UDC 519.22 (045)

Igor Prokopenko[1]
Igor Martynchuk[2]

OPTIMIZATION OF SMOOTHING PARAMETER BY KERNEL ESTIMATION OF PROBABILITY DENSITY DISTRIBUTION

[1,2]National Aviation University
Kosmonavta Komarova avenue,1, 03680, Kyiv, Ukraine
E-mail: [1]prokop-igor@yandex.ru; [2] igor-martynchuk@hotmail.com

Abstract. *In this paper the problem smoothing parameter choice for probability density kernel estimation is considered. The task of optimal smoothing parameter search according to predetermined criteria is formulated.*
Keywords: kernel estimation of probability density distribution; probability density distribution; smoothing parameter.

1. Analysis of researches and publications

Estimation of the data probability density is the central problem of mathematical statistics. The most known nonparametric estimation of probability density is histogram. Theoretical results about consistency and convergence speed of nonparametric density estimates, which develop histogram estimations, began to emerge from the late fifties. It was found that the best convergence speed of nonparametric estimations is slower than for parametric models. Convergence speed depends on the a priori density class.

In [1] the basic attention is given working out of the theory, studying a consistency of estimation f^* of density f and speed of L_1 metric convergence

$$R = \int_{-\infty}^{\infty} \left| f^*(x) - f(x) \right| dx.$$

The received fundamental results then are effectively applied.

2. Problem statement

Kernel probability density function (PDF) estimations on experimental data $\eta_0, \ldots, \eta_{N-1}$ are widely used in the assumption of its continuity and smoothness [1]. This estimation looks like

$$f_h^*(x) = \frac{1}{N \cdot h} \cdot \sum_{n=0}^{N-1} K\left(\frac{x - \eta_n}{h} \right), \qquad (1)$$

where $K(x)$ – is the kernel function, possessing property

$$\int_{-\infty}^{\infty} K(x)\, dx = 1,$$

N is the sample size; h is the smoothing parameter or scale parameter; $\eta_0, \ldots, \eta_{N-1}$ is the sample of a random variable.

Rectangular and Gaussian kernels

$$K(x) = \begin{cases} 0{,}5 & if \quad -1 \le x \le 1, \\ 0 & otherwise; \end{cases}$$

$$K(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2} \right),$$

find a wide application at construction of kernel estimations.

Article purpose is research of dependence of optimum smoothing parameter h_{opt} from sample size N for various probability density f. On the basis of researches it is necessary to pick up model of regress of this dependence which would optimum approximate $h_{opt}(N)$ for various form of probability densities f.

3. Research of dependence of optimum smoothing parameter h_{opt} from sample size N

Quality of the estimation (1) depends on value of smoothing parameter *h*. Criterion of quality in this work is an intersection of areas under theoretical PDF $f(x)$ and its estimation $f_h^*(x)$. The criterion essence is clear from Fig. 1.

We define a problem of search of optimum smoothing parameter for the specified sample size at kernel probability density estimation by a method of statistical modelling.

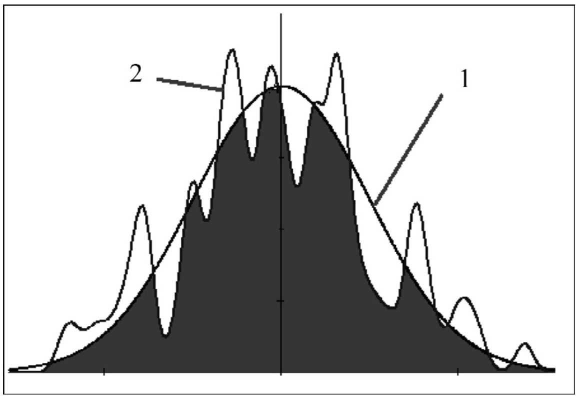

Fig. 1. An intersection of areas under theoretical PDF (1) and its kernel estimation (2)

Let there is an one-dimensional theoretical probability density with the given necessary parameters $f(x, a_0, a_1, \ldots, a_{K-1})$. This function is truncated at the left and on the right to borders $x_{\min} = F^{-1}(\alpha_L)$ and $x_{\max} = F^{-1}(\alpha_R)$ accordingly, where $F^{-1}(\alpha)$ is inverse function to probability distribution function. Borders are chosen so that the probability density normalizing condition with the chosen accuracy was satisfied. On Neumann's method (a principle of Monte-Carlo) the sample of random numbers of size N is generated, which distributed under the given theoretical distribution law. It also will be an initial material for modelling.

