

KERNEL DENSITY ESTIMATORS FOR GAUSSIAN MIXTURE MODELS

Tomas Ruzgas¹, Indrė Drulytė²

Kaunas University of Technology, Department of Applied Mathematics

Address: 50 Studentų St., LT-51368 Kaunas, Lithuania

E-mail: ¹tomas.ruzgas@ktu.lt, ²drulyte.indre@inbox.lt

Received: March 2013

Revised: October 2013

Published: November 2013

Abstract. The problem of nonparametric estimation of probability density function is considered. The performance of kernel estimators based on various common kernels and a new kernel K (see (14)) with both fixed and adaptive smoothing bandwidth is compared in terms of the symmetric mean absolute percentage error using the Monte Carlo method. The kernel K is everywhere positive but has lighter tails than the Gaussian density. Gaussian mixture models from a collection introduced by Marron and Wand (1992) are taken for Monte Carlo simulations. The adaptive kernel method outperforms the smoothing with a fixed bandwidth in the majority of models. The kernel K shows better performance for Gaussian mixtures with considerably overlapping components and multiple peaks (double claw distribution).

Keywords: nonparametric density estimators, kernel method, Monte Carlo method, adaptive smoothing, symmetric mean absolute percentage error.

1. Introduction

Nowadays there are a lot of methods for estimating the probability density function; however, one might face problems when choosing an effective estimation procedure and applying it in practice if the probability density of sampled data is multimodal and the sample size is small or moderate. Kernel smoothing is the most frequently used nonparametric estimation method (see [18], [20], [24]). Thus far, there is no generally accepted kernel estimation method which outperforms the other in all cases. Although many adaptive selection procedures have been proposed [2], [17], [28], their efficiency has not been well established yet, especially for samples of a moderate size.

The applications of the Gaussian mixture model (GMM) are often encountered in various fields. For example, in medical research, to classify the mobility of limbs using myoelectric signals, Yonghong, Englehart, Hudgins, Chan [27] compared GMM with other three classifiers: linear discriminant analysis, linear perceptrons network, and multilayer perceptrons neural network. In the said paper, the investigation made by using GMM showed an extreme accuracy of the classification. Such researchers as Henthorn, Benitez, Avram, Martinez, Llerena, Cobaleda, Krejcie and Gibbons in their investigation [14] analysed the ability to hydroxylate (the degree of urinary metabolic) the debrisoquine and the dextromethorphan of O-demethylate with the GMM assistance. Zimmerman, Povinelli, Johnson and Ropella [29] applied the Gaussian mixture model for ECG analysis by using ST segments and T waves to determine the ischemic disease. In the geological researches of Stachniss, Plagemann, Lilienthal [25], the Gaussian mixture model was used for the prediction of gas concentration based on drillings. Moreover, the approximation of GMM was used in ecological investigations to study the contamination of soil. This idea was implemented by Gruszczyński [10]. In an applied sociology paper, Chernova and Veloso [4] described the interactive policy learning results by using Gaussian mixture models as well.

In this paper, the Monte-Carlo method was used to compare the performance of various kernel estimators both with fixed and adaptive smoothing parameters. The following four kernel functions were applied: 1) Gaussian distribution density, 2) Epanechnikov kernel (12), 3) triweight kernel (centred beta (4,4) distribution density, see (13)), and 4) a new kernel K , obtained as a simple transformation of Gaussian distribution density (14). The kernel K is everywhere positive but has lighter tails than Gaussian (see (14)). Gaussian mixture models from a collection introduced by Marron and Wand (1992) are taken for Monte Carlo simulations. The accuracy of estimators is measured by the symmetric mean absolute percentage error.

The paper consists of 5 sections. Section 2 reviews the kernel density estimators and kernel functions; Section 3 describes the simulation experiment; Section 4 contains the simulation results. The detailed results of the study can be found in the appendix and supplementary material. Concluding remarks are presented in Section 5.

