PAN POLSKA AKADEMIA NAUK

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

M&MS

KERNEL ESTIMATION OF PROBABILITY DENSITY FUNCTION: PROPERTIES AND PARAMETERS OPTIMIZATION

Bartosz Jerzman [1)], Wiesław Kiciński [2)]

1) Naval Information Technology Group,Arciszewskiego 25, 81-103 Gdynia, Poland (b.jerzman@mw.mil.pl)

2) Nicolaus Copernicus University, Institute of Physics,Grudziącka 5, 87-100 Toruń,, Poland (✉ w.kicinski@fizyka.umk.pl, +48 56 611 3343)

Abstract

The article deals with the kernel estimation of the probability density function. The main subject of the research is the optimization of parameters of the estimator. In the particular case, the research focused on the estimation of the univariate, unimodal data representative for the normal distribution.

Keywords: probability density function, kernel estimation.

1. Introduction

The probability density function (PDF) depicts the structure of the statistical population and characterizes the distribution of the random variable [1]. It enables the construction and verification of the statistical models which are essential for data analysis and measurements. Its multimodality and boundary values associated with the tails are indispensable for the identification of the structure of the measuring data.

The probability density function for a one-dimensional case is defined by equation

$$P(a \le X \le b) = \int_a^b f(x)dx \tag{1}$$

and indicates the probability of the value X from a particular interval $[a,b]$.

In multivariate data analysis, the determination of probability for every n-dimensional vector enables the calculation of probability for any event in the multidimensional space.

There are two methods of t probability density estimation:

– parametric,
– non-parametric.

In parametric statistics it is assumed that variable is derived from a specific probability distribution for which distribution parameters are calculated. The limited number of known and described distributions is the main drawback of the parametric inference.

Non-parametric methods do not require any assumptions concerning the distribution of the random variable. Therefore, these methods are commonly applied in case of no a priori information about the random variable distribution or when there is no possibility of modeling the density by any standard distribution.

The main reason for the development of the non-parametric methods was the need for complex probabilistic models and the increase of computer processing power. More recent work [2,3,4] indicates that PDF can be a useful tool in transient detection in the measured signals or in picture recognition.

Statistical inference deals with the characteristic features of a population by estimating its values or a function from the data derived from a random sample. Estimators are therefore an approximation of the values or of a function [1].

Among the non-parametric methods of estimating the probability density function the following may be distinguished: histogram, naive estimator, kernel density estimator, nearest-neighbor estimator, Fourier estimator.

The theoretical backgrounds of PDF estimation are widely discussed in literature [1,5,6,7]. This paper deals with the kernel estimation and its main goal is to present some simulation results of kernel density estimation. The presented results of numerical experiments are addressed to users who face the problem of the kernel function choice and the adjustment of its parameters. We study the accuracy of various methods of kernel density estimation and try to focus our attention on the optimization of the quality of the kernel estimation, their parameters and the numerical complexity.

The organization of this paper is as follows. Section 2 introduces the definition and most important properties of the kernel estimation. Reviews briefly the main roles of construction of effective kernel estimators. In Section 3 we draw experimental performance evaluations of PDF models in order to indicate the advantages and limitations of various kernels and the methods of choosing the smoothing parameter. Section 4 contains the concluding remarks.

2. Kernel estimation

A kernel density estimate comes from the preceding methods of estimation: histogram and naive estimator. The histogram is the simplest density estimator, where given an origin x and a bin width h, the bins of the histograms H_k are defined as the intervals $\left[x+mh, x+(m+1)h\right)$. Thus, the histogram is defined by [1]

$$\hat{f}(x) = \frac{\#\{x_i \in H_k\}}{mh}, \tag{2}$$

where # denotes the number of $x_i \in H_k$.

Figure 1 shows histograms of the same random variable x for different bin widths h. When the value of parameter h is too big, the characteristic features of the density are lost. On the contrary, a small value of h results in the exposition of too many details what blurs the density picture (two modes). What is more, the choice of origin of the first bin has a significant impact on the quality of the histogram.

The naive estimator is void of this drawback. The histogram is constructed under the assumption that every value of the random variable is the origin of a new bin. The naive estimator function is discontinuous, it has a zero derivative in all points except $X_i \pm h$ and is defined by

$$\hat{f}(x) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{h}w\left(\frac{x-X_i}{h}\right), \tag{3}$$

where $w(x) = \frac{1}{2}$, when $|x| < 1$ or $w(x) = 0$ in other cases [1].

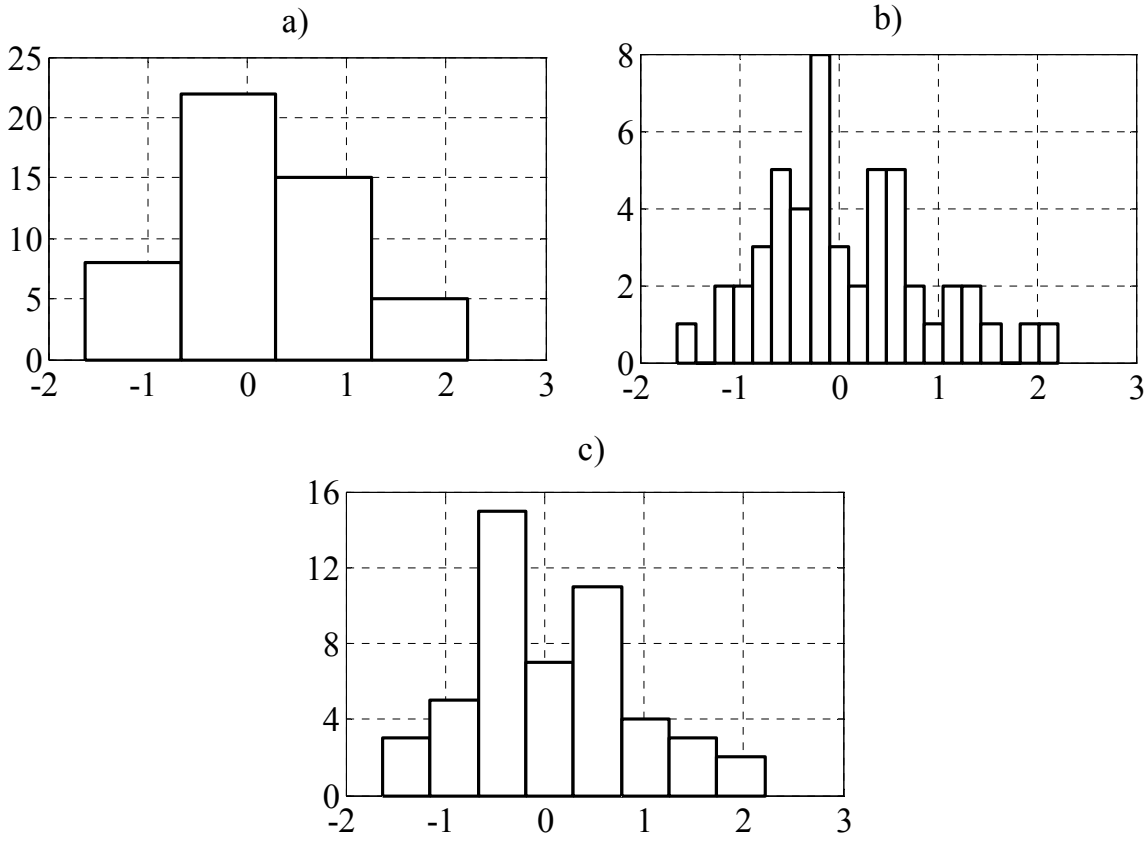

Fig. 1. Histograms of the probability density function.

