform	0.84 (0.000725)	0.13 (0.031556)	0.21 (0.011178)	0.78 (0.000888)	1.99 (0.009188)	0.74 (0.001273)	2.03 (0.000027)	0.019 (1.6551001)	0.053 (0.06167)
Triweight	1.57 (0.000677)	0.24 (0.028592)	0.44 (0.010893)	1.54 (0.000889)	3.70 (0.009814)	1.45 (0.001131)	3.77 (0.000017)	0.038 (1.706436)	0.094 (0.060112)
Cosinus	1.16 (0.000670)	0.18 (0.028325)	0.27 (0.011057)	0.93 (0.000903)	2.74 (0.009632)	1.07 (0.001123)	2.81 (0.000017)	0.025 (1.701697)	0.067 (0.061466)

III. CONCLUSIONS

The main conclusions resulting from the conducted experiments are the following:

1. Gaussian and Epanechnikov kernels used in kernel estimation need smoothing parameters of the same values. These kernels are characterized by great properties of smoothing.
2. Changing of location, scale or shape parameter in some types of distribution causes changing values of smoothing parameter minimizing mean squared error in estimation of density function.

REFERENCES

Domański Cz., Pruska K., Wagner W. (1998), *Wnioskowanie statystyczne przy nieklasycznych założeniach*, Wyd. UŁ, Łódź.

Hardle W. (1991), *Smoothing Techniques. With Implementation in S*, Springer Verlag, New York.

Priestley M., Chao M. (1972), Nonparametric function fitting, *J. R. Statist. Soc.*, Ser. B, **34**, 385–392.

Rosenblatt M. (1956), Remarks on some nonparametric estimation of a density function, *Ann. Math. Statist.*, **27**, 832–837.

Silverman B. W. (1996), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Wand M., Jones M. (1995), *Kernel Smoothing*, Chapman and Hall, London.

Aleksandra Baszczyńska

UWAGI O WYBORZE FUNKCJI JĄDRA W ESTYMACJI FUNKCJI GĘSTOŚCI

Streszczenie

Funkcja gęstości jest jedną z podstawowych charakterystyk opisujących zachowanie się zmiennej losowej. Najczęściej wykorzystywaną metodą nieparametrycznej estymacji jest estymacja jądrowa. W procesie konstrukcji estymatora konieczne są dwie decyzje, dotyczące parametrów metody: wybór funkcji jądra $K(u)$ oraz wybór parametru wygładzania h. W pracy nacisk położono na wpływ wyboru funkcji jądra na wielkość parametru wygładzania. Eksperyment Monte Carlo dotyczy siedmiu funkcji jądra (gausowskiej, równomiernej, trójkątnej, epanechnikowa, dwukwadratowej, trójkwadratowej i kosinusowej) w estymacji jądrowej funkcji gęstości.