2. Kernel estimator

A random variable $X \in \mathbf{R}$ satisfies a mixture model if its distribution density function $f(x)$ is given by the equality:

$$f(x) = \sum_{k=1}^q p_k f_k(x) = f(x, \theta). \quad (1)$$

The parameter q is the number of components in the mixture. The component weights p_k are called *a priori* probabilities and satisfy conditions:

$$p_k > 0, \quad \sum_{k=1}^q p_k = 1. \quad (2)$$

Function $f_k(x)$ is the distribution density function of the k^{th} component and θ is the vector of parameters of mixture model (1). Suppose a simple sample $\mathbf{X} = (X(1), \dots, X(n))$ of size n from X is given. The estimation of the distribution density of an observed random variable is one of the main statistical tasks.

Histogram is one of the simplest and the oldest density estimators. As far as it is known, for the first time the data were shown as a bar chart (without graphic visualization) in the context of the *mortality histogram* constructed by John Graunt (1662) from the Bills of Mortality during the plague years [26]. The circumstantial evidence that Graunt actually invented the histogram while looking at these mortality data seems quite strong, although there is a reason to infer that Galileo had used histogram-like diagrams earlier. Hald [11] recounts a portion of Galileo's *Dialogo*, published in 1632, in which Galileo summarized his observations on the star that appeared in 1572. This graphical representation was first introduced by Karl Pearson [23] in 1891. For the approximation of density $f(x)$, the number of observations $X(t)$ falling within the range of Ω is calculated and divided by n and the volume of area Ω .

Shaping the histogram, every $X(t)$ should be imagined as a separate bar with a height equal to $1/n$. Then, logically, the midpoint of the bar can be changed to the $X(t)$ and obtain a function:

$$\hat{f}(x) = \frac{1}{n \cdot h} \sum_{t=1}^n I_{C_h(X(t))}(x), \quad (3)$$

where C_h is a bin with the midpoint $X(t)$ and the width h . This statistic is often named as a simple estimator. In fact, we can use a smooth “bump”, known as a kernel function, instead of the indicator function. Thus, the fixed-width kernel density estimator (FKDE) $\hat{f}(x)$ with a kernel function K and a fixed (global) bandwidth h is derived:

$$\hat{f}(x) = \frac{1}{nh} \sum_{t=1}^n K\left(\frac{x - X(t)}{h}\right). \quad (4)$$

It is one of the most widespread non-parametric density estimators in data analysis (see e.g. [5], [16], [22]). The kernel function $K(u)$ should satisfy the condition:

$$\int_{\mathbf{R}} K(u) du = 1. \quad (5)$$

Usually, but not always, $K(u)$ will be a symmetric probability density function [24].

At first, the data is usually prescaled in order to avoid large differences in data spread. A natural approach [8] is first to standardize the data by a linear transformation yielding data with zero mean and unit variance. As a result, equation (4) is applied to the standardized data. Let Z denote standardized values of random X :

$$Z = S^{-1/2}(X - \bar{X}), \quad (6)$$

where \bar{X} is the empirical mean, and S is the empirical variance. Applying the FKDE method to the standardized data $\mathbf{Z} = (Z(1), \dots, Z(n))$ yields the following estimator of density function $f(x)$:

$$\hat{f}_Z(z) = \frac{1}{nh} \sum_{t=1}^n K\left(\frac{z - Z(t)}{h}\right), \quad (7)$$

$$\hat{f}(x) = \frac{S^{-1/2}}{nh} \sum_{t=1}^n K\left(S^{-1/2} \frac{x - X(t)}{h}\right). \quad (8)$$

The optimal FKDE bandwidth h^* is determined through the minimization of the mean integrated squared error (MISE) [24]. For example, when both the kernel and the distribution of observations is Gaussian with the unit variance, h^* can be found by Silverman's [24] equation:

$$h^* = An^{-\frac{1}{5}}, \text{ where } A = (4/3)^{\frac{1}{5}}. \quad (9)$$

More advanced methods for determining the kernel bandwidth, such as the least-square cross-validation method, are also available with respectively increasing complexity and computation resources [3], [7], [13]. In practice, the bandwidth is often selected by testing. If the value of h is too small, then the density estimator is irregular and has a lot of modes. In the ultimate case, the modes are located at each of the observations. The greater values of h mean greater smoothing of an estimator.