In kernel density estimators the weight function $w(x)$ is replaced by the kernel function K which satisfies the following conditions [5]:

1. $\int_{-\infty}^{\infty} K(x)dx = 1$.
2. Is symmetrical function, i.e.
   $K(x) = K(-x)$ for every $x \in \Re$.
3. Has maximum for x=0, i.e.

$K(0) \geq K(x)$ for every $x \in \Re$.

For random sample $X_1, X_2, ..., X_m$ the kernel estimator $\hat{f}(x)$ for every $x \in \Re$ is defined by

$$\hat{f}(x) = \frac{1}{mh} \sum_{i=1}^{m} K\left(\frac{x - X_i}{h} \right), \tag{4}$$

where h is the smoothing parameter or bandwidth - equivalent to the bin width in case of histograms. The choice criteria for this parameter are similar to those of histograms.

The kernel density estimate can be interpreted as the sum of single estimates, hence the sum of successive kernel functions K shifted by a vector X_i and scaled by the factor $\frac{1}{mh}$. Moreover, $\hat{f}(x)$ satisfies the conditions of integrality and differentiability, which are imposed

on the kernel function K. Most applications exploit the non-negative, symmetrical due to zero kernel which has its maximum there, usually it is the normal density kernel [5].

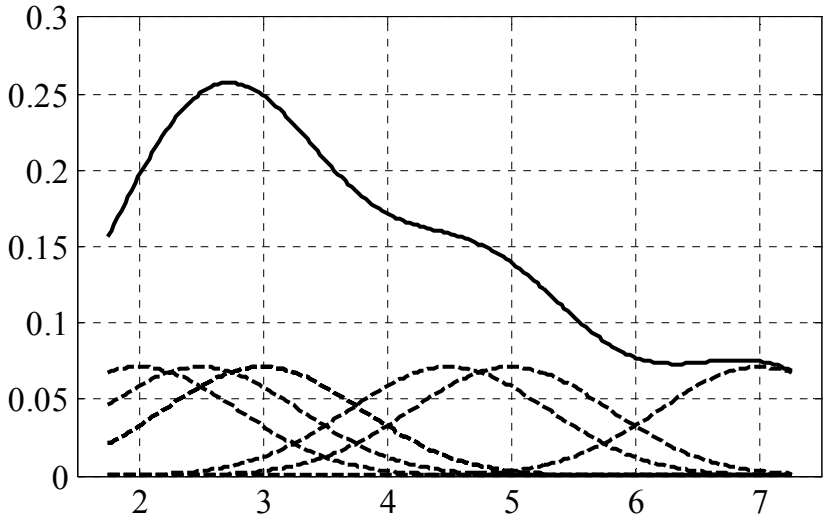

Fig. 2. An examples of the estimate of the probability density function (solid curve) constructed with the kernel estimators (dotted curves).

The concept of kernel estimation is presented in Fig. 2, for the m-element random sample with values X_i = [-2; -1,5; 0; 0; 1; 2]. In this example the estimate was constructed with the normal kernels. Fig. 2 is the graphical interpretation of (4). The appropriate choice of the bandwidth, hence the shape of the kernel, will have the fundamental significance for the quality of the density estimate.

The quality of the estimate is determined by two factors: smoothing parameter h and kernel function K.

A criterion of estimation of the kernel estimate quality is the bias estimator defined as the difference between the expected value of estimators and the "true" values of the estimated parameter

$$E\left(\hat{f}\right) - f \,. \tag{5}$$

Zero value of the bias, thus the case when the estimator is unbiased, means that the estimated values fluctuate between the "true" values of the distribution and that the estimator has no tendency to over- or underestimate. In practice, the estimated values oscillate between true values, while the variance $VAR\left(\hat{f}\right)$ is the indicator of these oscillations. The estimator is consistent if

$$\lim_{m\to\infty} VAR\left(\hat{f}\right) = 0 \tag{6}$$

and asymptotically unbiased if

$$\lim_{m\to\infty} E\left(\hat{f}\right) = f \qquad . \tag{7}$$

The analysis of bias is the basis for two criteria of evaluation of the estimate quality: Mean Squared Error (MSE) and Mean Integrated Squared Error (MISE).

MSE is defined by

$$MSE_x = E\left(\hat{f}(x) - f(x)\right)^2 \tag{8}$$

or

$$MSE_x = \left[E\left(\hat{f}(x)\right) - f(x)\right]^2 + VAR\left(\hat{f}(x)\right). \tag{9}$$

Mean squared error enables the evaluation of the difference between the mean value of the estimator and the true values, whereas the variance determines the measure of spread of the estimated values with respect to the mean value.

MISE is a global indicator of the estimate quality not only for a given x as in the case of MSE. MISE is defined as the integral of MSE (in the $\Re^n$ space):

$$MISE = \int_{\Re^n} E\left(\hat{f}(x) - f(x)\right)^2 dx \tag{10}$$

or

$$MISE = \int_{\Re^n} \left[E\left(\hat{f}(x)\right) - f(x)\right]^2 + VAR\left(\hat{f}(x)\right) dx. \tag{11}$$

The bias of the kernel estimator after extension into a Taylor series for the given x is defined by

$$E\left(\hat{f}(x)\right) - f(x) \approx \frac{1}{2} h^2 f''(x) \int_{-\infty}^{\infty} x^2 K(x) dx + o(h), \tag{12}$$

while the variance has the form

$$VAR\left(\hat{f}(x)\right) \approx \frac{1}{mh} f(x) \int_{-\infty}^{\infty} K(x)^2 dx + o\left(\frac{1}{mh}\right), \tag{13}$$

where *o(h)* and *o(*1*/mh)* represent the infinitely small higher-order terms of *h* and *(*1*/mh),* respectively.

The following terms can be inserted in order to simplify the notation [5]

$$U(K) = \int_{-\infty}^{\infty} x^2 K(x) dx. \tag{14}$$

$$W(K) = \int_{-\infty}^{\infty} K(x)^2 dx. \tag{15}$$

$$Z(f) = \int_{-\infty}^{\infty} f''(x)^2 dx. \tag{16}$$

Insertion of (14), (15) and (16) into (12) gives the mean integrated square error of the form:

$$MISE \approx \frac{1}{4} h^4 U(K)^2 Z(f) + \frac{1}{mh} W(K). \tag{17}$$

The expression (17) implies that the smaller value of the smoothing parameter (which gives amore sharpened structure of the data) results in the decrease of the bias at the same time leading to the increase of the estimator's variance. Increasing the values of h causes the

opposite effect. The main issue concerning the choice of the bandwidth is finding the optimal value that is a compromise between the bias and the variance.

