On Local Bandwidth Selection for Density Estimation

Rimantas RUDZKIS, Mindaugas KAVALIAUSKAS

Statistics Center of Vytautas Magnus University and Institute of Mathematics and Informatics
Vileikos 8, 3000 Kaunas, Lithuania
e-mail: rudzkis@ktl.mii.lt

Abstract. This article gives ideas for developing statistics software which can work without user intervention. Some popular methods of bandwidth selection for kernel density estimation (the nearest neighbour, least squares cross-validation, "plug-in" technique) are discussed. Modifications of the cross-validation criterion are proposed. Two-stage estimators combining these methods with multiplicative bias correction are investigated by simulation means.

Key words: kernel density estimation, local bandwidth selection, cross-validation, multiplicative bias correction.

1. Introduction

This work is related to statistics software developing problems of nonparametric distribution density estimating. Procedures where estimation parameters are calculated from the sample and not chosen by user becomes more and more popular. The computation of such procedures is longer but it is not important because of fast computers characteristics improvement. The priority is facilitation of user work. Besides that, such software can use people without special mathematical education. This article is related with paper by Jakimauskas (1997) from this point of view. Kernel density estimator is perhaps the most popular nonparametric density estimation method. This method is widely described in books by Prakasa Rao (1983), Silverman (1986) and Devroye and Györfi (1985). Although this method has been used for several decades, attention of theorists and practicians is still focused on it. New modifications of well known procedures have been proposed and investigated seeking for universal and adaptive estimator. A term "second generation methods" is even used in the survey by Jones *et al.* (1996). A bandwidth selection method with good asymptotic efficiency in the class of all densities in meaning of L_1 error is proposed in article by Devroye and Lugosi (1995). An interesting bias reduction procedure is investigated in paper by Jones *et al.* (1995). This procedure is easy to compute and analyze. Locally adaptive bandwidth function selection based on popular cross-validation method and splines is discussed in paper by Fan *et al.* (1996).

This article discusses nonparametric distribution density estimation in the case where the unknown density is a mixture of densities with different smoothness. An example of

such densities is a Gaussian mixture with different variance component model in clustering analysis discussed in paper by Rudzkis and Radavičius (1995). It's expedient to use local smoothing in this situation. Several density estimation procedures combining variuos ideas presented below are discussed. Simulation results are given.

2. Methods and Motivations

Let $X_1, X_2, \ldots$ be i.i.d. with the distribution density function $f(x)$, and $\mathbf{X} = (X_1, \ldots, X_n)$ a sample. The kernel density estimator $\widehat{f}_h$ is defined as

$$\widehat{f}_h(x) = n^{-1} \sum_{t=1}^{n} K_h(X_t - x), \tag{1}$$

where $K_h(y) = \frac{1}{h} K\left(\frac{y}{h}\right)$.

In the expression above, $h = h(x, \mathbf{X})$ is the bandwidth, $K(y)$ is the kernel function satisfying the condition $\int K(y)dy = 1$. We assume that K is a symmetric, non-negative function supported on $[-1, \cdot 1]$.

The main problem in practical density estimation is selection of h. Usually h selection is based on the analysis of smoothness of f. Here we discuss some h selection ideas. We use a term "parameter h" if the bandwidth is independent of x, and "function h" if $h = h(x)$.

2.1. *Asymptotic Analysis and "Plug-in" Technique*

Recall some well known formulas when h is independent of the sample, i.e., $h = h(x, n)$. The bias of estimator (1) is

$$b_h(x) = \mathbf{E}\widehat{f}_h(x) - f(x) = \int [f(x + hy) - f(x)]K(y)dy, \tag{2}$$

and the variance

$$\sigma_h^2(x) = \mathbf{D}\widehat{f}_h(x) = \frac{1}{nh} \int K^2(y)f(x + hy)dy - \frac{\left(\mathbf{E}\widehat{f}_h(x)\right)^2}{n}. \tag{3}$$