Although FKDE is a widely used method of nonparametric density estimation, it has some practical flaws [24]; for example, the inability to deal satisfactorily with the tails of distributions without oversmoothing the main part of the density. A significant improvement of FKDE is an adaptive kernel density estimator (AKDE) [24]. The construction of AKDE looks like FKDE; however, now the kernel bandwidth h varies together with the argument x . This method includes two stages: computing an adaptive bandwidth and the density estimation by kernel method. The algorithm is as follows:

Step 1. The elements of a sample $\mathbf{X} = (X(1), \dots, X(n))$ are standardized producing data $\mathbf{Z} = (Z(1), \dots, Z(n))$ with the sample mean $\hat{\mathbf{E}}[Z] = \mathbf{0}$ and the sample variance $\hat{\mathbf{E}}[Z^2] = 1$.

Step 2. Calculate the pilot FKDE (7) estimator $\tilde{f}_Z(z)$ satisfying the condition $\tilde{f}_Z(Z(t)) > 0, \forall t$.

Step 3. Find the values of a local width $\lambda_t = (\tilde{f}_Z(Z(t))/g)^{-\gamma}$; here g is the geometric mean of $\tilde{f}_Z(z)$, i.e., $\log g = \frac{1}{n} \sum_{i=1}^n \log \tilde{f}_Z(Z(t))$, and γ is the sensitivity parameter, $0 \leq \gamma \leq 1$.

Step 4. Obtain the adaptive kernel estimator $\hat{f}_Z(z)$ by

$$\hat{f}_Z(z) = \frac{1}{n} \sum_{t=1}^n h^{-1} \lambda_t^{-1} K\left(\frac{z - Z(t)}{h \lambda_t}\right), \quad (10)$$

where h is the same global smoothing parameter as used in equation (4). The larger the γ , the more sensitive performance to the selection of kernel. The typical value is $\gamma = \frac{1}{2}$ [1], [24]. This recommended value will be used in the further analysis.

The comparative analysis of estimation accuracy was made for four different types of kernels. The first three kernels are classical, while the last one is new.

The Gaussian kernel consistent with the distribution of normal $\varphi(x)$ [9], [19] selection:

$$K(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \quad (11)$$

The Epanechnikov kernel is the second order polynomial, corrected to satisfy the properties of the density function [6]:

$$K(x) = \frac{3}{4}(1-x^2)\mathbf{1}_{\{|x|\leq 1\}}. \quad (12)$$

The triweight kernel proposed by Tapia and Thompson [26] has better smoothness properties and finite support. Its detailed investigation was done by Hall [12]:

$$K(x) = \frac{35}{32}(1-x^2)^3\mathbf{1}_{\{|x|\leq 1\}}. \quad (13)$$

The kernel K is everywhere positive, but has lighter tails than Gaussian distribution density and was introduced by authors of this article:

$$K(x) = \varphi(x(c+|x|)^q) \cdot ((c+|x|)^q + |x|^q(c+|x|)^{q-1}), \quad c > 0, \quad q > 0. \quad (14)$$

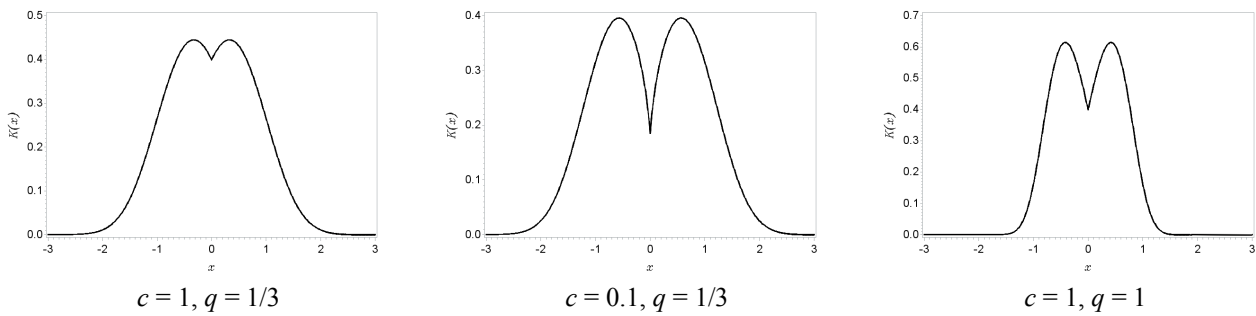


Fig. 1. Form of kernel K function dependence on parameters

This kernel function depends on parameters c and q . In simulations, the chosen values of the parameters c and q were, respectively, 0.01, 0.1, 1 and 1/3, 1/2, 1. The first two and the last two values in the first and the second collection, respectively, produce worse accuracy results. Therefore, only the results for $c = 1$ and $q = 1/3$ are reported here. Detailed results are presented in the supplementary material [21].