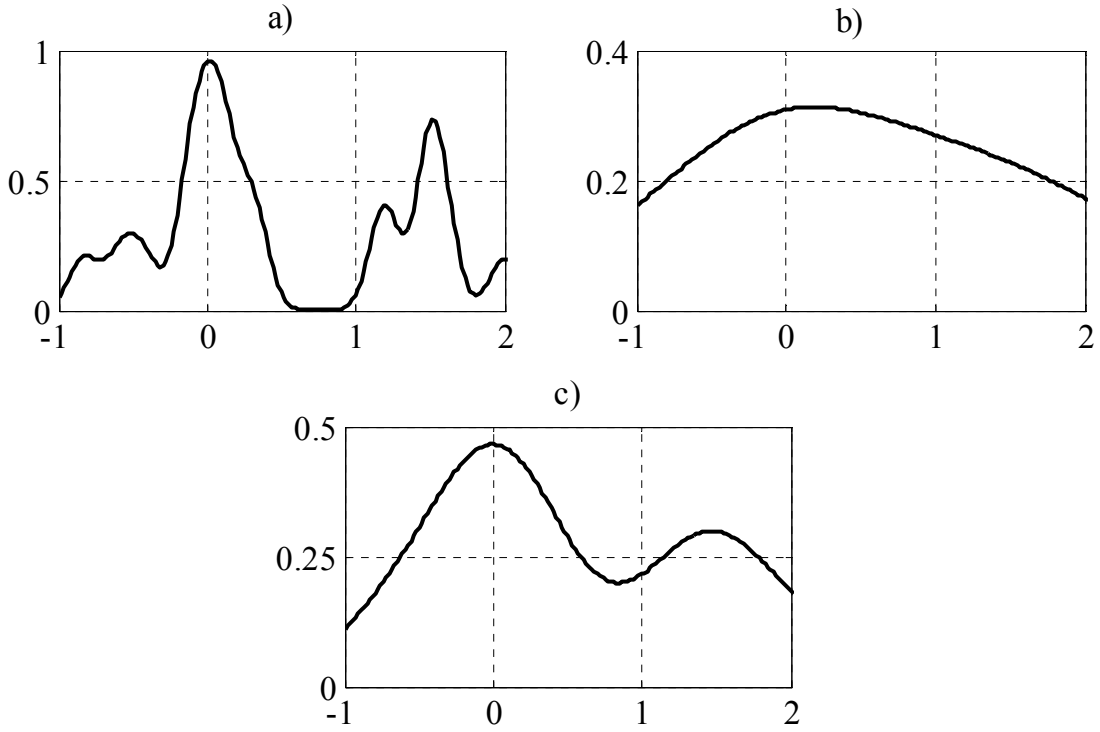

Fig. 3. An example of kernel density estimates with different bandwidths: a) *h*=0,1; b) *h*=0,8; c) *h*=0,4.

This optimal value minimizing the MISE criterion is defined as follows [5]

$$h_{opt} = \left(\frac{W(K)}{mU(K)^2 Z(f)} \right)^{\frac{1}{5}}. \tag{18}$$

This expression implies that the optimal value depends on the unknown density-$Z(f) = \int_{-\infty}^{\infty} f''(x)^2 \, dx$, and that the optimal bandwidth tends to zero as the sample size increases. Secondly, the term $Z(f)$ describing the rate of change of the probability density indicates that the small value of h should be adopted in the multimodal distributions. Due to the unknown factor f this expression cannot however be applied. The natural conception would lead to the conclusion that the term $Z(f)$ is determined by the arbitrary assumption regarding one of the standard distributions. In most cases it is a normal distribution.

2.1. Methods of optimal bandwidth determination

We may distinguish between three general methods for the determination of the optimal bandwidth: approximate, plug-in and cross-validation methods.

The approximate method assumes that $Z(f)$ complies with the normal distribution. Denoting the variance of normal distribution as σ^2, $Z(f)$ has the form [5]

$$Z(f) = \int_{-\infty}^{\infty} f''(x)^2 \, dx = \frac{3}{8}\pi^{-\frac{1}{2}}. \tag{19}$$

If the normal kernel function is applied, then the optimal bandwidth is

$$h_{opt} = (4\pi)^{-\frac{1}{10}} \left(\frac{3}{8}\pi^{-\frac{1}{2}} \right)^{-\frac{1}{5}} \sigma n^{-\frac{1}{5}} = 1{,}06\sigma n^{-\frac{1}{5}}. \tag{20}$$

The above expression is known as the Silverman method of Rule of Thumb – ROT [1]. Unfortunately, for the multimodal distributions which are the general object of the kernel density estimation, too large overestimation results in covering of the characteristic features of the distribution. In practice, it is applied initially to obtain the approximate value of the probability distribution function.

Plug-in methods consist of iterative determination of the function value $Z(f)$. Firstly, due to the presence of the second derivative of the probability density function f'' in the above expression, the bandwidth of the estimator of the *k*-order derivative of the density function should be calculated using the approximate method. In order to determine the bandwidth of function f, the smoothing parameters for $f^{(2k-2)}$, $f^{(2k-4)}$ and up to f'' are calculated. The accuracy of this method depends on the number of steps *k.* In practice two steps are sufficient [8]. The study applies a representative plug-in method proposed by Hall, Sheather, Jones and Marron in [9].

Cross-validation methods are based on the concept of minimizing MISE. They belong to the so-called classical methods that are used in parametric modeling. The smoothing parameter h for the *m*-element sample size is determined on the basis of the function estimate obtained for the $m-1$ observations [5].

This is a very accurate method due to the minimization of MISE, however the tendency for underestimation may occur resulting in too many local extremes. On the other hand, plug-in methods are characterized by good smoothing properties of the density estimate. Therefore, the choice of the bandwidth selection method depends on its intended application [7,8].

Additionally, the estimators may be optimized with reference to the "local" density of the data. An example of such optimization is the local version of the cross-validation method where the nearest neighbor method is adapted to control the amount of smoothing. In effect, the value of the bandwidth is not constant and varies in accordance to the density of distribution [1].

2.2. Choice of the kernel function

The second factor affecting the quality of the estimation is the choice of kernel. The analysis of the formula (17) proves that after substituting the value of the optimal smoothing parameter (18), the mean square error is calculated as follows

$$MISE = C(K)Z(f)^{\frac{1}{5}} m^{-\frac{4}{5}}, \tag{21}$$

where the coefficient $C(K)$ is defined by

$$C(K) = U(K)^{\frac{2}{5}} W(K)^{\frac{4}{5}}. \tag{22}$$

The coefficient $C(K)$ characterizes the kernel function K of the estimator. The kernels with small value of this coefficient should be chosen since it leads to the minimization of MISE for the given optimal bandwidth h_{opt}.

Table 1. Basic kernel functions and their efficiency [5].

Kernel type	$K(x)$	$U(K)$	$W(K)$	The difference in efficiency in comparison to the Epanachnikov kernel
Epanachnikov	$\frac{3}{4}\left(1-x^2\right)$ for $x \in [-1,1]$	$\frac{1}{5}$	$\frac{3}{5}$	-
Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$	1	$\frac{1}{2\sqrt{\pi}}$	5%
Rectangular	$\frac{1}{2}$ for $x \in [-1,1]$	$\frac{1}{3}$	$\frac{1}{2}$	7%
Biweight	$\frac{15}{16}\left(1-x^2\right)^2$ for $x \in [-1,1]$	$\frac{1}{7}$	$\frac{5}{7}$	1%
Triangular	$1-\lvert x\rvert$ for $x \in [-1,1]$	$\frac{1}{6}$	$\frac{2}{3}$	1%

The kernel that minimizes $C(K)$ is the Epanachnikov function defined as follows [1]

$$K_e = \begin{cases} \frac{3}{4}\left(1-x^2\right), & \text{for } x \in [-1,1] \\ 0, & \text{for } x \in \left[(-\infty,-1)\cup(1,\infty)\right] \end{cases}. \tag{23}$$

The other kernel functions are suboptimal in comparison to the Epanachnikov function. Their efficiency is determined as follows [5]

$$ef(K) = \left(\frac{C(K_e)}{C(K)}\right)^{\frac{5}{4}}. \tag{24}$$

For a large sample size the mean squared error is the same for *m* observations using any kernel functions and for a sample size of $m \cdot ef$ using the Epanachnikov kernel.