Hereafter, the integration interval is $(-1, \cdot 1)$ if not specified otherwise. Usually the bandwidth h is selected so that it minimizes the mean square error. Since $\mathbf{E}(\widehat{f}_h(x) - f(x))^2 = b_h^2(x) + \sigma_h^2(x)$, denote

$$h_{opt}(x) = \arg\min_h \left(b_h^2(x) + \sigma_h^2(x)\right). \tag{4}$$

If the sample size n is large, then it is natural to analyze asymptotic expressions besides (4). Let $n \to \infty$, and the bandwidth $h = h(x, n)$ satisfies well known consistency conditions

$$h \to 0, \quad nh \to \infty. \tag{5}$$

Suppose that f is continuous. Then (3) could be written

$$\sigma_h^2(x) = \frac{c_2 f(x)}{nh} + o\left(\frac{1}{nh}\right), \quad c_2 = \|K\|_2^2, \tag{3'}$$

and if the second density derivative f'' is continuous, then

$$b_h(x) = \frac{c_1 f''(x) h^2}{2} + o(h^2), \quad c_1 = \int y^2 K(y) dy. \tag{2'}$$

We denote by $\|\cdot\|_p$ a norm in the function space L_p. If $f(x) \neq 0$ and $f''(x) \neq 0$, then, from (2') and (3'), we obtain an expression of an asymptotically optimal bandwidth function h

$$h_{AS}(x) = \left(\frac{c_2 f(x)}{c_1^2 (f''(x))^2 n}\right)^{1/5}, \tag{6}$$

and optimal bandwidth parameter h

$$h_{AS} = \left(\frac{c_2}{c_1^2 \|f''\|_2^2 n}\right)^{1/5}. \tag{7}$$

The popular usage of expression (7) in statistical density estimation is "plug-in" estimators. The parameter h is calculated by plugging an estimate of the unknown $\|f''\|_2$ in expression (7). If the kernel K has the second derivative, f'' estimate is obtained analogously to (1)

$$\widehat{f}_\Delta''(x) = n^{-1} \sum_{t=1}^{n} K_\Delta''(X_t - x), \tag{8}$$

where the bandwidth Δ is usually selected to be larger than the expected h_{AS} value. Survey (Jones *et al.*, 1996) proposes $\Delta = \Delta(h)$ selection method based on f''' statistical analysis and approximation of the density f by a parametric family of densities. It's then recommended to solve the equation $h^5 = \dfrac{c_2}{c_1^2 \|f_{\Delta h}''\|_2^2 n}$. Paper by Jones *et al.* (1996) gives very good results in favour of this method. Unfortunately, when the smoothing bandwidth selected locally, i.e., using (6) instead of (7), the method proposed in paper Jones *et al.* (1996) gives seemingly good practical results only in the case of very large sample sizes as rather accurate estimator of the third derivative $f'''(x)$ is needed.

To our mind, there is no need of using asymptotic expression (2') when selecting the kernel bandwidth $h(x)$. It's better to select the bandwidth $h(x)$ by estimating the bias

$b_h(x)$ directly. The "plug-in" estimator studied in this paper is obtained from expression (4) after substitution of $b_h(x)$ and $\sigma_h^2(x)$ by their estimates

$$\widehat{\sigma}_h^2(x) = \frac{c_2 \widehat{f_h}(x)}{nh} \tag{9}$$

and

$$\widehat{b}_{h,\Delta}(x) = \int \left[\widehat{f}_\Delta(x + hy) - \widehat{f}_\Delta(x) \right] K(y) dy, \quad \Delta = \Delta(h). \tag{10}$$

Thus we define the "plug-in" bandwidth function $\widetilde{h}_{PI}(x)$ as follows.

$$\widetilde{h}_{PI}(x) = \arg\min_h \left[\widehat{\sigma}_h^2(x) + \widehat{b}_{h,\Delta(h)}^2(x) \right]. \tag{11}$$

Calculation of the function $\Delta(h)$ is based on asymptotical analysis. Under condition (5), estimation of $b_h(x)$ for large n is equivalent to estimation of f''. It's recommended to use a larger bandwidth for estimation of the density derivatives than for estimation of the density itself (see, e.g., Jones *et al.*, 1996), i.e., $\Delta(h) > h$. In this case, asymptotical expressions are readily obtained

$$\mathbf{D}\widehat{b}_{h,\Delta}(x) = \frac{f(x) c_1^2 c_3}{4} \frac{h^4}{n\Delta^5} + o\left(\frac{h^4}{n\Delta^5}\right), \quad c_3 = \|K''\|_2^2, \tag{12}$$

and

$$\mathbf{E}\widehat{b}_{h,\Delta}(x) = b_n(x) + o(h^2), \quad \text{if } \Delta \to 0. \tag{13}$$