On the obtained sample the kernel estimation of PDF is constructed at the given smoothing parameter h and the value of quality criterion is calculated

$$R(h) = \int_{x_{\min}}^{x_{\max}} \min\left\{f_h^*(x), f(x)\right\} dx . \qquad (2)$$

Further such value h_{opt} is found at which the maximum of criterion (2) is achieved.

4. Regression model $h_{opt}(N)$

On the obtained dependence of optimum smoothing parameter on sample size the regression model is constructed by the method of least squares [2].

Let's bring results of modelling (fig. 2, 3) for normal probability density with mean $\mu = 0$, and standard deviation $\sigma = 1$ at following parameters:

1) a minimum and maximum value of x variable of theoretical probability density *f(x)* – $x_{\min} = F^{-1}(0,001)$, $x_{\max} = F^{-1}(0,999)$;

2) a number of discretization intervals of the kernel and theoretical probability density (for numerical integration) – 100;

3) a method of numerical integration (at calculation of intersection of areas) – middle rectangles;

4) a kernel – $K(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$;

5) the estimation of smoothing parameter for each sample size was carry out on 10 experiments;

6) a maximum sample size – 10 000;

7) a search method of maximum in functional dependence (2) – a golden section [3].

The search accuracy of maximum in function $R(h)$ on the segment [$h_{\min}$; $h_{\max}$] is $\varepsilon = 1 \cdot 10^{-3}$. The segment [$h_{\min}$; $h_{\max}$] is selected thus that quantity of intervals of discretization of kernel probability density was enough for correct execution of numerical integration.

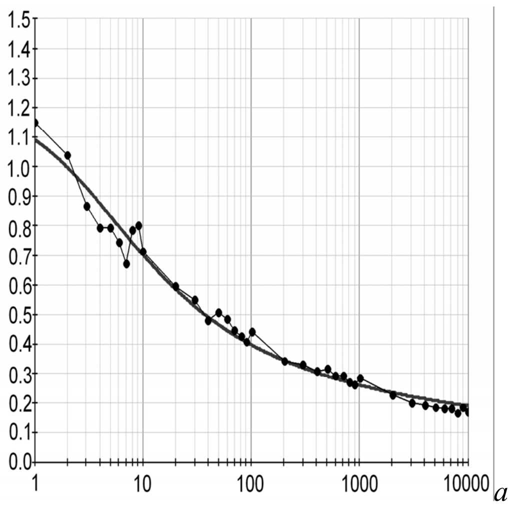

Fig. 2. Dependence of the optimum value of the scale parameter on the sample size (pointwise line) and its regression approximation (solid line) for normal probability density

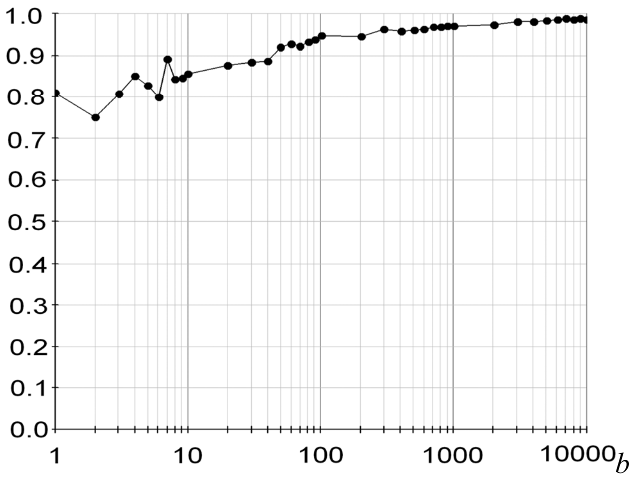

Fig. 3. Dependence of the quality criterion value (with optimum scale parameter) on the sample size for normal probability density

In the given work the regression model has been fitted, which enough precisely approximates dependence of optimum value of scale parameter on sample size

$$h_{opt}(N)=\frac{a}{\lg(N+c)}+b .$$

Such probability densities have been involved in modelling: Beta, Cauchy, Chi-squared, Exponential, F, Gamma, Logistic, Log-Normal, Normal, Rayleigh, Student's T, Uniform, Weibull.