On the basis of the MISE criterion, the optimal choice is the Epanachnikov kernel if the quality of the estimate is taken into consideration. However, it has to be noticed that the differences between the efficiency of various kernels are insignificant. In practice, the values from Table 1 respond to the increase in the sample size for other kernel functions in order to achieve inaccuracy corresponding to the estimate achieved using the Epanachnikov kernel.

3. Results of simulation analysis

3.1. Statistics

The simulation studies concentrated on the examination of the influence of the type of kernel and the bandwidth on the quality of the estimation. The aim of the research was the assessment of conformity between the curves representing the kernel estimator and the one-dimensional normal distribution. The following indicators of the curve fitting were applied:

– the mean square error

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(x_i^n - x_i^e\right)^2, \tag{25}$$

where

- x_i^n - the value of the ordinate of the one-dimensional normal distribution in point *i*,
- x_i^e - the value of the ordinate of the kernel estimator in point *i*,
- the variance (it characterizes the spread of the difference between the theoretical and the estimated curve)

$$VAR = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{x} - x_i\right)^2, \tag{26}$$

where

$$x_i = x_i^n - x_i^e \quad \text{and} \quad \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

- the mean square error of the censored distribution on *t*-level

$$MSE_t^o = \frac{1}{N_2 - N_1}\sum_{i=N_1}^{N_2}\left(x_i^n - x_i^e\right)^2, \tag{27}$$

- the censored variance on t-level

$$VAR_t^o = \frac{1}{N_2 - N_1}\sum_{i=N_1}^{N_2}\left(\bar{x} - x_i\right)^2, \tag{28}$$

where

N_1, N_2 – left and right limits of the one-dimensional normal distribution curve and the kernel estimator on the threshold value *t*.

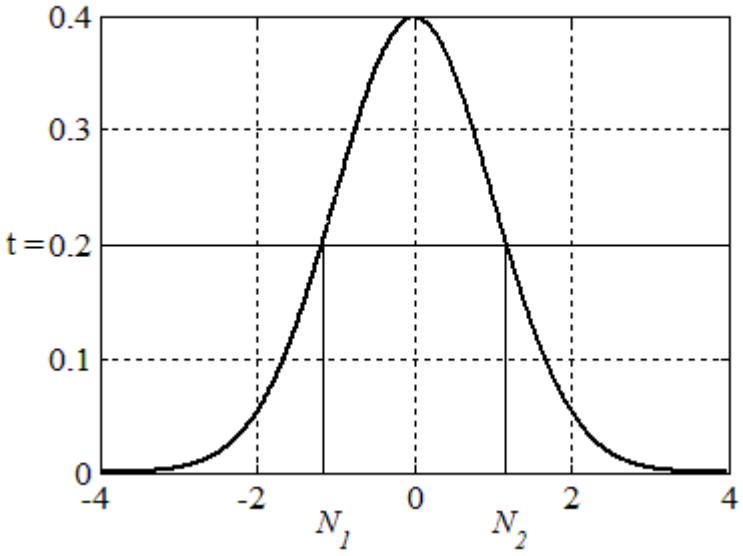

Fig. 4. Graphical interpretation of the censored distribution parameters

The censored mean square error as well as the censored variance enable the assessment of the distribution conformity for small values of the random variable. At a certain level *t*, the amount of spread of the values of the kernel estimator with reference to the normal

distribution curve is calculated for the interval limited by the ordinates of the normal distribution curve (Fig. 4). The use of these statistics is justified by the fact that the measure of curves fitting is dependent on the conformity for small values of the random variable. Equations (25) and (26) relate to the whole range of values of the random variable. Hence the results are biased by the curves accordance for large values of the random variable.

3.2. Results of simulations

The results of the research can be divided into three groups. The first two focus on the influence of the method of choosing the smoothing parameter and the length of the sample size on the quality of estimation. The last one presents the time consumption of the estimation algorithms.

The research was conducted on pseudorandom samples of the standardized normal distribution.

The impact of the bandwidth choosing method was carried for the samples of length 1024 using four methods described in part 2.1:

- approximate method (ROT),
- likelihood cross-validation method (LCV),
- plug-in method (HSJM),
- optimized method LCV (LOCAL)

Figure 5 presents the results of the kernel estimation for the Epanachnikov kernel, the corresponding statistics are included in Table 2. The results for the Gauss kernel are shown in Fig. 6 and in Table 3. The distribution curves representing the results for the Laplace kernel are depicted in Fig. 7 and Table 4 contains the corresponding statistic results.

The analysis of the curves depicted in the Fig. 5-7 and the results of the kernel estimations included in Tables 2-4 indicate that the best quality of estimation is obtained for methods LCV and ROT, independently of the chosen kernel. Estimates using the optimized LCV method (LOCAL) lead to the exposure of too many details, which does not mirror the true character of the distribution. It may be concluded that the LCV (LOCAL) method as well as the HSJM would be more suitable for estimates of the non-standard distributions, especially multimodal.

The impact of the sample size on the quality of estimation was analyzed for the three types of kernels using the LCV method for choosing the value of the bandwidth. The simulation results are shown in Fig. 8-10. According to expectations, the increase of the sample size leads to the minimization of the mean square error and the variance. It should be stated that a sample size larger than 1024 does not improve the quality of the estimation. Knowledge of the results for small sample sizes is crucial for the construction of the real-time classification algorithms for the multivariate data [2].

The evaluation of the time consumption of the procedures was realized for the two methods: ROT and LCV for which the best results of estimations were obtained. Table 8 contains the mean computing times for the PC computer with a CPU frequency of 2×1.73GHz. According to expectations, the fastest method is ROT, practically the time of computation is independent from the type of kernel. The computation times for the LCV method are significantly longer with exception for the combination of this method with the Epanachnikov kernel. The double increase of the sample size leads to a double increase of the computation time for the ROT method and a quadruple increase for the LCV method in combination with the Epanachnikov kernel.

4. Concluding remarks

The choice of kernel is determined by the subsequent factors: the numerical complexity, differentiability, integrability. Hence, in kernel estimate applications the normal kernel is generally used and the inaccuracy is compensated with the sample size increase. The advantages of the normal kernel are as following: low numerical complexity and the possibility to determine the differential of any order, which is of great importance for the determination of other characteristics of the distribution, *e.g.* the quantile or the cumulative distribution function.

The simulation results indicate that the approximate method of estimation – ROT is as well accurate as the LCV method for the estimates of unimodal distributions. Moreover, the CPU power load for this method is at a very low level, thus the quality of the estimate can be significantly improved by increasing the sample size.

The study presented in this paper is a part of the work of the authors on applying the kernel estimation in the exploratory data analysis of the measuring signals observed in the underwater environment. There is no trivial solution for the choice of the optimal kernel estimator parameters for the specific application. The presented results of numerical experiments are useful for users who face the problem of the kernel function choice and the adjustment of its parameters.