Expressions (12) and (13) are derived by analogy with (3') and (2') (Silverman, 1986). It's important that

$$\mathbf{D}\widehat{b}_{h,\Delta}(x) \leqslant b_h^2(x). \tag{14}$$

In case $h(x)$ is close to $h_{AS}(x)$, the condition

$$\Delta(h) \geqslant c_4 h, \quad c_4 = \left(\frac{c_3 c_1^2}{c_2}\right)^{1/5} \tag{17}$$

follows from (2'), (6), (12) and (14). We analyzed the density estimator defined by (1) and (11) in the case where

$$\Delta(h) = \alpha c_4 h. \tag{18}$$

This estimator depends on the parameter α that can be calculated using cross-validation method descussed below.

2.2. *Cross-Validation Method*

Cross-validation is a very popular method used to choose parameters of statistical estimators. The estimation performance criterion is the integrated mean square error.

An estimate $\widehat{f}$ is optimal if it minimizes

$$Q(\widehat{f}) \to \min, \quad \text{where} \quad Q(\widehat{f}) = \mathbf{E}\|\widehat{f} - f\|_2^2 - \|f\|_2^2. \tag{19}$$

Let F be the distribution function of a random variable X. Then

$$Q(\widehat{f}) = \|\widehat{f}\|_2^2 - 2 \int \widehat{f}(x) dF(x). \tag{20}$$

An estimate of the functional Q is obtained from (20) after substituting F by the empirical distribution function

$$\widehat{Q}(\widehat{f}) = \|\widehat{f}\|_2^2 - \frac{2}{n} \sum_{t=1}^{n} \widehat{f}(X_t|t), \tag{21}$$

where $\widehat{f}(x|t)$ denotes the value of the estimate calculated using the sample data with X_t removed. Such a modification reduces the bias of the estimate $\widehat{Q}$. The essence of cross-validation method is to choose the density estimator parameters in such a way that the requirement

$$\widehat{Q}(\widehat{f}) \to \min \tag{22}$$

is fulfilled. Note that theorists recommend to use the smoothed empirical distribution function to estimate the functional Q (smoothed cross-validation). Empirical research shows that the largest local minimizer gives better performance than the global minimizer (Jones *et al.*, 1996).

Selection of the bandwidth $h(x)$ using the cross-validation idea has been investigated in paper by Fan *et al.* (1996). This paper proposes the following algorithm:

1) fix argument x grid $x_1, \ldots, x_p$;

2) the function $h(x)$ is defined as the third-order spline with the nodes $(h_1, x_1), \ldots,$ (h_p, x_p), i.e., we have $h(x) = h(x, h_1, \ldots, h_p)$;

3) $h_1, \ldots, h_p$ are chosen to minimize $\widehat{Q}(\widehat{f})$.

We denote by $\widetilde{h}_{CV(x)}$ the bandwidth $h(x)$ calculated using this algorithm. If $p = 1$ we have the parameter h_{CV} calculated using the cross-validation approach.

2.3. *The Nearest Neighbour Method*

The nearest neighbour method is one of the oldest methods of local bandwidth $h(x)$ selection.

An integer $k < n$ is chosen (for example $k \cong n^{2/3}$). We denote $\rho_k(x) = \min\left\{a : \sum_{t=1}^{n} \mathbf{1}_{\{|X_t - x| \leqslant a\}} = k\right\}$, i.e., $\rho_k(x)$ is the distance from x to the kth nearest sample element. The bandwidth is defined as follows

$$\widetilde{h}_{NN}(x) = \rho_k(x). \tag{23}$$

The relative error of the estimate $\widehat{f}_{\widetilde{h}_{NN}}(x)$ defined by (1) and (23) is $\delta(x) = \dfrac{\widehat{f}(x) - f(x)}{f(x)}$. It is independent of the scale parameter, and its variance is asymptotically independent of x, $\mathbf{D}\delta(x) = \dfrac{c_2}{k} + o\left(\dfrac{1}{k}\right)$. The bandwidth function $\widetilde{h}_{NN}(x)$ defined by (23) and the estimate $\widehat{f}_{\widetilde{h}_{NN}}(x)$ are unsmooth, and their construction don't take into consideration local smoothness of f. The nearest neighbour estimate is fully defined by the parameter k which can be chosen using cross-validation method.

