Nonclassical kernel functions in kernel procedures of statistical inference

Aleksandra Baszczyńska[1]

Abstract

Choice of kernel function is one of the basic (besides choice of smoothing parameter) issue in kernel procedures concerning, not only the estimation of numerical and functional characteristics of a random variable, but also testing statistical hypotheses. There is an increasing number of applications where nonclassical kernel function is used. The paper will identify the conditions that are repealed for nonclassical kernel function and their basic properties will be presented. The results of simulation studies concerning the comparison of properties of kernel procedures with classical and nonclassical kernel function will be presented. This will allow the identification of areas where procedures of mathematical statistics with nonclassical kernel functions are optimal.

Keywords: *kernel function, density estimation, smoothing parameter, statistical inference*

JEL Classification: C12, C13

1. Introduction

Kernel function is a device used in varied statistical inference procedures devoted to estimation (e.g. density, distribution function and regression function estimation) or to hypothesis verification in e.g. normality, independence and goodness-of-fit tests. In application of kernel inference methods it is necessary to make decisions about two parameters of the kernel method, where the choice of the smoothing parameter is treated as the crucial one. There are works (cf. Li and Racine, 2007; Baszczyńska, 2014) that discuss this problem widely indicating these methods of choice of smoothing parameter that may be regarded as optimal in determined research situation.

For random sample $X_1, X_2, ..., X_n$ chosen from population with unknown density function $f(x)$, kernel density function estimator is defined as (cf. Rosenblatt, 1956; Parzen 1962):

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

where h is smoothing parameter, $K(u)$ is kernel function.

In classical kernel approach, symmetric, second order kernels $K(u)$ are used. They fulfil the following conditions:

[1] Corresponding author: University of Łódź, Department of Statistical Methods, Rewolucji 1905 r. 41, 90-214 Łódź, Poland, e-mail: albasz@uni.lodz.pl.

$$\int_{-\infty}^{+\infty} K(u)du = 1, \tag{2}$$

$$\int_{-\infty}^{+\infty} uK(u)du = 0, \tag{3}$$

$$\int_{-\infty}^{+\infty} u^2 K(u)du = \kappa_2 \neq 0. \tag{4}$$

The second order of kernel function $K(u)$ indicates that for this function, second moment is that one which is the first nonzero.

The bias term (taking the assumption that $f(x)$ is three times differentiable) of the classical kernel density estimator (with kernel function satisfying conditions (2)-(4)) is of the following form:

$$Bias\left[\hat{f}_h(x)\right] = \frac{h^2}{2} f^{(2)}(x) \int_{-\infty}^{+\infty} u^2 K(u)du + O\left(h^3\right). \tag{5}$$

The analysis shows that reducing the bias can be done both by decreasing the value of smoothing parameter h or using higher order kernel functions. In addition, applying kernels of order higher than 2 results in reducing mean squared error of kernel estimator. While weakening the assumption about the differentiability of $f(x)$ to being twice differentiable leads to:

$$Bias\left[\hat{f}_h(x)\right] = \frac{h^2}{2} f^{(2)}(x) \int_{-\infty}^{+\infty} u^2 K(u)du + o\left(h^2\right). \tag{6}$$

For the bias of order h^2, kernel estimator (1) is asymptotically unbiased (Wand and Jones, 1995). The dependence of the bias (6) on the true density function $f(x)$ indicates that for large absolute value of the second derivative of $f(x)$, the bias is large. It can be noticed in the regions where the curvature of the density is high.