3. Monte Carlo experiment

A comprehensive simulation study was conducted with the aim to compare kernel methods of statistical density estimation and the kernel functions described before. In the comparative analysis, Gaussian mixture models from a collection suggested by J. S. Marron and M. P. Wand [20] were used as the target densities:

- 1) Gaussian $N(0, 1)$;
- 2) skewed unimodal $\frac{1}{5}N(0, 1) + \frac{1}{5}N(\frac{1}{2}, (\frac{2}{3})^2) + \frac{3}{5}N(\frac{13}{12}, (\frac{5}{9})^2)$;
- 3) strongly skewed $\sum_{l=0}^7 \frac{1}{8}N(3\{(\frac{2}{3})^l - 1\}, (\frac{2}{3})^{2l})$;
- 4) kurtotic unimodal $\frac{2}{3}N(0, 1) + \frac{1}{3}N(0, (\frac{1}{10})^2)$;
- 5) outlier $\frac{1}{10}N(0, 1) + \frac{9}{10}N(0, (\frac{1}{10})^2)$;

- 6) bimodal $\frac{1}{2}N(-1, (\frac{2}{3})^2) + \frac{1}{2}N(1, (\frac{2}{3})^2)$;
- 7) separated bimodal $\frac{1}{2}N(-\frac{3}{2}, (\frac{1}{2})^2) + \frac{1}{2}N(\frac{3}{2}, (\frac{1}{2})^2)$;
- 8) skewed bimodal $\frac{3}{4}N(0, 1) + \frac{1}{4}N(\frac{3}{2}, (\frac{1}{3})^2)$;
- 9) trimodal $\frac{9}{20}N(-\frac{6}{5}, (\frac{3}{5})^2) + \frac{9}{20}N(\frac{6}{5}, (\frac{3}{5})^2) + \frac{1}{10}N(0, (\frac{1}{4})^2)$;
- 10) claw $\frac{1}{2}N(0, 1) + \sum_{l=0}^4 \frac{1}{10}N(l/2 - 1, (\frac{1}{10})^2)$;
- 11) double claw $\frac{49}{100}N(-1, (\frac{2}{3})^2) + \frac{49}{100}N(1, (\frac{2}{3})^2) + \sum_{l=0}^6 \frac{1}{350}N((l-3)/2, (\frac{1}{100})^2)$;
- 12) asymmetric claw $\frac{1}{2}N(0, 1) + \sum_{l=-2}^2 (2^{1-l}/31)N(l + \frac{1}{2}, (2^{-l}/10)^2)$;
- 13) asymmetric double claw $\sum_{l=0}^1 \frac{46}{100}N(2l-1, (\frac{2}{3})^2) + \sum_{l=1}^3 \frac{1}{100}N(-l/2, (\frac{1}{100})^2) + \sum_{l=1}^3 \frac{7}{100}N(l/2, (\frac{7}{100})^2)$;
- 14) smooth comb $\sum_{l=0}^5 (2^{5-l}/63)N(\{65 - 96(\frac{1}{2})^l\}/21, (\frac{32}{63})^2/2^{2l})$;
- 15) discrete comb $\sum_{l=0}^2 \frac{2}{7}N((12l-15)/7, (\frac{2}{7})^2) + \sum_{l=8}^{10} \frac{1}{21}N(2l/7, (\frac{1}{21})^2)$.

These fifteen densities have been carefully chosen because they thoroughly represent many different types of challenges to curve estimators [20]. The first five represent different types of problems that can arise for unimodal densities. The rest of the densities are multimodal. Densities from 6 to 9 are mildly multimodal and one might hope to be able to estimate them fairly well with a data set of a moderate size. The remaining densities are strongly multimodal and, for moderate sizes, it is difficult to recover even their shape; yet, they are well worth studying because the issue of just how much of them can be recovered is an important one.

