ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 196, 2006

*Aleksandra Baszczyńska**

CHOICE OF THE SMOOTHING PARAMETER IN KERNEL DENSITY ESTIMATION

Abstract. Kernel density estimation is one of the main methods available for univariate density estimation. The problems of choosing the kernel function and choosing the smoothing parametr are of crucial importance in density estimation. Various methods, used in practice, for choosing smoothing parametr are discussed. Some of them are simple, some complicated in calculations, but it must be emphasized that the appropriate choice of method for choosing parameter depends on the purpose for which the density estimate is to be used.

Monte Carlo study is presented, where three "practical rules" and two forms of cross-validation (maximum likelihood CV and least-squares CV) are used in density estimation. The values of smoothing parameters are compared with the "optimal" one, which is obtained by minimizing mean squared error. In all mentioned studies the accuracy of the estimation, measured by mean squared error, is considered.

Key words: density estimation, kernel function, smoothing parameter, practical rules, cross-validation.

1. INTRODUCTION

Kernel density estimation is one of the methods most widely used for nonparametric density estimation. It is defined by:

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

where $x_1, \ldots, x_n$ is a random sample, $K(u)$ is a kernel function satisfying $\int_{-\infty}^{+\infty} K(u)du = 1$, h is smoothing parameter called bandwidth or the window width.

* Ph.D., Chair of Statistical Methods, University of Łódź.

In the process of constructing the estimator we have to choose two parameters of the method: kernel function $K(u)$ and smoothing parameter h (bandwidth). The kernel functions used in practice are symmetric around 0 and integrate to 1, and since the kernel is a density, estimator (1) is also density function. Some of the best known kernel functions are presented in Domański et al. (1998). The properties of some kernel functions are explored by Baszczyńska (2005). The smoothing parameter $h\,(h>0)$ regulates the degree of smoothness in estimation of the density function. In other words, for small values of h we get less noisy density estimate, but for big values of h we get very smooth density estimate.

The appropriate choice of smoothing parameter depends on the purpose for which the density estimate is to be used.

Subjective choice of smoothing parameter is a natural method, where several kernel estimations are made with different parameters h and such an estimation is chosen, which is most in accordance with prior ideas about density. This method is used in situation, where the main purpose of density estimation is only to explore the data. Of course, it may be connected with mistakes and errors, when unexperienced user takes advantages with such method of smoothing parameter.

Automatic methods for choosing the bandwidth are used in estimation of density function, which is the base for presenting conclusions, and as a starting point in subjective adjustment. They are also used in situations, where density estimation is based on a large number of data sets or when density estimation is only a part of larger procedure.

2. BANDWIDTH CHOOSING

Assume that:

- given a sample $X_1, \ldots, X_n$, $\hat{f}(x)$ is the estimator of continuous density function
- kernel function is of second order:

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

$$\int_{-\infty}^{+\infty} uK(u)du = 0, \tag{2}$$

$$\int_{-\infty}^{+\infty} u^2K(u)du = k_2 \neq 0.$$

"Practical rules", known also as "reference to a standard distribution", is a simple automatic chooice of parameter h in density estimation. Parametr h is "optimal" when minimizing the approximate mean integrated square error is made. It depends on unknown density. So, for simplicity, it is assumed that the unknown distribution is normal with parameters μ and σ.

The practical rules are the following:

Practical rule I

$$h = 1.06\,\sigma n^{-\frac{1}{5}},$$

where σ is estimated from a sample

Practical rule II

$$h = 0.79\,Rn^{-\frac{1}{5}},$$

where R is interquartile range

Practical rule III

$$h = 0.9\,An^{-\frac{1}{5}},$$

where $A = \min\left(\sigma, \frac{R}{1.34}\right)$

Cross-validation is a well-known method to choose the smoothing parametr in kernel density function. There are two forms of cross-validation: maximum likelihood cross-validation and least-squares cross-validation.

Least-squares cross-validation method of choosing the smoothing parametr consists in minimizing the following:

$$CVNK(h) = \frac{1}{n^2}\frac{1}{h}\sum_{i=1}^{n}\sum_{j=1}^{n} K\cdot K\left(\frac{X_j - X_i}{h}\right) - 2\frac{1}{n}\sum_{i=1}^{n}\frac{1}{n-1}\frac{1}{h}\sum_{j\pm 1}^{n} K\left(\frac{X_i - X_j}{h}\right), \quad (3)$$

where $K\cdot K$ denotes the convolution of the kernel with itself.