References

[1] B.W. Silverman: *Density estimation for statistics and data analysis*, Chapman and Hall, London, 1986.

[2] W.J. Krzanowski, T.C. Bailey, T. Spatinas, K.J. Powell: "Signal detection in underwater sound using wavelets". *Journal of the American Statistical Association*, vol. 93, no.441, 1998, pp. 73-82.

[3] W. Zhang, L. Wenyin, K. Zhang: "Symbol recognition with Kernel Density Matching". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, 2006, pp. 2020-2024.

[4] T. Parag, A.Elgammal, A.Mittal: "A Framework for Feature Selection for Background Subtraction". *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* vol. 2, 2006, pp. 1916-1923.

[5] P. Kulczycki: *Kernel estimators in system analysis*, WNT, Warsaw, 2005. (in Polish)

[6] S.J. Sheather: "Density Estimation". *Statistical Science*, vol. 19, no. 4, 2004, pp. 588-597.

[7] D.W. Scott: *Multivariate Density Estimation. Theory, Practice and Visualization*, Wiley, New York 1992.

[8] C. R. Loader: "Bandwidth selection: classical or plug-in?" *The Annals of Statistics*, vol. 27, no. 2, 1999, pp. 415-438.

[9] P. Hall, S.J. Sheather, M.C. Jones, J.S. Marron: "On optimal data-based bandwidth selection in kernel density estimation", *Biometrika*, vol. 78, no. 2, 1991, pp. 263-269.

Figures and Tables

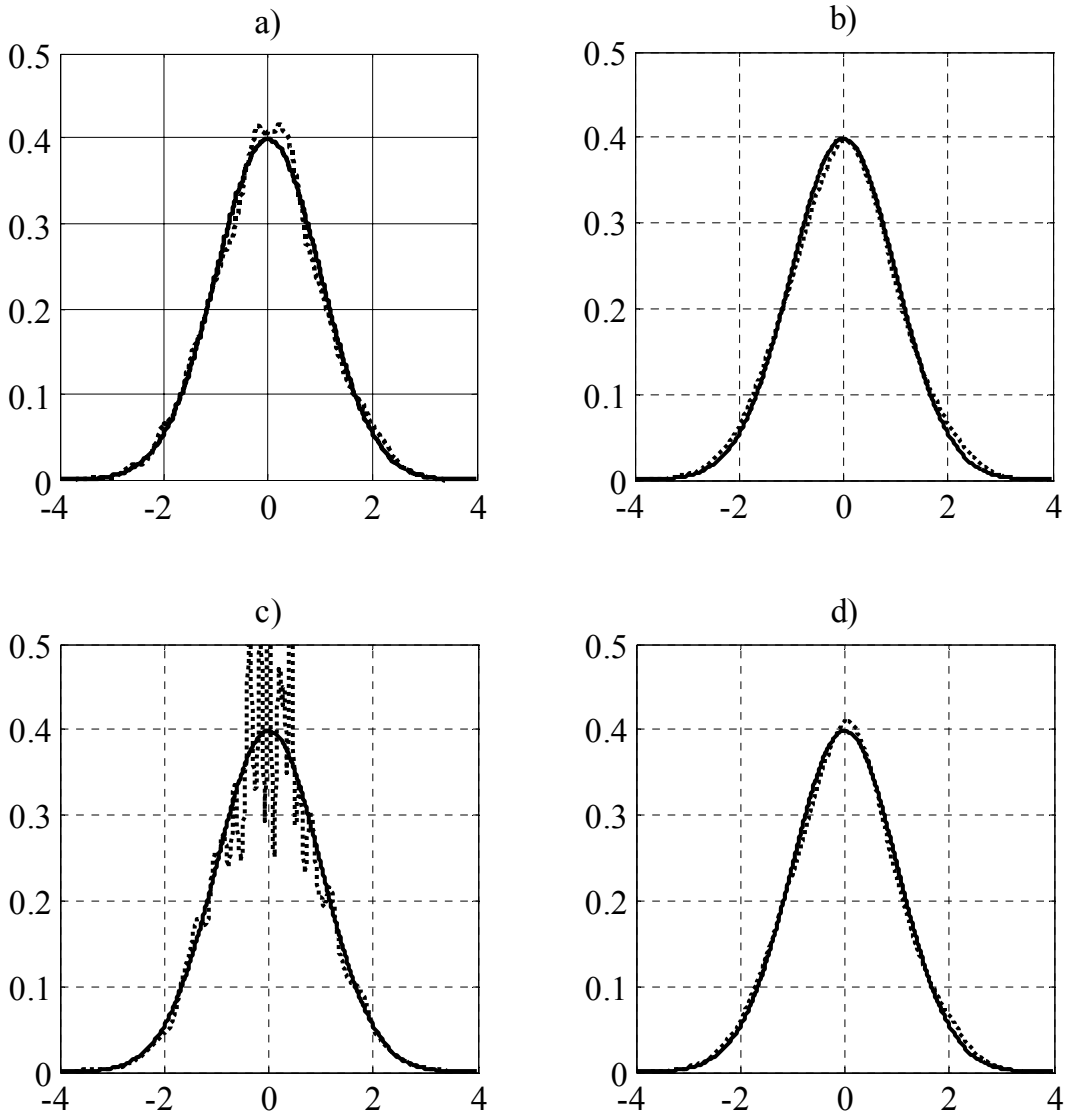

Fig. 5. The curves of the theoretical normal distribution and the kernel estimate for the Epanachnikov kernel
a) HSJM; b) LCV; c) LOCAL; d) ROT.

Table 2. Indicators of curve accordance for the Epanachnikov kernel.

	HSJM	LCV	LOCAL	ROT
MSE	8.28×10^{-4}	8.29×10^{-4}	0.0020	7.91×10^{-4}
VAR	6.00×10^{-4}	6.03×10^{-4}	0.0018	5.63×10^{-4}
$MSE^{o}_{t=0.2}$	0.0013	8.64×10^{-4}	0.0054	9.57×10^{-4}
$VAR^{o}_{t=0.2}$	0.0012	8.65×10^{-4}	0.0052	9.57×10^{-4}