In the simulation study, low-size and moderate-size samples (16, 32, 64, 128, 256, 512, 1024, 2048) were used. 500 replications were generated in each case. The estimation accuracy is measured by the symmetric mean absolute percentage error:

$$\text{SMAPE} = \frac{2}{n} \sum_{t=1}^n \frac{|f(X(t)) - \hat{f}(X(t))|}{|f(X(t)) + \hat{f}(X(t))|} \cong \int |f(x) - \hat{f}(x)| dx.$$

An additional bias is introduced when the kernel estimator is calculated at the sample values, i.e. for $x = X(t)$. Therefore, to eliminate this bias estimator (4) in the SMAPE, the formula is changed to:

$$\hat{f}^*(x) = \frac{1}{(n-1)h} \sum_{t=1}^n K\left(\frac{x - X(t)}{h}\right) - \frac{K(0)}{(n-1)h}. \quad (15)$$

4. Results of Monte Carlo simulations

The errors in diagrams that are obtained by fixed width kernel density estimator are represented by solid line, and the errors obtained by adaptive width kernel density estimator are represented by dashed line. Symbols G, E, T, and K mark, accordingly, the Gaussian, the Epanechnikov, the triweight and the proposed kernel functions. Full comparative analysis results are given in the tables of supplementary material [21]. The estimator of fixed kernel density in the tables is denoted FK. The estimator of adaptive kernel density is denoted AK. The results (mean and standard deviation) of the SMAPE for 500 Monte Carlo replications are presented in supplementary material. All the kernel functions applied to the fixed width and adaptive kernel density estimators used are described in Section 2. In addition to the kernel functions mentioned, other functions were investigated as well: Cauchy and Laplace distributions, cosine kernel. The application of these kernel functions results in worse accuracy in this computer experiment; therefore, the respective results are not reported. The outcomes of this research confirmed the advantage of AK over FK, which had been previously shown by other researches [15]. The Epanechnikov kernel function suits well for symmetrical distributions with clearly expressed one or several peaks. The triweight kernel function is most suitable for several or high number of

needles. However, for the complex mixture models investigated in this study, these peaks show up only for the samples of size $n \geq 128$. Therefore, the use of this kernel is not recommended for estimation in case of a small sample size. The proposed kernel function K is statistically significantly better for distributions where several components have large variances and the components are strongly overlapping. From the fifteen distributions offered by J. S. Marrton and M. P. Wand [20] (see Section 3), better results using the K kernel were obtained for two of them for samples with a sample size $n > 16$ and for four distributions when $n < 256$. As an example, the results for double claw distribution are presented in Table 1.

Table 1. Performance of density estimators for double claw distributions

Kernel	Method	Sample size							
		16	32	64	128	256	512	1024	2048
Gaussian	FK	0.30524	0.24427	0.21904	0.16844	0.13402	0.11485	0.09427	0.08076
		0.08109	<i>0.06844</i>	<i>0.04993</i>	<i>0.03855</i>	<i>0.02821</i>	<i>0.02232</i>	<i>0.01894</i>	<i>0.01391</i>
	AK	0.31247	0.24521	0.21728	0.16069	0.12231	0.09947	0.08076	0.06626
		0.08885	<i>0.07125</i>	<i>0.05641</i>	<i>0.03928</i>	<i>0.02650</i>	<i>0.01955</i>	<i>0.01625</i>	<i>0.00806</i>
Epanechnikov	FK	0.45021	0.31451	0.23839	0.18902	0.14184	0.11320	0.08684	0.07157
		<i>0.13135</i>	<i>0.07915</i>	<i>0.05941</i>	<i>0.03222</i>	<i>0.02636</i>	<i>0.02276</i>	<i>0.01649</i>	<i>0.00916</i>
	AK	0.45928	0.32957	0.24654	0.19485	0.14710	0.11496	0.08885	0.07285
		<i>0.12894</i>	<i>0.08170</i>	<i>0.06013</i>	<i>0.03343</i>	<i>0.02757</i>	<i>0.02402</i>	<i>0.01653</i>	<i>0.00868</i>
Triweight	FK	0.53563	0.38827	0.27989	0.22144	0.16453	0.12845	0.09926	0.08016
		<i>0.12367</i>	<i>0.07688</i>	<i>0.05974</i>	<i>0.03444</i>	<i>0.02693</i>	<i>0.02340</i>	<i>0.01608</i>	<i>0.00861</i>
	AK	0.55140	0.39922	0.28775	0.22777	0.17054	0.13129	0.10188	0.08199
		<i>0.11811</i>	<i>0.07937</i>	<i>0.05732</i>	<i>0.03487</i>	<i>0.02764</i>	<i>0.02492</i>	<i>0.01616</i>	<i>0.00896</i>
K	FK	0.32027	0.23969	0.20870	0.15469	0.12045	0.10119	0.08242	0.06953
		<i>0.10304</i>	0.07485	0.06261	0.03973	0.02784	0.02160	0.01841	0.01268
	AK	0.32831	0.24180	0.20826	0.15335	0.11783	0.09725	0.07881	0.06621
		<i>0.11446</i>	0.08287	0.06684	0.03694	0.02696	0.02076	0.01728	0.00751