3. Modifications of the Methods

3.1. *The Modification of Cross-Validation Method*

As mentioned above, the risk function in cross-validation method is $\mathbf{E}\|\widehat{f} - f\|_2^2$. Unfortunately, a value of this functional has linear dependency on the scale parameter. This criterion suggests to construct an estimate of the density in such a way that the relative errors would be small in the region where the density values are large. For example, if we have a mixture of the Gaussian densities with different means and variances $\sigma_1^2 \ll \sigma_2^2$, then usage of this criterion gives much larger relative errors of estimation of the first component in comparison with those of the second component. The criterion based on the mean losses in space L_1 $\mathbf{E}\|\widehat{f} - f\|_1$ does not have this drawback, but statistical estimation of a value of such losses causes problems. A good way out of this situation is to define an error $Q^*(\widehat{f}) = \mathbf{E}\left\|\dfrac{\widehat{f} - f}{\sqrt{f}}\right\|_2^2 - 1$ instead of (19). Note that a value of the functional Q^* is independent of scale. If the density is defined parametrically $f(x) \in \{f(x, \theta),\ \theta \in \Theta\}$ then a pseudoestimate $\widehat{\theta} = \arg\min_{\theta} Q^*(f(\cdot, \theta))$ of the parameter θ is asymptotically equivalent to the maximum likelihood estimate (under natural conditions). It is an important argument to use the functional Q^*. It is easy to calculate a statistical estimate of Q^*. Let $\widehat{f}_0$ be some fixed density estimate and let $g(x) = \widehat{f}_0(x) + \dfrac{\varepsilon}{\sqrt{n}\ \max\limits_{i,j=\overline{1,n}} |X_i - X_j|}$, where $\varepsilon \geqslant 0$ is a parameter. Then

$$\widehat{Q}^*(\widehat{f}) = \int_{-\infty}^{+\infty} \frac{\widehat{f}^2(x)}{g(x)}\,dx - \frac{2}{n}\sum_{t=1}^{n} \frac{\widehat{f}(X_t|t)}{g(X_t)}. \tag{24}$$

We define the modified cross-validation (MCV) method analogously as in (22) by substituting of $\widehat{Q}$ by $\widehat{Q}^*$.

3.2. *"Plug-in" and the Nearest Neighbour Methods Modifications*

The bandwidth functions $\widetilde{h}_{PI}(x)$ and $\widetilde{h}_{NN}(x)$ are not sufficiently smooth, thus it is natural to use additional smoothing, by treating $\widetilde{h}(x)$ as noisy regression functions. The simplest way is to use the same kernel as in formula (1)

$$\overline{h}_{PI}(x) = \sum_{t=1}^{n} \widetilde{h}_{PI}(X_t) K_h(X_t - x) \Big/ \sum_{t=1}^{n} K_h(X_t - x), \quad h = \widetilde{h}_{PI}(x). \tag{25}$$

We define $\overline{h}_{NN}(x)$ analogously. We recomend to use the MCV method to calculate the parameters α and k used in the definition of $\overline{h}_{PI}$ and $\overline{h}_{NN}$.

Simulation results show that better performance is obtained using not the estimate (10) of the bias b_h but its modification

$$\left|\widehat{b}_{h,\Delta}(x)\right| = \int_0^1 \left|\widehat{f}_\Delta(x + hy) - \widehat{f}_\Delta(x - hy) + 2\widehat{f}_\Delta(x)\right| K(y) dy \tag{25'}$$

in expression (11). When the sample size n increases, estimates (10) and (25') are asymptotically equivalent. Statistics (25'), however, is more stable.

3.3. *Multiplicative Bias Reduction*

This bias reduction method is described in paper Jones *et al.* (1995). Let $\widehat{f}(x)$ be an estimate defined by (1) and $h(x)$ a bandwidth function. We denote

$$\widehat{f}^*(x) = \frac{\widehat{f}(x)}{n} \sum_{t=1}^{n} \frac{K_h(X_t - x)}{\widehat{f}(X_t)}. \tag{26}$$

We define $\widehat{f}_{PI}^*(x)$ and $\widehat{f}_{NN}^*(x)$ using (26) with $\widehat{f} = \widehat{f}_{\overline{h}_{PI}}$ and $\widehat{f} = \widehat{f}_{\overline{h}_{NN}}$ respectively.