For Gaussian and Epanechnikov kernel convolution of the kernel with itself are respectively:

$$K\cdot K(u) = \frac{1}{2\sqrt{\pi}}\exp\left(\frac{-u^2}{4}\right),$$

$$K\cdot K(u) = \begin{cases} \frac{3}{160}(32 - 40u^2 + 20|u|^3 - |u|^5) & \text{dla} \quad |u| \leqslant 2 \\ 0 \quad \text{otherwise} \end{cases}.$$

Maximum likelihood cross-validation consists in choosing the smoothing parametr, which maximizes the following:

$$CVNW(h) = \frac{1}{n}\sum_{i=1}^{n}\log\left[\sum_{j\pm 1}^{n} K\left(\frac{X_i - X_j}{h}\right)\right] - \log[(n-1)h]. \tag{4}$$

3. MONTE CARLO STUDY

Monte Carlo study was conducted to compare the values of smoothing parametrs h in kernel density estimation. Analysis of properties of estimator was done in three basic variants, depending on distribution, from which the data were chosen. These variants are the following:

- variant I: normal distribution $N(0, 0.2)$,
- variant II: mixture of normal distributions: $f(x) = 0.25f_1(x) + 0.75f_2(x)$, where $f_1(x)$ is density function $N(0, 0.2)$, $f_2(x)$ is density function $N(3, 0.5)$. Variant II presents two-modal distributions.
- variant III: mixture of normal distributions: $f(x) = 0.5f_1(x) + 0.25f_2(x) + 0.25f_3(x)$, where $f_1(x)$ is density function $N(0, 0.2)$, $f_2(x)$ is density function $N(3, 0.5)$, $f_3(x)$ is density function $N(7, 0.5)$. Variant III presents three-modal distributions.

In experiment we used some measures:

- mean squred error

$$BSK = \frac{1}{n}\sum_{i=1}^{n}\left[f(x_i) - \hat{f}(x_i)\right]^2,$$

- $MR = \max_i |f(x_i) - \hat{f}(x_i)|$,
- P is a number of cases, where the value of estimator is greater than value of density function in this point (over smoothing),
- N is a number of cases, where the value of estimator is less than value of density function in this point (under smoothing).

Experiment A consists in using in density estimation values of smoothing parameters, which are calculated from „practical rules". In that way estimators (with Gaussian and Epanechnikov kernels and with smoothing parameters h_{I}, h_{II}, h_{III}) are compared with true density functions descibed by variant I, II and III. Optimal smoothing parameter is determined by minimizing BSK in estimation. The results of this part of the study are presented in Tables 1, 2 and 3.

Table 1. Values of smoothing parameters from „practical rules" for variant I

Kernel function	Smoothing parameter	*BSK*	*MR*	*P*	*N*	Optimal smoothing parameter
Gaussian	$h_{\mathrm{I}} = 0.085814$	0.030399	0.397500	44	84	0.08
	$h_{\mathrm{II}} = 0.085716$	0.030393	0.397796	44	84	
	$h_{\mathrm{III}} = 0.072368$	0.034431	0.456595	71	57	
Epanechnikov	$h_{\mathrm{I}} = 0.085814$	0.030365	0.307472	40	88	0.08
	$h_{\mathrm{II}} = 0.085716$	0.030322	0.307761	39	89	
	$h_{\mathrm{III}} = 0.072368$	0.029081	0.386202	55	73	

Table 2. Values of smoothing parameters from „practical rules" for variant II

Kernel function	Smoothing parameter	*BSK*	*MR*	*P*	*N*	Optimal smoothing parameter
Gaussian	$h_{\mathrm{I}} = 0.818362$	0.046498	0.376785	25	103	0.18
	$h_{\mathrm{II}} = 0.560079$	0.028571	0.329700	30	98	
	$h_{\mathrm{III}} = 0.472863$	0.022256	0.302722	32	96	
Epanechnikov	$h_{\mathrm{I}} = 0.818362$	0.053581	0.396041	25	103	0.16
	$h_{\mathrm{II}} = 0.560079$	0.032578	0.352379	30	98	
	$h_{\mathrm{III}} = 0.472863$	0.025426	0.327119	32	96	

Table 3. Values of smoothing parameters from „practical rules" for variant III

Kernel function	Smoothing parameter	*BSK*	*MR*	*P*	*N*	Optimal smoothing parameter
Gaussian	$h_{\mathrm{I}} = 1.267892$	0.198267	0.833802	18	110	0.11
	$h_{\mathrm{II}} = 1.164849$	0.191606	0.822093	19	109	
	$h_{\mathrm{III}} = 0.983456$	0.176536	0.794683	19	109	
Epanechnikov	$h_{\mathrm{I}} = 1.267892$	0.212725	0.858054	17	111	0.09
	$h_{\mathrm{II}} = 1.164849$	0.207093	0.848407	16	112	
	$h_{\mathrm{III}} = 0.983456$	0.193565	0.825899	19	109	

In variant I all values of smoothing parameters from „practical rules" are similar to optimal value. In variants II and III differences are big. It means that „practical rules" can be used only in situations when the distribution of population is normal. In other cases, „practical rules" do not provide appropriate estimation of density function. This uncomplicated method of choosing a smoothing parameter, used mostly in applications, turned out to be useless when there is another distribution than the normal one.

In experiment B estimation of density function is made using parameters calculated by methods based on cross-validation (maximum likelihood cross-validation and least-squares cross-validation). Values of smoothing parameters from cross-validation methods and optimal value of parametr h in density estimation (for seven different kernel functions) are presented in Table 4.