The decreasing parameter c usually increases the SMAPE. However, its increase on average improves the SMAPE for the skewed distributions and the small sample size. The increase in the value of the parameter q has a positive effect when estimating densities with sharp peaks, i.e. when the triweight kernel function is most suitable. The Gaussian kernel was the best solution only for very smooth skewed mixtures when the size of the sample was less than 32 observations.

5. Conclusion and future work

The comparative simulation analysis has shown that the new kernel K is preferable to the well-known classical kernels, when the components of estimated mixture are considerably overlapping. The research was conducted using the Gaussian mixture models commonly used for data simulations by other researchers as they typify many different types of challenges of density estimators. The future work can address the problem how the local behaviour of the kernel in the neighbourhood of zero and its tails influence the accuracy of kernel estimators for data distributions of various shapes in moderate sample sizes.

Acknowledgements

We would like to thank the anonymous reviewer for his/her suggestions for improving this paper. We are also deeply grateful to Marijus Radavičius.

Appendix

The density estimation results are presented. Each figure corresponds to a different sample distribution. Both axes are expressed in logarithmic scale.

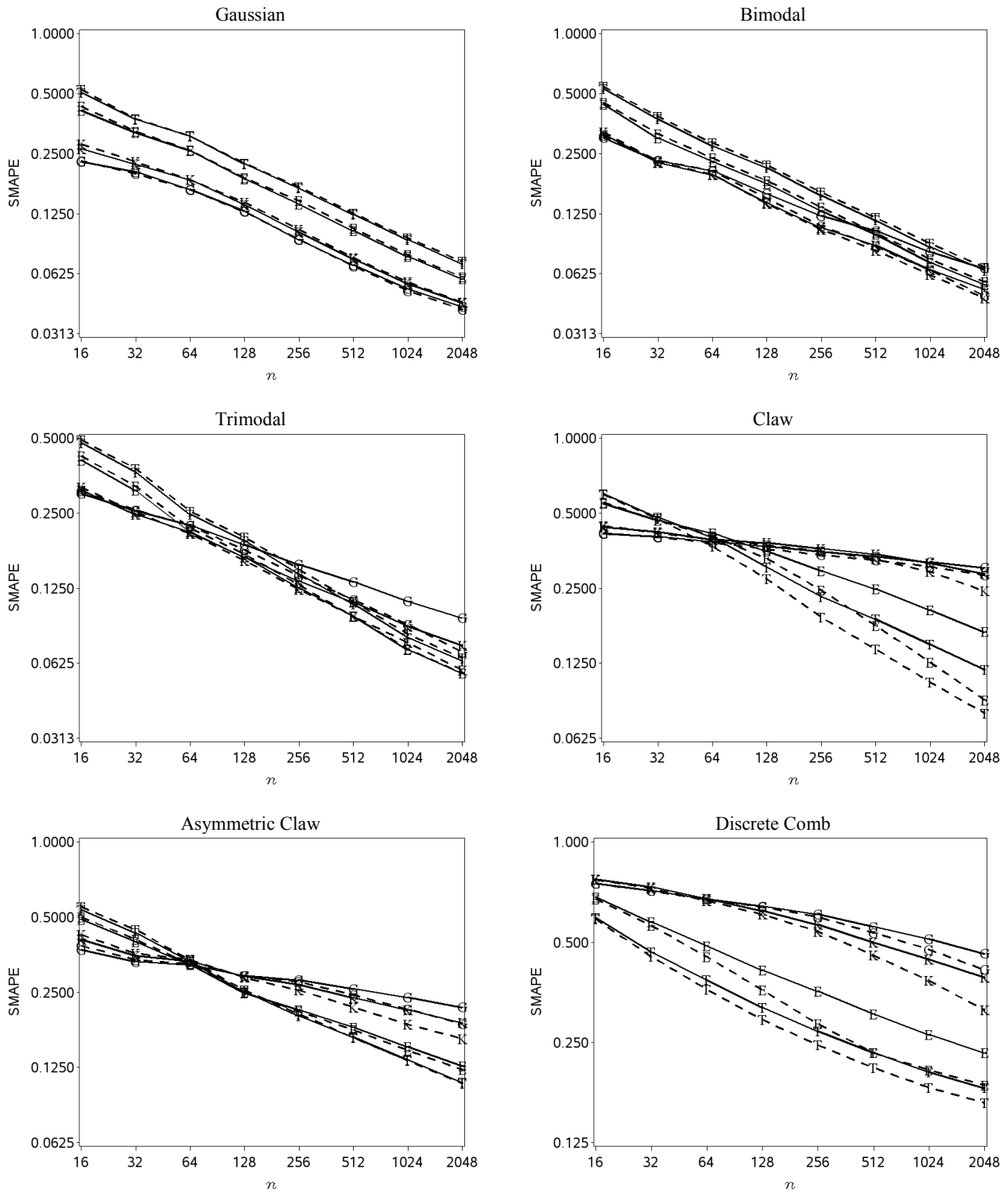


Fig. 2. Relationship between the sample size and errors

References

1. Abramson I. 1982: On Bandwidth Variation in Kernel Estimates – A Square Root Law, *The Annals of Statistics*, 10, p. 1217–1223.
2. Bashtannyk D.M., and Hyndman R.J. 1998: Bandwidth Selection for Kernel Conditional Density Estimation, *Technical report*, Department of Econometrics and Business Statistics, Monash University.
3. Breiman L., Meisel W., and Purcell E. 1977: Variable Kernel Estimates of Multivariate Densities, *Technometrics*, 19, p. 135–144.
4. Chernova S., and Veloso M. 2007: Confidence-Based Policy Learning from Demonstration Using Gaussian Mixture Models, *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA.
5. Elgammal A., Duraiswami R., Harwood D., and Davis L.S. 2002: Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance, *IEEE*, 90 (7).
6. Epanechnikov, V.A. 1969: Nonparametric Estimation of a Multivariate Probability Density, *Theory of Probability and Its Applications*, 14, p. 153–158.