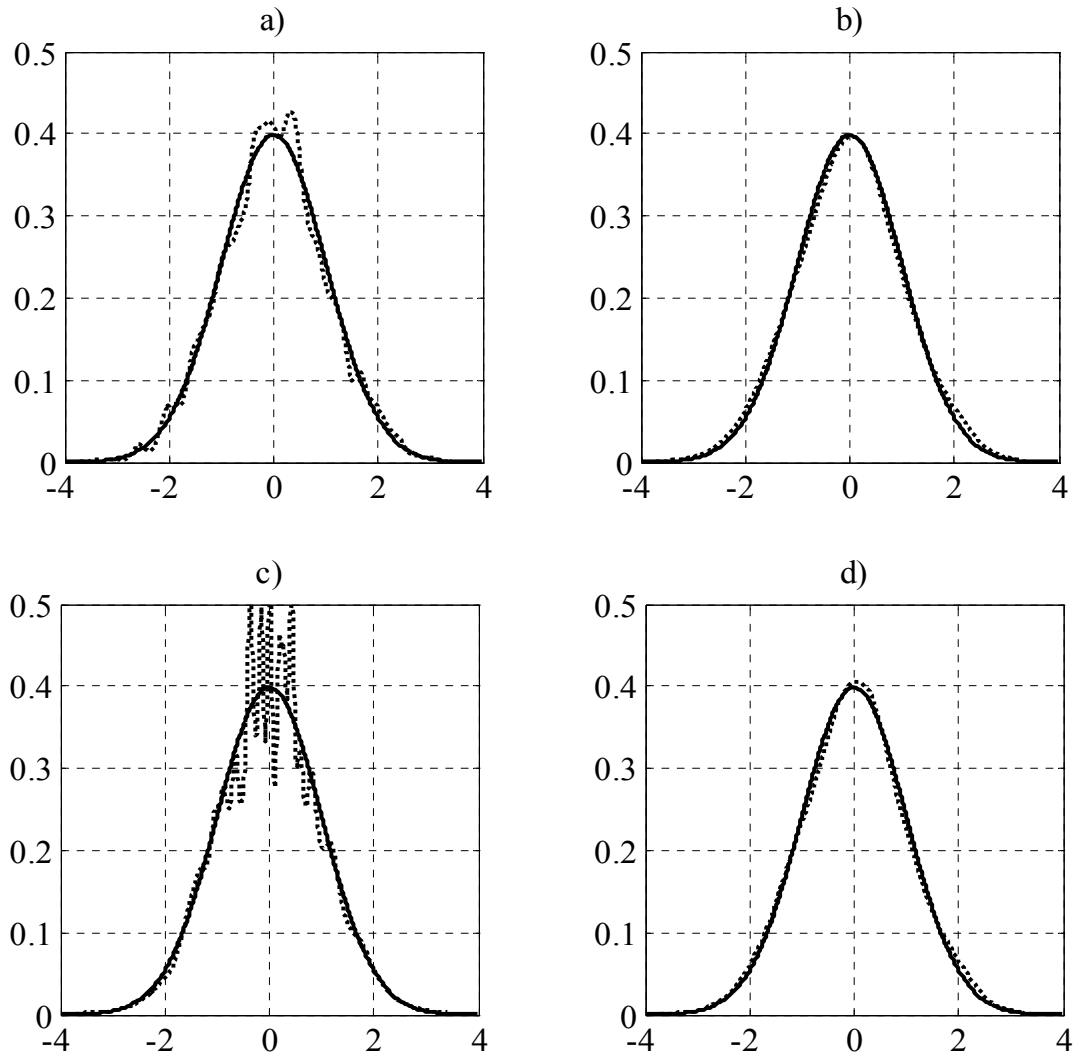

Fig. 6. The curves of the theoretical normal distribution and the kernel estimate for the Gauss kernel a) HSJM; b) LCV; c) LOCAL; d) ROT.

Table 3. Indicators of curve accordance for the Gauss kernel

	HSJM	LCV	LOCAL	ROT
MSE	9.17×10^{-4}	8.21×10^{-4}	0.0017	7.93×10^{-4}
VAR	6.89×10^{-4}	5.95×10^{-4}	0.0014	5.66×10^{-4}
$MSE^{o}_{p=0.2}$	0.0015	8.69×10^{-4}	0.0042	9.85×10^{-4}
$VAR^{o}_{p=0.2}$	0.0014	8.68×10^{-4}	0.0040	9.51×10^{-4}

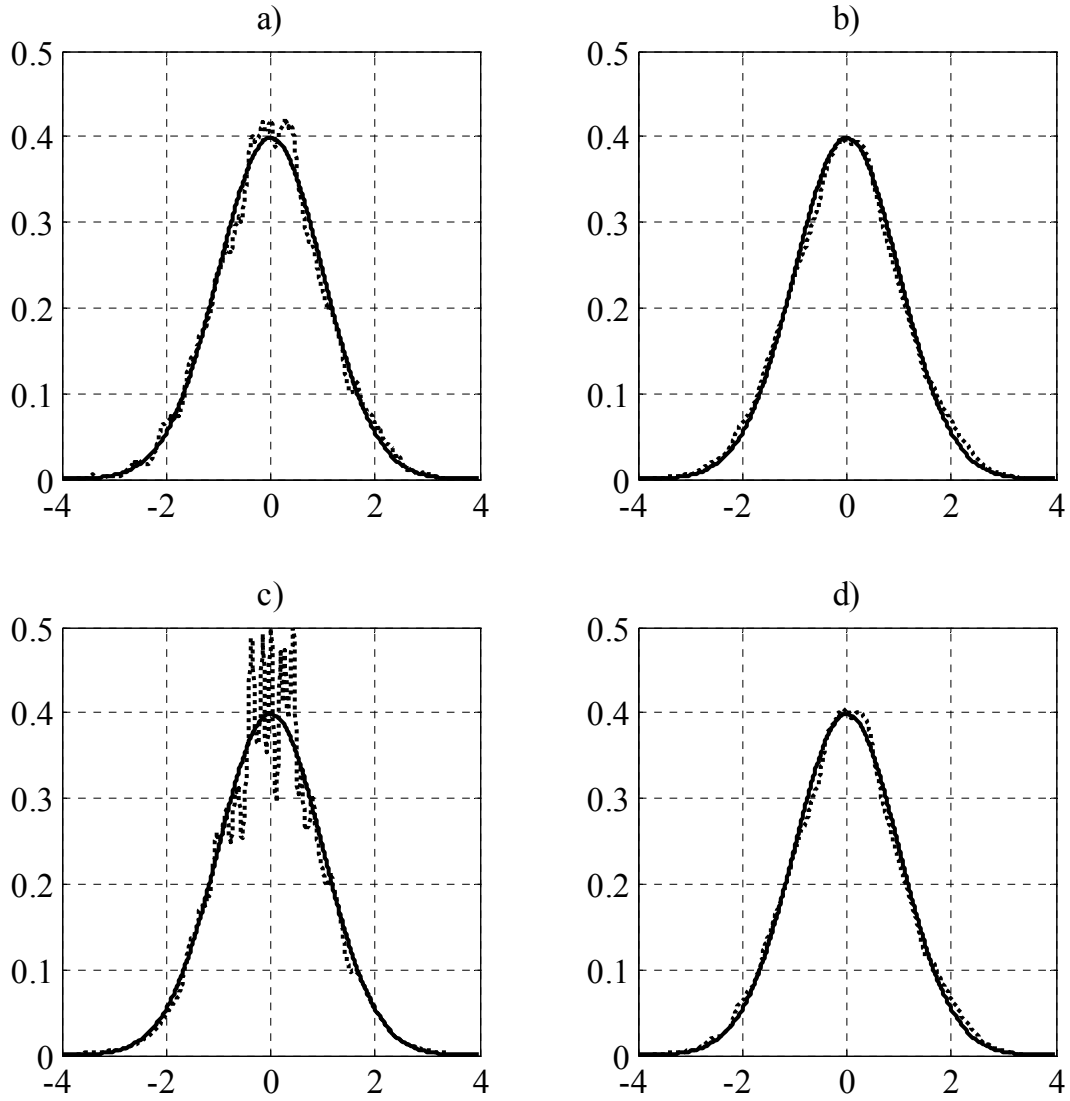

Fig. 7. The curves of the theoretical normal distribution and the kernel estimate for the Laplace kernel a) HSJM; b) LCV; c) LOCAL; d) ROT.