Table 4. Values of smoothing parameters determined by cross-validation methods for variant I, II and III

Kernel function	Variant I		Variant II		Variant III	
Gaussian	min*BSK*	$h = 0.0840$	min*BSK*	$h = 0.1340$	min*BSK*	$h = 0.1180$
	max*CVNW*	$h = 0.1003$	max*CVNW*	$h = 0.1949$	max*CVNW*	$h = 0.1385$
	min*CVNK*	$h = 0.0917$	min*CVNK*	$h = 0.1935$	min*CVNK*	$h = 0.1125$
Epanechnikov	min*BSK*	$h = 0.1770$	min*BSK*	$h = 0.3100$	min*BSK*	$h = 0.2530$
	max*CVNW*	$h = 0.2071$	max*CVNW*	$h = 0.4110$	max*CVNW*	$h = 0.3016$
	min*CVNK*	$h = 0.2107$	min*CVNK*	$h = 0.4030$	min*CVNK*	$h = 0.2171$
Triangle	min*BSK*	$h = 0.1916$	min*BSK*	$h = 0.3245$	min*BSK*	$h = 0.2757$
	max*CVNW*	$h = 0.2167$	max*CVNW*	$h = 0.4379$	max*CVNW*	$h = 0.3464$
Uniform	min*BSK*	$h = 0.1355$	min*BSK*	$h = 0.2410$	min*BSK*	$h = 0.2008$
	max*CVNW*	$h = 0.1701$	max*CVNW*	$h = 0.3429$	max*CVNW*	$h = 0.2897$
Quartic	min*BSK*	$h = 0.2075$	min*BSK*	$h = 0.3429$	min*BSK*	$h = 0.3018$
	max*CVNW*	$h = 0.2425$	max*CVNW*	$h = 0.4835$	max*CVNW*	$h = 0.3693$
Triweight	min*BSK*	$h = 0.2356$	min*BSK*	$h = 0.3873$	min*BSK*	$h = 0.3452$
	max*CVNW*	$h = 0.2914$	max*CVNW*	$h = 0.5534$	max*CVNW*	$h = 0.4124$
Cosinus	min*BSK*	$h = 0.1783$	min*BSK*	$h = 0.3137$	min*BSK*	$h = 0.2586$
	max*CVNW*	$h = 0.2113$	max*CVNW*	$h = 0.4168$	max*CVNW*	$h = 0.3105$

In all examined cases (seven kernel functions and different distributions of population denoted as variants I, II, III) value of smoothing parameter derived from cross-validation methods is bigger than optimal value (last column in Tables 1, 2 and 3). It means that criterium of cross-validation causes less smoothness of kernel density estimator. Independently of distribution of population, from which sample is derived, Gaussian kernel is connected with smaller parametr h – it can indicate some smoothing properties of this kernel.

REFERENCES

Baszczyńska A. (2005), *Some Remarks on the Choice of the Kernel Function in Density Estimation, Acta Universitatis Lodziensis*, Folia Oeconomica.

Domański Cz., Pruska K., Wagner W. (1998), *Wnioskowanie statystyczne przy nieklasycznych założeniach*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Hardle W. (1991), *Smoothing Techniques. With Implementation in S*, Springer-Verlang, New York.

Priestley M., Chao M. (1972), "Nonparametric Function Fitting", *Journal of the Royal Statistical Society B*, **34**, 385–392.

Rosenblatt M. (1956), "Remarks on Some Nonparametric Estimation of a Density Function", *Annals of Mathematical Statistics*, **27**, 832–837.

Silverman B. W. (1996), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Wand M., Jones M. (1995), *Kernel Smoothing*, Chapman and Hall, London.

Aleksandra Baszczyńska

WYBÓR PARAMETRU WYGŁADZANIA W ESTYMACJI JĄDROWEJ FUNKCJI GĘSTOŚCI

(Streszczenie)

Jądrowa estymacja jest jedną z podstawowych metod nieparametrycznej estymacji funkcji gęstości. Zagadnienie wyboru funkcji jądra oraz wyboru właściwej wartości parametru wygładzania traktowane są jako zasadnicze w estymacji funkcji gęstości. W pracy rozważane są różne metody wyboru parametru wygładzania w estymacji jądrowej, od metod najprostszych do nieco bardziej złożonych. Należy podkreślić jednak, iż wybór metody wyboru parametru wygładzania zależy od celu dokonywanej estymacji charakterystyki funkcyjnej.

W artykule przedstawiono również wyniki z przeprowadzonego eksperymentu Monte Carlo, gdzie rozważano trzy „praktyczne zasady" wyboru parametru wygładzania oraz dwie metody *cross-validation* (największej wiarygodności i najmniejszych kwadratów). Wartości tak otrzymanych parametrów wygładzania są porównywane z parametrem otrzymanym poprzez minimalizację błędu średniokwadratowego, traktowanym jako parametr „optymalny".