7. Fan J., and Gijbels I. 1995: Data-Driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation, *Journal of the Royal Statistical Society*, B, 57, p. 371–394.
8. Fukunaga K. 1972: *Introduction to Statistical Pattern Recognition*. New York, Academic Press.
9. Gasser T., Müller, H.G., and Mammitzsch V. 1985: Kernels for Nonparametric Curve Estimation, *Journal of the Royal Statistical Society*, B, 47, p. 238–252.
10. Gruszczyński S. 2005: The Assessment of Variability of the Concentration of Chromium in Soils with the Application of Neural Networks, *Polish Journal of Environmental Studies*, Vol. 14, No. 6, p. 743–751.
11. Hald A. 1990: *A History of Probability and Statistics and their Applications Before 1750*, New York, Wiley.
12. Hall P. 1985: Kernel Estimation of a Distribution Function. *Communications in Statistics, Theory and Methods*, 14, p. 605–620.
13. Hall P., Sheather S.J., Jones M.C., and Marron J.S. 1991: On Optimal Data-Based Bandwidth Selection in Kernel Density Estimation, *Biometrika*, 78, p. 263–269.
14. Henthorn T. K., Benitez J., Avram M. J., Martinez C., Llerena A., Cobaleda J., Krejcie T. C., and Gibbons R. D. 1989: Assessment of the Debrisoquin and Dextromethorphan Phenotyping Tests by Gaussian Mixture Distributions Analysis, *Clinical Pharmacology and Therapeutics*, 45, p. 328–333.
15. Hwang J.N., Lay S.R., and Lippman A. 1994: Nonparametric Multivariate Density Estimation: A Comparative Study, *IEEE Transactions on Signal Processing*, 42, p. 2795–2810.
16. Jones M.C. 1989: Discretized and Interpolated Kernel Density Estimates, *Journal of the American Statistical Association*, 84, p. 733–741.
17. Jones M.C. 1992: Potential for Automatic Bandwidth Choice in Variations of Kernel Density Estimation, *Statistics and Probability Letters*, 13, p. 351–356.
18. Jones M.C., Marron J.S., and Sheather S.J. 1996: A Brief Survey of Bandwidth Selection for Density Estimation, *Journal of the American Statistical Association*, 91, p. 401–407.
19. Marron J.S., and Nolan D. 1988: Canonical Kernels for Density Estimations, *Statistics and Probability Letters*, 7 (3), p. 195–199.
20. Marron J.S., and Wand M.P. 1992: Exact mMean Integrated Squared Error, *Annals of Statistics*, 20, p. 712–736.
21. Ruzgas T., and Drulytė I. 2013: Kernel Density Estimators for Gaussian Mixture Models: Supplementary Material, <http://www.personalas.ktu.lt/~truzgas/straipsniai/LJS2013.pdf>.