The list of probability distributions is explained by their applications in the radio engineering. Let's bring the Table 1 with various distributions and formulas for $h_{opt}(N)$. At modeling the minimum and maximum value of variable x of theoretical probability density $f(x)$ was set in the following limits:

$x_{\min}=F^{-1}(0{,}1)$, $x_{\max}=F^{-1}(0{,}9)$ – for Cauchy distribution;

$x_{\min}=F^{-1}(0{,}001)$, $x_{\max}=F^{-1}(0{,}9)$ – for log-normal distribution;

$x_{\min}=F^{-1}(0{,}001)$, $x_{\max}=F^{-1}(0{,}999)$ – for all other distributions.

Table 1. Regression approximation of dependence $h_{opt}(N)$.

Probability density name	Probability density function *f(x)*	Probability density parameters	Regression approximation of dependence $h_{opt}(N)$
Beta	$\frac{\Gamma(s0+s1)}{\Gamma(s0)\cdot\Gamma(s1)}\cdot x^{s0-1}\cdot(1-x)^{s1-1}$	$s0=2$ $s1=2$	$\frac{0{,}196}{\lg(N+3{,}796)}-8{,}079\cdot10^{-3}$
Cauchy	$\left[\pi\cdot s\cdot\left[1+\left(\frac{x-l}{s}\right)^2\right]\right]^{-1}$	$l=0$ $s=1$	$\frac{0{,}516}{\lg(N+1{,}832)}+0{,}402$
Chi-squared	$\frac{\exp\left(-\frac{x}{2}\right)}{2\cdot\Gamma\left(\frac{d}{2}\right)}\cdot\left(\frac{x}{2}\right)^{\frac{d}{2}-1}$	$d=100$	$\frac{8{,}834}{\lg(N+1{,}769)}+0{,}838$
Exponential	$r\cdot\exp(-r\cdot x)$	$r=1$	$\frac{0{,}559}{\lg(N+3{,}616)}-0{,}12$
F	$\frac{d0^{\frac{d0}{2}}\cdot d1^{\frac{d1}{2}}\cdot\Gamma\left(\frac{d0+d1}{2}\right)}{\Gamma\left(\frac{d0}{2}\right)\cdot\Gamma\left(\frac{d1}{2}\right)}\times\frac{x^{\frac{d0}{2}-1}}{(d1+d0\cdot x)^{\frac{d0+d1}{2}}}$	$d0=10$ $d1=100$	$\frac{0{,}306}{\lg(N+2{,}436)}+1{,}92\cdot10^{-3}$
Gamma	$\frac{x^{s-1}\exp(-x)}{\Gamma(x)}$	$s=2$	$\frac{0{,}919}{\lg(N+4{,}01)}-0{,}086$
Logistic	$\frac{\exp\left(-\frac{x-l}{s}\right)}{s\cdot\left(1+\exp\left(-\frac{x-l}{s}\right)\right)^2}$	$l=1$ $s=1$	$\frac{1{,}364}{\lg(N+3{,}887)}-0{,}041$