Note that the bias correction usually increases the variance of estimate. Empirical analysis shows that modification (26) decreases smoothness of estimate. It's natural to use an extra smoothing procedure

$$\widehat{\widehat{f}^*}(x) = \sum_{t=1}^{n} K_{\dot{h}}(X_t - x)\widehat{f}^*(X_t) \Big/ \sum_{t=1}^{n} K_{\dot{h}}(X_t - x), \tag{27}$$

where $\dot{h}(x) = \beta h(x)$. The parameter β could be calculated using MCV method or simple taken $\beta = 1$.

4. Simulation Results

The "plug-in" and the nearest neighbour estimators and parameter selection using cross-validation and MCV has been explored. The density estimate modifications defined by

(25)–(27) have also been studied. We give preliminary results only. Comprehensive research results including analysis of the estimate modifications proposed in paper by Fan *et al.* (1995) will be published in a sequel of this paper.

We give the concrete simulation description. Estimation accuracy was tested in the metrics of the L_1 space $\Delta_1(\widehat{f}) = \|\widehat{f} - f\|_1$. We give errors in the metrics of the L_2 space $\Delta_2(\widehat{f}) = \|\widehat{f} - f\|_2$, too. The sample $X_1, \ldots, X_n$ has the distribution density

$$f(x) = p\varphi_1(x) + (1-p)\varphi_2(x), \tag{28}$$

where φ_i is the density of Gaussian distribution $\mathbf{N}(m_i, \cdot\sigma_i)$. Values of parameters σ_i were chosen significantly different, because only in these cases where smoothness properties are dependent on x values it is expedient to use estimates with the bandwidth dependent on x. We used the Parzen kernel

$$K(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & |u| < 1/2, \\ 2(1 - |u|)^3, & 1/2 \leqslant |u| < 1, \\ 0, & |u| \geqslant 1, \end{cases} \tag{29}$$

which possesses two derivatives.

The "plug-in" and the nearest neighbour estimates were also compared with the pseudoestimate calculated using the theoretically optimal bandwidth function $h_{opt}(x)$ defined by (9)–(11) with the density estimates substituted by true density. They were also compared with the cross-validation estimate with the bandwidth h_{CV} which gives good results in the space L_2 metrics. The errors of these estimates are given in Table 1.

The results show that the nonmodified "plug-in" estimator gives smaller errors in metrics L_1 than the cross-validation and the nearest neighbour method. Cross-validation gives better results in metrics L_2. Investigations show that the nearest neighbour method is inaccurate in the edge of the density where it gives too large value of the bandwidth. Note that adaptive estimates give better results in comparison with the fixed bandwidth estimates in the case where component variances σ_i^2 are significantly different.

It's important that a density estimate is not only accurate in the meaning of any metrics, but also gives a good visual result. Therefore the estimates were compared visually, too. We present the figures of the mentioned estimates which were calculated using the following mixture: $p = 0.5$, $\varphi_1 = N(0, \cdot 1)$, $\varphi_2 = N(20, \cdot 6)$.

The figures show that the modified "plug-in" method restores the density shape much better than other methods. Fig.1 shows that the fixed bandwidth estimate adapts to the cluster with the smaller variance, but gives an inaccurate estimate of the other cluster. As mentioned above, modifications (25)–(27) were analyzed also. Monte-Carlo simulations showed that the additional bandwidth smoothing procedure defined by (25) decreases the error of both the nearest neighbour (by 2–10%) and "plug-in" estimates (by 10–20%). Multiplicative bias reduction was the second procedure applied to density estimators. This modification decreases the "plug-in" estimate error up to 20%, but it is unfit for the nearest neighbour estimate and increases error up to 1.5 times. The reason of this is unsmoothness of the nearest neighbour estimate (see Fig. 4). Multiplicative bias correction makes the estimate even more unsmooth. Therefore such a correction is unsuitable