Table 4. Indicators of curve accordance for the Laplace kernel.

	HSJM	LCV	LOCAL	ROT
MSE	8.79×10^{-4}	8.26×10^{-4}	0.0015	8.14×10^{-4}
VAR	6.51×10^{-4}	6.00×10^{-4}	0.0012	5.87×10^{-4}
$MSE^{o}_{t=0.2}$	0.0014	9.16×10^{-4}	0.0037	0.0010
$VAR^{o}_{t=0.2}$	0.0013	9.13×10^{-4}	0.0034	9.89×10^{-4}

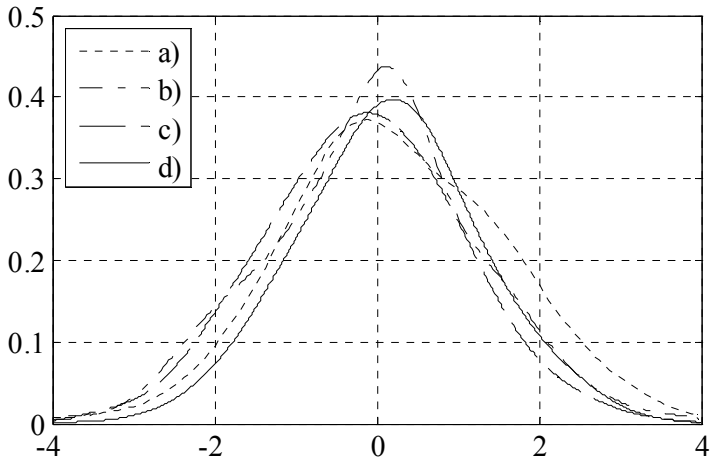

Fig. 8. The curves of the estimates using the Epanachnikov kernel; sample sizes:
a) 128; b) 256; c) 512; d) 1024.

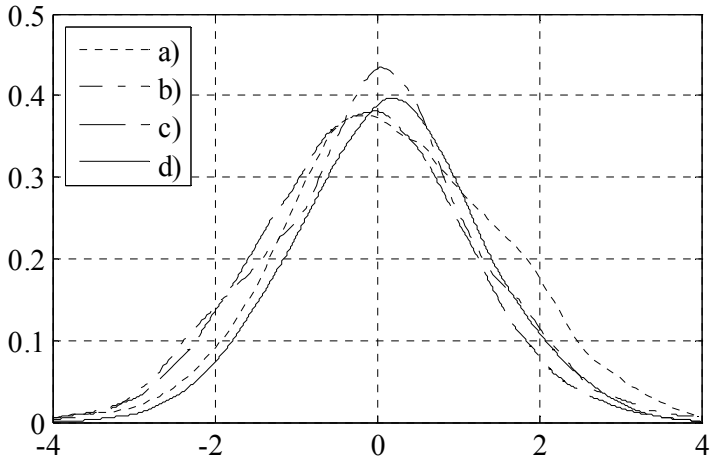

Fig. 9. The curves of the estimates using the Gauss kernel; sample sizes:
a) 128; b) 256; c) 512; d) 1024.

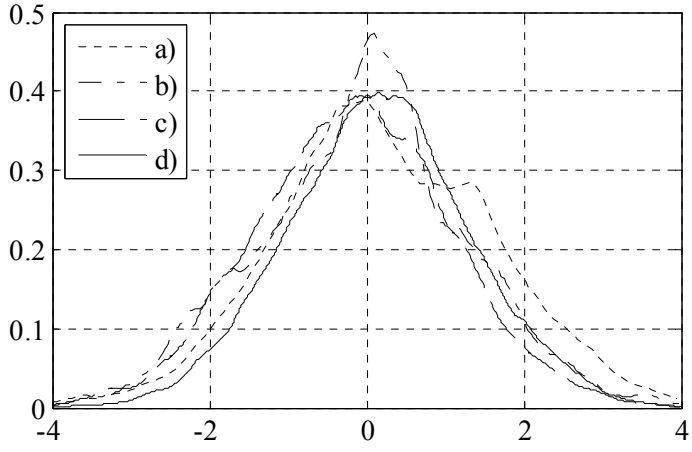

Fig. 10. The curves of the estimates using the Laplace kernel; sample sizes:
a) 128; b) 256; c) 512; d) 1024.

Table 5. The comparison of the curve accordance indicators for the Epanachnikov kernel.

Sample size 128				
	HSJM	LCV	LOCAL	ROT
MSE	0.0030	0.0027	0.0096	0.0027
VAR	0.0018	0.0015	0.0085	0.0015
$MSE^{o}_{t=0.2}$	0.0013	6.65×10^{-4}	0.0140	6.27×10^{-4}
$VAR^{o}_{t=0.2}$	0.0011	6.53×10^{-4}	0.0143	5.98×10^{-4}
Sample size 256				
	HSJM	LCV	LOCAL	ROT
MSE	0.0025	0.0017	0.0056	0.0017
VAR	0.0014	6.58×10^{-4}	0.0047	6.81×10^{-4}
$MSE^{o}_{t=0.2}$	0.0030	5.45×10^{-4}	0.0153	6.51×10^{-4}
$VAR^{o}_{t=0.2}$	0.0026	2.72×10^{-4}	0.0146	3.56×10^{-4}
Sample size 512				
	HSJM	LCV	LOCAL	ROT
MSE	0.0013	0.0014	0.0029	0.0013
VAR	7.03×10^{-4}	8.48×10^{-4}	0.0023	7.65×10^{-4}
$MSE^{o}_{t=0.2}$	7.07×10^{-4}	6.35×10^{-4}	0.0053	6.28×10^{-4}
$VAR^{o}_{t=0.2}$	4.82×10^{-4}	6.20×10^{-4}	0.0050	5.49×10^{-4}
Sample size 1024				
	HSJM	LCV	LOCAL	ROT
MSE	8.28×10^{-4}	8.29×10^{-4}	0.0020	7.91×10^{-4}
VAR	6.00×10^{-4}	6.03×10^{-4}	0.0018	5.63×10^{-4}
$MSE^{o}_{t=0.2}$	0.0013	8.64×10^{-4}	0.0054	9.57×10^{-4}
$VAR^{o}_{t=0.2}$	0.0012	8.65×10^{-4}	0.0052	9.57×10^{-4}
Sample size 2048				
	HSJM	LCV	LOCAL	ROT
MSE	4.74×10^{-4}	4.51×10^{-4}	0.0016	4.54×10^{-4}
VAR	3.56×10^{-4}	3.33×10^{-4}	0.0015	3.36×10^{-4}
$MSE^{o}_{t=0.2}$	7.08×10^{-4}	6.39×10^{-4}	0.0042	6.33×10^{-4}
$VAR^{o}_{t=0.2}$	6.38×10^{-4}	6.09×10^{-4}	0.0041	6.11×10^{-4}
Sample size 4096				
	HSJM	LCV	LOCAL	ROT
MSE	4.77×10^{-4}	4.17×10^{-4}	0.0011	4.25×10^{-4}
VAR	4.45×10^{-4}	3.86×10^{-4}	0.0011	3.93×10^{-4}
$MSE^{o}_{t=0.2}$	0.0010	9.16×10^{-4}	0.0029	9.26×10^{-4}
$VAR^{o}_{t=0.2}$	9.43×10^{-4}	7.53×10^{-4}	0.0029	7.86×10^{-4}

Table 6. The comparison of the curve accordance indicators for the Gauss kernel.