The generalization of conditions (2)-(4) for υ th order kernel (υ is integer) is the following:

$$\int_{-\infty}^{+\infty} K(u)du = \kappa_0 = 1, \tag{7}$$

$$\int_{-\infty}^{+\infty} u^l K(u)du = \kappa_l = 0 \quad \text{for} \quad l = 1,\ldots,\upsilon - 1, \tag{8}$$

$$\int_{-\infty}^{+\infty} u^\upsilon K(u)du = \kappa_\upsilon \neq 0. \tag{9}$$

Applying υth order kernel in kernel density estimator with taking the assumption that $f(x)$ is υth order differentiable results in:

$$Bias\left\lfloor \hat{f}_h(x) \right\rfloor = O\left(h^{\nu}\right). \quad (10)$$

It means the possibility of reducing the order of the bias of $\hat{f}(x)$ from $O\left(h^2\right)$ to $O\left(h^{\nu}\right)$ but in the situation when relaxing the restrictions that the kernel is a density function is acceptable. For example, when $K(u)$ is a density function, second moment of the kernel fulfills: $\kappa_2 > 0$, assuming that second derivative of unknown density function is continuous, square integrable and ultimately monotone. When this assumption is strengthened to continuous square integrable fourth derivative, for $\kappa_2 = 0$ kernel function does not need to be density function.

It is worth noticing that the require that kernel function is symmetric implies that υ is even.

2. Higher order kernel function construction

In the construction of kernel function, for example, of order υ=4, kernel of order υ=2 is used with setting up a polynomial in its argument and solving for the roots of the polynomial subject to desired moments constraints.

For a second order Gaussian and Epanechnikov kernels given by the formulas, respectively:

$$ {}_2^G K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-u^2}{2}\right), \quad {}_2^E K(u) = \begin{cases} \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}u^2\right) & for \quad u^2 < 5 \\ 0 \quad otherwise \end{cases}, $$

the corresponding fourth and sixth order Gaussian and Epanechnikov kernels are the following:

$$ {}_4^G K(u) = \left(\frac{3}{2} - \frac{1}{2}u^2\right) \frac{\exp\left(\frac{-u^2}{2}\right)}{\sqrt{2\pi}}, \quad {}_6^G K(u) = \left(\frac{15}{8} - \frac{5}{4}u^2 + \frac{1}{8}u^4\right) \frac{\exp\left(\frac{-u^2}{2}\right)}{\sqrt{2\pi}}, $$

$$ {}_4^E K(u) = \begin{cases} \frac{3}{4\sqrt{5}}\left(\frac{15}{8} - \frac{7}{8}u^2\right)\left(1 - \frac{1}{5}u^2\right) & for \quad u^2 < 5 \\ 0 \quad otherwise \end{cases}, $$

$$ {}_6^E K(u) = \begin{cases} \frac{3}{4\sqrt{5}}\left(\frac{175}{64} - \frac{105}{32}u^2 + \frac{231}{320}u^4\right)\left(1 - \frac{1}{5}u^2\right) & for \quad u^2 < 5 \\ 0 \quad otherwise \end{cases}. $$

As a generalization, the formula for the construction of the kernel function of higher order can be written in the following way:

$$ {}_{v+2}K(u) = \frac{3}{2}{}_{v}K(u) + \frac{1}{2}u\,{}_{v}K'(u) \tag{11} $$

where ${}_{v}K(u)$ denoting v th order kernel, is assumed to be differentiable.

A detailed discussion on constructing kernels of these type and indicating pros and cons of them including issues connected with interpretations and plausibility are widely presented in literature (e.g. Wand and Jones, 1995).

Another approach for constructing higher order kernels is based on second order Gaussian kernel. The higher order kernel is the following (Wand and Schucany, 1989, 1990):

$$ {}_{v}K(u) = \sum_{l=0}^{\frac{v}{2}-1} \frac{(-1)^l}{2^l l!}\phi^{(2l)}(x) \tag{12} $$

where $l = 0,2,4,...$, $\phi(x)$ denotes Gaussian kernel.

The approach of higher kernels can be extended to kernels for which $\int_{-\infty}^{+\infty} u^l K(u)du = \kappa_l = 0$ for all $l = 1,2,...$. The example of kernels of this type called the infinite order kernels is the sinc kernel (Bissantz and Holzmann, 2007; Glad et al., 2007):

$$ K(u) = \frac{\sin x}{\pi x}. \tag{13} $$

The sinc kernel is widely used both in the density estimation and regression estimation because of its optimality properties in terms of mean square error and mean integrated square error. But its good properties depends in fact on the degree of smoothness of density. In addition, applying this kernel to samples of finite sizes does not guaranteed satisfactory results because of its only good asymptotic performance.