Table 1

Simulation results

Metrics	Theoretical pseudo-estimate	Fixed bandwidth cross-validation estimate	The nearest neighbour estimate	"Plug-in" estimate	Modified "plug-in" estimate
	$p = 0.5,\quad \varphi_1 = N(0, \cdot 1),\quad \varphi_2 = N(15, \cdot 3),\quad n = 500$				
L_1	0.08971	0.12208	0.14253	0.12674	0.05953
L_2	0.02580	0.03387	0.05075	0.03950	0.01831
	$p = 0.5,\quad \varphi_1 = N(0, \cdot 1),\quad \varphi_2 = N(20, \cdot 6),\quad n = 500$				
L_1	0.09655	0.15690	0.15515	0.13882	0.06042
L_2	0.02556	0.03613	0.05305	0.04414	0.01776
	$p = 0.5,\quad \varphi_1 = N(0, \cdot 1),\quad \varphi_2 = N(35, \cdot 10),\quad n = 500$				
L_1	0.09049	0.19022	0.16643	0.13259	0.06343
L_2	0.02288	0.03587	0.04263	0.03941	0.01706
	$p = 0.5,\quad \varphi_1 = N(0, \cdot 1),\quad \varphi_2 = N(20, \cdot 6),\quad n = 2000$				
L_1	0.054113	0.09006	0.11379	0.08827	0.05096
L_2	0.01319	0.01940	0.02657	0.02523	0.01650

for them. Smoothing of the "plug-in" estimate defined by expression (27) was applied taking into consideration the decrease of smoothness that had been noticed in "plug-in" estimate after multiplicative bias reduction. $\widehat{f}$ smoothing procedure decreased the error of both the nearest neighbour and "plug-in" methods, but this decrease didn't compensate the increase of the error after multiplicative bias correction for the nearest neighbour estimate.

After all these modifications the nearest neighbour method was insignificantly improved in meaning of L_1 error but became worse in meaning of L_2. All modifications decreased the error of the "plug-in" estimate. Errors became close to the theoretical pseu-
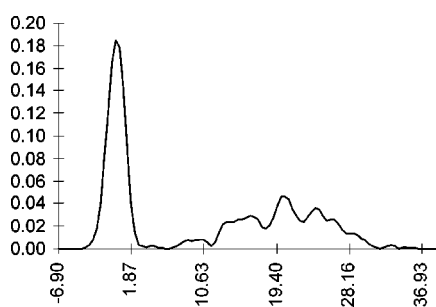

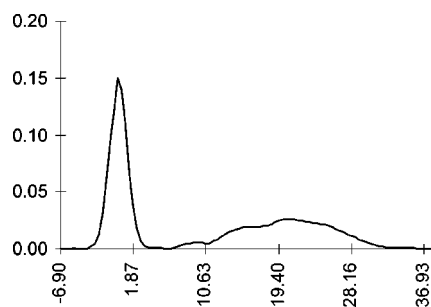Fig. 1. Cross-validation fixed bandwidth distribution density estimate.
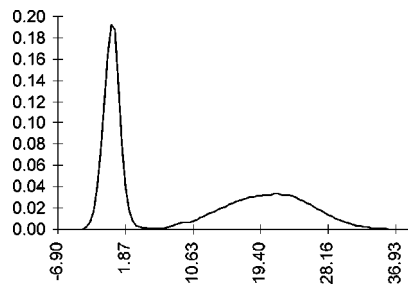
Fig. 2. "Plug-in" estimate.

Fig. 3. "Plug-in" estimate after smoothing and multiplicative bias correction.
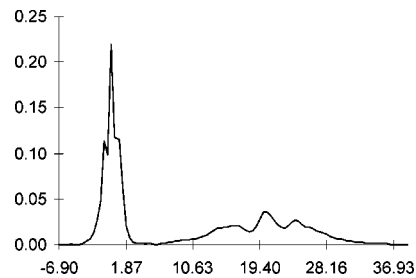

Fig. 4. The nearest neighbours estimate.

doestimate errors.

From the practical point of view it is very important to compare the methods with respect to the computation time. The nearest neighbour method didn't give good results but this method is very fast. The "plug-in" method is 20–100 times slower. This ratio of the computation time increases if the sample size grows. The fixed bandwidth cross-validation method needs less time than the "plug-in" estimate, but when the sample size is large it becomes much slower than the "plug-in" method.