Sample size 128				
	HSJM	LCV	LOCAL	ROT
MSE	0.0038	0.0028	0.0080	0.0027
VAR	0.0026	0.0017	0.0069	0.0016
$MSE^{o}_{t=0.2}$	0.0027	6.83×10^{-4}	0.0101	6.09×10^{-4}
$VAR^{o}_{t=0.2}$	0.0026	7.00×10^{-4}	0.0102	5.97×10^{-4}
Sample size 256				
	HSJM	LCV	LOCAL	ROT
MSE	0.0033	0.0018	0.0051	0.0018
VAR	0.0022	6.91×10^{-4}	0.0041	7.44×10^{-4}
$MSE^{o}_{t=0.2}$	0.0054	5.36×10^{-4}	0.0137	8.07×10^{-4}
$VAR^{o}_{t=0.2}$	0.0049	3.19×10^{-4}	0.0130	5.35×10^{-4}
Sample size 512				
	HSJM	LCV	LOCAL	ROT
MSE	0.0015	0.0014	0.0025	0.0013
VAR	9.08×10^{-4}	8.52×10^{-4}	0.0020	7.63×10^{-4}
$MSE^{o}_{t=0.2}$	0.0012	6.46×10^{-4}	0.0046	6.21×10^{-4}
$VAR^{o}_{t=0.2}$	9.44×10^{-4}	6.34×10^{-4}	0.0042	5.45×10^{-4}
Sample size 1024				
	HSJM	LCV	LOCAL	ROT
MSE	9.17×10^{-4}	8.21×10^{-4}	0.0017	7.93×10^{-4}
VAR	6.89×10^{-4}	5.95×10^{-4}	0.0014	5.66×10^{-4}
$MSE^{o}_{t=0.2}$	0.0015	8.69×10^{-4}	0.0042	9.85×10^{-4}
$VAR^{o}_{t=0.2}$	0.0014	8.68×10^{-4}	0.0040	9.51×10^{-4}
Sample size 2048				
	HSJM	LCV	LOCAL	ROT
MSE	5.39×10^{-4}	4.54×10^{-4}	0.0013	4.56×10^{-4}
VAR	4.21×10^{-4}	3.35×10^{-4}	0.0012	3.38×10^{-4}
$MSE^{o}_{t=0.2}$	8.25×10^{-4}	6.35×10^{-4}	0.0032	6.24×10^{-4}
$VAR^{o}_{t=0.2}$	7.46×10^{-4}	6.04×10^{-4}	0.0031	6.04×10^{-4}
Sample size 4096				
	HSJM	LCV	LOCAL	ROT
MSE	5.18×10^{-4}	4.17×10^{-4}	9.29×10^{-4}	4.26×10^{-4}
VAR	4.86×10^{-4}	3.85×10^{-4}	8.97×10^{-4}	3.94×10^{-4}
$MSE^{o}_{t=0.2}$	0.0011	9.13×10^{-4}	0.0024	9.25×10^{-4}
$VAR^{o}_{t=0.2}$	0.0010	7.47×10^{-4}	0.0023	7.85×10^{-4}

Table 7. The comparison of the curve accordance indicators for the Laplace kernel.

Sample size 128				
	HSJM	LCV	LOCAL	ROT
MSE	0.0034	0.0030	0.0063	0.0030
VAR	0.0022	0.0018	0.0053	0.0018
$MSE^{o}_{t=0.2}$	0.0017	7.06×10^{-4}	0.0086	8.09×10^{-4}
$VAR^{o}_{t=0.2}$	0.0016	7.02×10^{-4}	0.0085	8.27×10^{-4}
Sample size 256				
	HSJM	LCV	LOCAL	ROT
MSE	0.0028	0.0020	0.0049	0.0020
VAR	0.0017	9.53×10^{-4}	0.0039	9.75×10^{-4}
$MSE^{o}_{t=0.2}$	0.0039	0.0013	0.0130	0.0014
$VAR^{o}_{t=0.2}$	0.0035	0.0011	0.0122	0.0011
Sample size 512				
	HSJM	LCV	LOCAL	ROT
MSE	0.0014	0.0014	0.0020	0.0013
VAR	8.48×10^{-4}	8.29×10^{-4}	0.0015	8.01×10^{-4}
$MSE^{o}_{t=0.2}$	9.64×10^{-4}	6.38×10^{-4}	0.0037	6.56×10^{-4}
$VAR^{o}_{t=0.2}$	7.92×10^{-4}	6.12×10^{-4}	0.0032	5.99×10^{-4}
Sample size 1024				
	HSJM	LCV	LOCAL	ROT
MSE	8.79×10^{-4}	8.26×10^{-4}	0.0015	8.14×10^{-4}
VAR	6.51×10^{-4}	6.00×10^{-4}	0.0012	5.87×10^{-4}
$MSE^{o}_{t=0.2}$	0.0014	9.16×10^{-4}	0.0037	0.0010
$VAR^{o}_{t=0.2}$	0.0013	9.13×10^{-4}	0.0034	9.89×10^{-4}
Sample size 2048				
	HSJM	LCV	LOCAL	ROT
MSE	5.15×10^{-4}	4.73×10^{-4}	0.0012	4.72×10^{-4}
VAR	3.97×10^{-4}	3.55×10^{-4}	0.0011	3.54×10^{-4}
$MSE^{o}_{t=0.2}$	7.78×10^{-4}	6.66×10^{-4}	0.0029	6.33×10^{-4}
$VAR^{o}_{t=0.2}$	7.13×10^{-4}	6.31×10^{-4}	0.0028	6.17×10^{-4}
Sample size 4096				
	HSJM	LCV	LOCAL	ROT
MSE	4.93×10^{-4}	4.23×10^{-4}	8.15×10^{-4}	4.33×10^{-4}
VAR	4.61×10^{-4}	3.92×10^{-4}	7.81×10^{-4}	4.00×10^{-4}
$MSE^{o}_{t=0.2}$	0.0011	9.25×10^{-4}	0.0020	9.37×10^{-4}
$VAR^{o}_{t=0.2}$	9.77×10^{-4}	7.58×10^{-4}	0.0020	7.95×10^{-4}

Table 8. Mean times of the estimates [s].

128		
	LCV	ROT
Epanechnikov kernel	0.0128	0.0010
Gauss kernel	0.0558	0.0018
Laplace kernel	0.0635	0.0010
256		
	LCV	ROT
Epanechnikov kernel	0.0327	0.0013
Gauss kernel	0.2041	0.0023
Laplace kernel	0.2392	0.0013
512		
	LCV	ROT
Epanechnikov kernel	0.1075	0.0018
Gauss kernel	0.8059	0.0027
Laplace kernel	0.9209	0.0019
1024		
	LCV	ROT
Epanechnikov kernel	0.4091	0.0032
Gauss kernel	3.0264	0.0049
Laplace kernel	3.5363	0.0032
2048		
	LCV	ROT
Epanechnikov kernel	1.3222	0.0062
Gauss kernel	12.6249	0.0071
Laplace kernel	13.6975	0.0062
4096		
	LCV	ROT
Epanechnikov kernel	5.2299	0.0127
Gauss kernel	48.7009	0.0134
Laplace kernel	0.0131	0.0131