3. Optimal kernels

The optimality of kernel choice is the crucial issue in both theoretical and practical point of view. Because of the dependence between kernel function and measure of the estimation error, being regarded not only at a single point but also at the whole real line, the optimal measure of the error is discussed in the literature (Hansen 2005; Jones and Signorini, 1997). One of the approaches to this problem is based on a class of minimum of the asymptotic variance of kernel estimators. The minimum variance kernel is the solution of $\min \int_{-\infty}^{+\infty} K^2(u)du$

subject to $K \in S_{\nu,k}$, where $0 \le \nu \le k-2$, ν and k are of the same parity and

$$S_{\nu,k} = \begin{cases} K \in Lip[-1,1], \quad \text{support}(\mathrm{K}) = [-1,1], \\ \int_{-1}^{1} u^j K(u)du = \begin{cases} 0 & for \quad 0 \le j < k,\ j \ne \nu, \\ (-1)^{\nu} \nu! & for \quad j = \nu, \\ \kappa_{\nu} \ne 0 & for \quad j = k. \end{cases} \end{cases}$$

The formula for the optimal kernels $K \in S_{\nu,k}$ is the following (cf. Horová, 2000; Horová et al., 2012):

$$K(u) = \frac{(-1)^{\nu} \nu!}{2} \sum_{i=\nu}^{k-2} (2i+1) p_{\nu}^{i} \left(\sum_{r=0}^{i} p_r^i u^r - \sum_{k=0}^{i} p_r^k u^k \right). \tag{14}$$

It means that the optimal kernel is polynomial of degree k restricted on the interval $[-1,1]$. For even k this polynomial is symmetric while for odd k it is asymmetric. For example for $k = 2$, $\nu = 0$ we can get kernel function $K(u) = -\frac{3}{4}(u^2 - 1)$; for $k = 3$, $\nu = 1$ we can get kernel function $K(u) = \frac{15}{4} u (u^2 - 1)$; for $k = 4$, $\nu = 2$ we can get kernel function $K(u) = -\frac{105}{16}(u^2 - 1)(5u^2 - 1)$. All kernels functions are presented on $[-1,1]$.

Imposing some level of smoothness on kernel function by order of smoothness μ, the class of smooth kernels can be determined (cf. Müller, 1984): $K \in S_{\nu,k} \cap C^{\mu}[-1,1]$, $K^{(j)}(-1) = K^{(j)}(1) = 0$, $j = 0,1,..., \mu$. Some mostly used smooth optimal kernels are:

– quartic kernel $K \in S_{0,2}^{1}$ (for $k = 2$, $\nu = 0$, $\mu = 1$): $K(u) = \begin{cases} \frac{15}{16}(u^2 - 1)^2 & for \quad u \in [-1,1], \\ 0 & for \quad u \notin [-1,1]. \end{cases}$

– triweight kernel $K \in S_{0,2}^{2}$ (for $k = 2$, $\nu = 0$, $\mu = 2$):

$$K(u) = \begin{cases} -\frac{35}{32}(u^2 - 1)^3 & for \quad u \in [-1,1], \\ 0 & for \quad u \notin [-1,1]. \end{cases}$$

– second derivative of triweight kernel $K \in S_{2,4}^{0}$ (for $k = 4$, $\nu = 2$, $\mu = 0$):

$$K^{(2)}(u) = \begin{cases} -\frac{105}{16}(u^2 - 1)(5u^2 - 1) & for \quad u \in [-1,1], \\ 0 & for \quad u \notin [-1,1]. \end{cases}$$

4. Results of the simulation study

In the study density kernel estimators were analyzed to indicate the performance of density estimators depending on used higher order kernel functions. Samples of different sizes (n = 10, 20, …, 90, 100) were drawn from populations (with known density functions), next estimators od density functions (with different order of kernels) were calculated and compared with true density functions.