Some estimate aspects that need improvement were noticed during the simulation tests. First of all, it is the choice of the parameters k and α. The tests showed that cross-validation method didn't give an optimal k value. A result of this is increase of the error of the nearest neighbour method by 5–20%. MCV method didn't yield an optimal value, too. We used the optimal value of the "plug-in" method parameter α equal to 1, because the cross-validation and MCV methods required a lot of computing time and gave the value close to 1. The constraint problem has been encountered during the numerical calculations. It's necessary to have an interval $[h_{\min}(x), \cdot h_{\max}(x)]$ containing an optimal bandwidth value. The "plug-in" estimate becomes unstable if we use too wide interval. This interval was chosen using the results of the nearest neighbour method $h_{\min}(x) = \underline{c}\widetilde{h}_{NN}(x)$, $h_{\max}(x) = \overline{c}\widetilde{h}_{NN}(x)$.

5. Conclusions

The nearest neighbour method is fast and can be used to get a preliminary estimate. This method, however, needs a modification to give an more accurate results in the edge of the density and to adapt to local smoothness of the density.

The "plug-in" estimate and its modifications are accurate density estimates. All the modifications improved estimation performance in the cases investigated by the authors. The estimate is to be used in such problems where accurate probability density estimation is very important.

References

Devroye, L., and L. Györfi (1985). *Nonparametric Density Estimation: The L_1 View*. Wiley, New York.

Devroye, L., and G. Lugosi (1996). A universal acceptable smoothing factor for kernel density estimates. *The Annals of Statistics*, **24**(6), 2499–2512.

Fan, J., P. Hall, M.A. Martin and P. Patil (1996). On local smoothing of nonparametric curve estimators. *Journal of the American Statistical Association*, **91**(433), 258–266.

Jakimauskas, G. (1997). Efficient analysis of one estimation and clusterization procedure of one-dimentional Gaussian mixture. *Informatica*, **8**(3), 331–343.

Jones, M.C., O. Linton and J.P. Nielsen (1995). A Simple bias reduction method for density estimation. *Biometrica*, **82**(2), 327–338.

Jones, M.C., J.S. Marron and S.J. Sheather (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, **91**(433), 401–407.

Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*. Academic Press: New York.

Rudzkis, R., and M. Radavičius (1995). Statistical estimation of a mixture of Gaussian disributions. *Acta Applicandae Mathematicae*, **38**, 37–54.

Silverman, B.W. (1986). *Density Estimation for Statistics Data Analysis*. Chapman and Hall, London.

R. Rudzkis was born in Kaunas, in 1951. In 1973 he graduated from Kaunas Polytechnic Institute (speciality of computers). In 1978 he maintained a doctoral degree in mathematics at Vilnius University and in 1993 a degree of doctor habilius in mathematics at Institute of Mathematics and Informatics. He is a professor, expert-member of the Lithuanian Academy of Sciences. Currently he is head of Statistics Center of Vytautas Magnus University and Institute of Mathematics and Informatics. Also, he is head of the Applied Statistics department at Institute of Mathematics and Informatics. R. Rudzkis fields of scientific research interest are theoretical and applied statistics, econometrics.

M. Kavaliauskas, born in 1975, in 1997 graduated from Kaunas Tecnology University, Faculty of Fundamental Sciences, continues the studies to acquire a master's degree. Since 1996 he has also been working at the Statistics Center of Vytautas Magnus University and Institute of Mathematics and Informatics, as an engineer.

Apie lokalų glodinimo pločio parinkimą, vertinant pasiskirstymo tankį

Rimantas RUDZKIS, Mindaugas KAVALIAUSKAS

Darbe nagrinėjami pasiskirstymo tankio branduoliniai statistiniai įverčiai. Aptariami keli populiarūs branduolio pločio lokalaus parinkimo būdai (artimiausių kaimynų, "cross-validation", "plug-in"). Pasiūlytos šių metodų modifikacijos, papildomai panaudojančios pločio funkcijos įverčio glodinimą bei tankio įverčio poslinkio multiplikatyvų mažinimą. Pateikiami palyginamųjų tyrimų Monte-Karlo metodu rezultatai.