22. Scott D.W., and Wand M.P. 1991: Feasibility of Multivariate Density Estimates, *Biometrika*, 78, p. 197–205.
23. Scott D.W. 1992: *Multivariate Density Estimation: Theory, Practice, and Visualization*, New York, John Wiley.
24. Silverman B.W. 1986: *Density Estimation for Statistics and Data Analysis*, London, Chapman and Hall.
25. Stachniss C., Plagemann Ch., and Lilienthal A. J. 2009: Learning Gas Distribution Models Using Sparse Gaussian Process Mixtures, *Autonomous Robots*, 26 (2-3), p. 187–202.
26. Tapia R.A., and Thompson J.R. 1978: Nonparametric Probability Density Estimation. *Johns Hopkins Series in the Mathematical Sciences*. Baltimore and London, Johns Hopkins University Press.
27. Yonghong H., Englehart K.B., Hudgins B., and Chan A.D.C. 2005: A Gaussian Mixture Model Based Classification Scheme for Myoelectric Control of Powered Upper Limb Prostheses, *Biomedical Engineering, IEEE Transactions on*, 52 (11), p. 1801–1811.
28. Zhang X., King M.L., and Hyndman R.J. 2004: Bandwidth Selection for Multivariate Kernel Density Estimation Using MCMC, *Computational Statistics and Data Analysis*, 50, p. 3009–3031.
29. Zimmerman M.W., Povinelli R.J., Johnson M.T., and Ropella K.M. 2003: A Reconstructed Phase Space Approach for Distinguishing Ischemic from Non-Ischemic ST Changes Using Holter ECG Data, *Computers in Cardiology*, 21–24, p. 243–246.

BRANDUOLINIAI TANKIO ĮVERTINIAI GAUSO SKIRSTINIŲ MIŠINIŲ MODELIAMS

Tomas Ruzgas, Indrė Drulytė

Santrauka. Darbas skirtas pasiskirstymo tankiams įvertinti. Taikomi tiek fiksuotojo, tiek ir adaptyviojo pločio branduoliniai įvertiniai. Vertinimas paremtas įvairių populiarių branduolių, o taip pat naujo branduolio K palyginimu Monte Karlo metodu naudojant simetrinę vidutinę absoliučiąją paklaidą, išreikštą procentais. Branduolys K yra visada teigiamas, tačiau didelėms argumento reikšmėms jis yra mažesnis nei Gauso skirstinio tankis. Modeliuota naudojant Gauso mišinių modelių rinkinį, kurį pasiūlė J. S. Marronas ir M. P. Wandas (1992). Adaptyviojo pločio branduolinis metodas daugeliu atvejų pranašesnis už fiksuotojo pločio branduolinį metodą. Smarkiai persidengiančių Gauso mišinio komponentų atveju geresni rezultatai gaunami naudojant branduolį K .

Reikšminiai žodžiai: neparametrinis tankio įvertinys, branduolinis metodas, Monte Karlo metodas, adaptyvusis glodinimas, simetrinė vidutinė absoliučioji paklaida išreikšta procentais.