Completion of table 1

Log-Normal	$\frac{1}{\sqrt{2\pi}\cdot\sigma\cdot x}\times$ $\times\exp\left[-\frac{1}{2\sigma^2}\cdot\left(\ln(x)-\mu\right)^2\right]$	$\mu=0$ $\sigma=1$	$\frac{0{,}498}{\lg(N+2{,}48)}-0{,}063$
Normal	$\frac{1}{\sqrt{2\pi}\cdot\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$	$\mu=0$ $\sigma=1$	$\frac{0{,}841}{\lg(N+4{,}719)}-0{,}021$
Rayleigh	$\frac{x}{\sigma^2}\exp\left[-\frac{x^2}{2\sigma^2}\right]$	$\sigma=1$	$\frac{0{,}534}{\lg(N+4{,}498)}-0{,}019$
Student's T	$\frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)\sqrt{\pi\cdot d}}\cdot\left(1+\frac{x^2}{d}\right)^{-\frac{d+1}{2}}$	$d=100$	$\frac{0{,}782}{\lg(N+3{,}6)}-2{,}83\cdot10^{-3}$
Uniform	$\frac{1}{b-a}$	$a=-0{,}5$ $b=0{,}5$	$\frac{0{,}23}{\lg(N+2{,}853)}-0{,}045$
Weibull	$s\cdot x^{s-1}\cdot\exp(-x^s)$	$s=10$	$\frac{0{,}077}{\lg(N+2{,}703)}+2{,}28\cdot10^{-3}$

For an example with normal density ($\mu=0$, $\sigma=1$) the parameters of regression model will be the following: $a=0{,}841$, $b=-0{,}021$, $c=4{,}719$.

5. Conclusion

For each kind of probability density function there is a dependence of optimum smoothing parameter of a kernel estimation of this density on the sample size of experimental data.

At construction of kernel estimation the choice of optimum smoothing parameter of probability density function should be made proceeding from a priori information on a kind of this density.

References

[1] *Devroye L., Gyorfi L.* – Nonparametric Density Estimation: The L1 View. John Wiley & Sons, New York, 1985. (in Russian).

[2] *Demidenko E. Z.* – Linear and nonlinear regression, Moscow: Finansy i statistika, 1981, 302 p. (in Russian).

[3] *Panteleev A. V.* – Optimization methods in examples and problems, Moscow: Visshaja shkola, 2005, 544 p. (in Russian).

Received 10 June 2014

І. Г. Прокопенко[1], І. А. Мартинчук[2]. Оптимізація параметра згладжування ядерних оцінок щільності розподілу ймовірності

[1,2]Національний авіаційний університет, просп. Космонавта Комарова, 1, Київ, Україна, 03680

E-mail: [1]prokop-igor@yandex.ru, [2]igor-martynchuk@hotmail.com

В роботі проводиться обгрунтування вибору параметра згладжування для ядерної оцінки щільності ймовірності. Формулюється оптимізаційна задача пошуку параметра згладжування по заданому критерію. Отримано аналітичні вирази залежності оптимального параметра згладжування від об’єму вибірки для моделей розподілів, які мають найбільше застосування в обробці радіолокаційних сигналів.

Ключові слова: параметр згладжування; щільність розподілу ймовірності; ядерна оцінка щільності розподілу ймовірності.

И. Г. Прокопенко[1], И. А. Мартынчук[2]. Оптимизация параметра сглаживания ядерных оценок плотности распределения вероятности
[1,2]Национальный авиационный университет, просп. Космонавта Комарова, 1, Киев, Украина, 03680
E-mail: [1]prokop-igor@yandex.ru, [2]igor-martynchuk@hotmail.com
В работе приводится обоснование выбора параметра сглаживания для ядерной оценки плотности вероятности. Формулируется оптимизационная задача поиска параметра сглаживания по заданному критерию. Получены аналитические выражения зависимости оптимального параметра сглаживания от объёма выборки для моделей распределений, которые имеют наибольшее применение в обработке радиолокационных сигналов.
Ключевые слова: параметр сглаживания; плотность распределения вероятности; ядерная оценка плотности распределения вероятности.

Prokopenko Igor.
Doctor of Engineering. Professor.
Aviation radio-electronic complexes department, National Aviation University, Kyiv, Ukraine.
Education: Kyiv institute of engineers of civil aviation, Kyiv, Ukraine.
Research area: Signal and data processing.
Publications: 200.
E-mail: prokop-igor@yandex.ru

Martynchuk Igor.
Engineer. Aviation radio-electronic complexes department, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine.
Research area: Signal and data processing.
Publications: 8.
E-mail: igor-martynchuk@hotmail.com