In the simulation study fifteen populations, introduced by Marron and Wand (1992) are taken into regard. This is a collection of Gaussian mixture models used widely in works concerning analysis of kernel methods estimation. It represents variety of distributions including symmetric, asymmetric, unimodal and multimodal ones, for example, distributions (1)-(7) represent symmetric distributions (with Gaussian distribution), (1)-(5) unimodal distributions, (8)-(15) multimodal distributions, (12)-(15) asymmetric distributions. This collection allows to take into account populations with different sides and levels of distribution characteristics.

When a sample is drawn from Gaussian standardized distribution, optimal kernels for $\nu = 2$ results in unimodal density estimator. Even when parameter $\mu = -1$ (denoting noncontinuous kernel) is used. It is worth noticing that for values of μ bigger than 2, the shape of the resulting estimators do not change, though along with the increasing values of the μ, the values of smoothing parameter become bigger. It means that increasing parameter μ influences on the level of density estimator smoothing. Figures 1-3 present density estimators for different optimal kernels when n = 50 and smoothing parameter is chosen by reference rule.

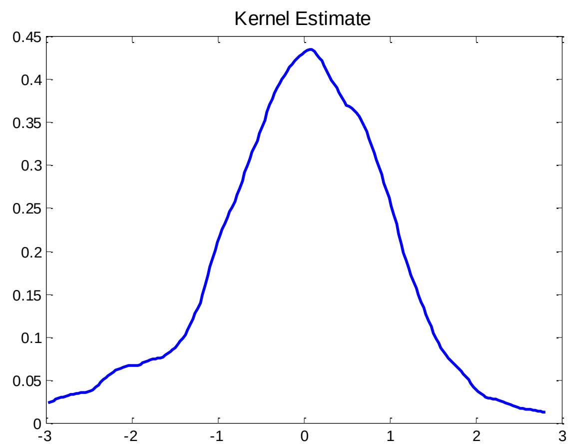

Fig. 1. Kernel density estimator (sample from Gaussian density; Epanechnikov kernel – optimal kernel with $\nu = 0$, $k = 2$, $\mu = 0$; n = 50; $h = 0.84409$).

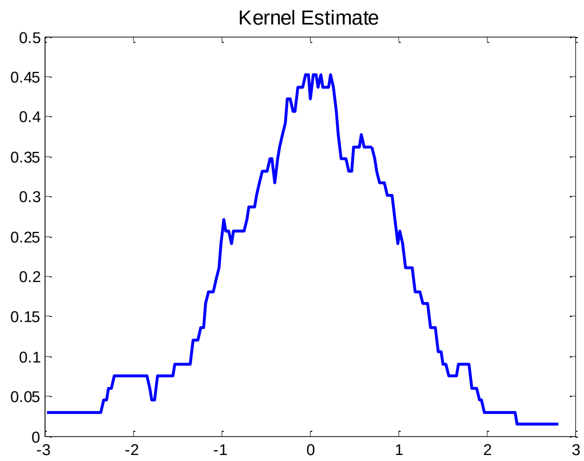

Fig. 2. Kernel density estimator (sample from Gaussian density; optimal kernel with $\nu = 0$, $k = 2$, $\mu = -1$; n = 50; $h = 0.66346$).

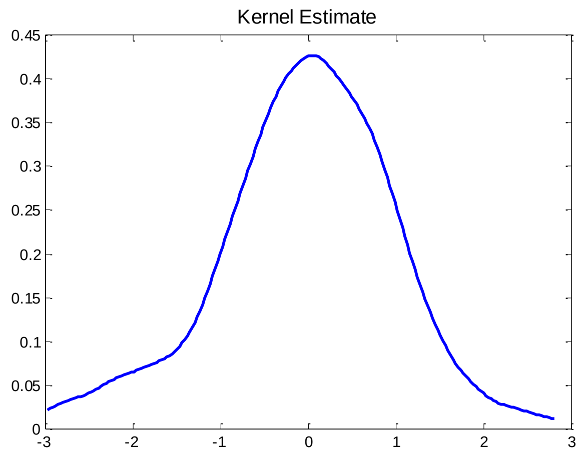

Fig. 3. Kernel density estimator (sample from Gaussian density; optimal kernel with $\nu = 0$, $k = 2$, $\mu = 1$; n = 50; $h = 1.1355$).

The similar results can be obtained when samples are drawn from other unimodal distributions (1)-(5) from Marron and Wand collection and for other sample sizes. What is interesting that basing on even small samples the resulting estimators properly indicate the shape of the true density. For kernel parameter $\nu = 2$ (kernel function is bimodal and is not a density function) density estimator imitates the shape of kernel function.

When samples from multimodal distributions are taken into account it is very difficult to find such kernel for which density estimator shows all significant features of distribution, including number of modes. Figures 4-6 shows density estimators for samples from distribution #15 (asymmetric and multimodal).

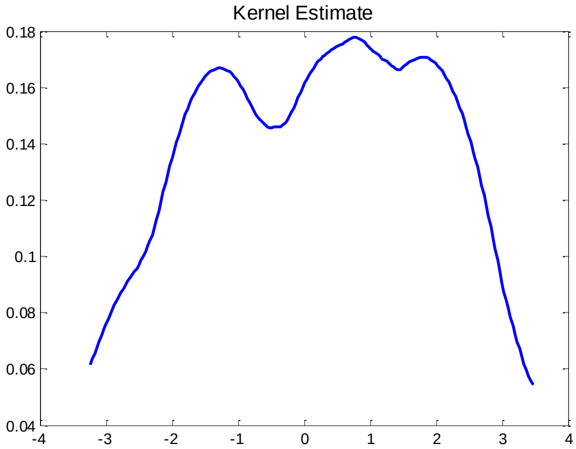

Fig. 4. Kernel density estimator (sample from density #15; optimal kernel with $\nu = 0$, $k = 2$, $\mu = 0$; n = 50; $h = 1.8532$).

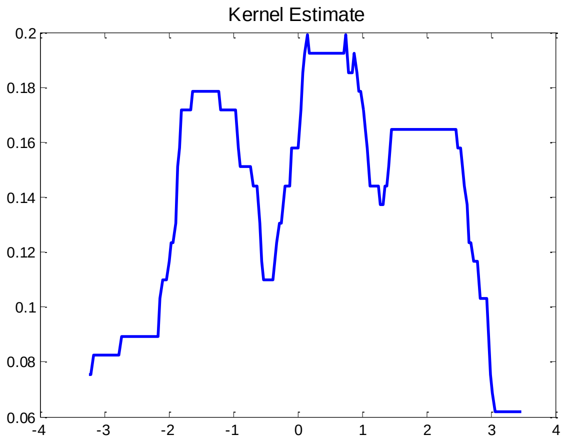

Fig. 5. Kernel density estimator (sample from density #15; optimal kernel with $\nu = 0$, $k = 2$, $\mu = -1$; n = 50; $h = 1.4566$).

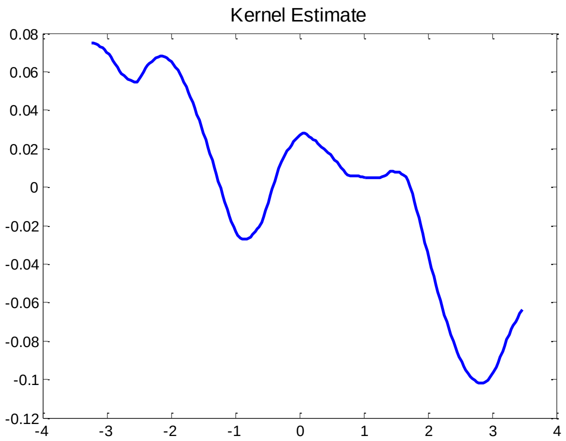

Fig. 6. Kernel density estimator (sample from density #15; optimal kernel with $\nu = 1$, $k = 3$, $\mu = 0$; n = 50; $h = 2.1045$).

5. Empirical illustrations

Regarded methods were applied to data sets concerning the employment of the largest companies in Central and Eastern Europe in 2013 (top 500 CCE, *http://www.coface.com* [22.03.2015]). The samples of various sizes were drawn and kernel density estimators were calculated. Figure 7 presents the illustrative estimators for sample size n = 50.

Optimal kernel parameters

$\nu = 0,\ k = 2,\ \mu = 0$ $\quad$ $\nu = 0,\ k = 2,\ \mu = -1$ $\quad$ $\nu = 1,\ k = 3,\ \mu = 0$ $\quad$ $\nu = 1,\ k = 3,\ \mu = -1$

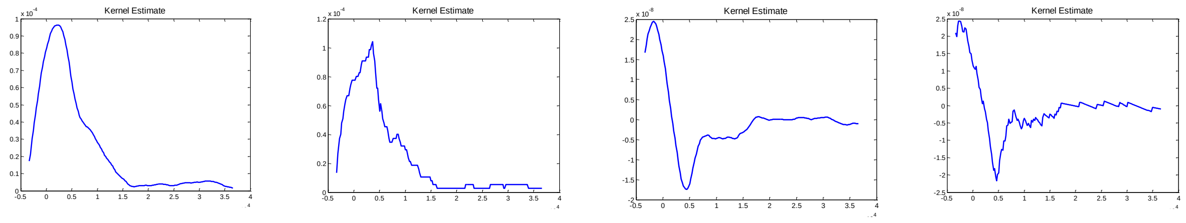

Fig. 7. Kernel density estimators for real data sets (n = 50).

Conclusion

The main aim of using higher order kernels is to reduce the bias in kernel procedures of estimation. But this tool should be used very carefully. Simulation study shows that higher order kernels are very useful when distribution of population from which the sample is drawn is unimodal and symmetric. In the case of even slightly changes from this state results in density estimator which does not answer the requirements of the researches. Nonclassical kernels of higher order certainly broaden the applications of procedures based on kernel methods. It seems to be easier to have some additional information about population and use it in the process of choosing the proper kernel function in kernel inference procedures. The results of simulation study and the trial of application of regarded methods for real data indicate that further research should be referred to the influence between shape of kernel function and value of smoothing parameter.

References

Baszczyńska, A. (2014). Testing significance of peaks in kernel density estimator by SiZer map. In: Papież, M., & Śmiech, S. (eds.), *Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*. Cracow: Foundation of the Cracow University of Economics, 9-17.

Bissantz, N., & Holzmann, H. (2007). Estimation of a quadratic regression functional using the sinc kernel, *Journal of Statistical Planning and Inference*, *137*(3), 217-219.

Glad, I., Hjort, N., & Ushakov, N. (2007). Density estimation using the sinc kernel. Retrieved from http://www.academia.edu/3392551/Density_Estimation_Using_the_Sinc_Kernel.

Hansen, B. (2005). Exact mean integrated squared error of higher order kernel estimators, *Econometric Theory*, *21*, 1031-1057.

Horová, I. (2000). Some remarks on kernels, *Journal of Comput, Analysis and Applications*, The University of Memphis, USA: Plenum Publishing Corporation, *2*(2), 253-263.

Horová, I., Koláček, J., & Zelinka, J. (2012). *Kernel smoothing in Matlab. Theory and practice of kernel smoothing*, New Jersey: World Scientific.

Jones, M., & Signorini, D. (1997). A comparison of higher-order bias kernel density estimators, *Journal of the American Statistical Association*, *92*(439), 1063-1073.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics. Theory and practice*, Princeton and Oxford: Princeton University Press.

Marron, J., & Wand, M. (1992). Exact Mean Integrated Squared Error. *The Annals of Statistics*, *20*(2), 712-736.

Müller, H-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics*, *12*(2), 766-774.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, *33*(3), 1065-1076.

Rosenblatt, M. (1956). Remarks on some nonparametric estimation of a density function. *Annals of Mathematical Statistics*, *27*(3), 832-837.

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing*, London: Chapman and Hall.

Wand, M. P., & Schucany, W. R. (1989). Gaussian-based kernels for curve estimation and window width selection. Retrieved from http://www.smu.edu/search?q =wand %20schucany.

Wand, M. P., & Schucany, W. R. (1990). Gaussian-based kernels. *The Canadian Journal of Statistics*, *18*, 197-204.