ON INTERVAL ESTIMATING REGRESSION

Marcin Michalak

Institute of Informatics, Silesian University of Technology, Gliwice, Poland
`Marcin.Michalak@polsl.pl`

ABSTRACT

This paper presents a new look on the well-known nonparametric regression estimator – the Nadaraya-Watson kernel estimator. Though it was invented 50 years ago it still being applied in many fields. After these yearsfoundations of uncertainty theory – interval analysis – are joined with this estimator. The paper presents the background of Nadaraya-Watson kernel estimator together with the basis of interval analysis and shows the interval Nadaraya-Watson kernel estimator.

KEYWORDS

Interval Analysis, Nonparametric Regression, Nadaraya-Watson Kernel Estimator.

1. INTRODUCTION

The main difference between the mathematical and physical interpretation of a number is that from the mathematical point of view the number is a well-defined point in some space while in physics number (a value) cannot be measured without nonzero level of the uncertainty: in the macro world it is the limitation of our eye precision and the rule precision, for example during measuring the apple diameter, and in the micro world it was well described by Werner Heisenberg and his uncertainty principle.

For a long time scientists have been trying to describe the uncertainty in the mathematical way and applying it in the data processing. As the most famous approaches fuzzy sets [1] (with fuzzy numbers [2]) rough sets[3] or interval analysis [4][5] should be mentioned.

The main motivation of the research presented in this paper is the 50th anniversary of very simple nonparametric regression function estimator – the Nadaraya-Watson kernel estimator. It was invented independently by Nadaraya [6] and Watson [7] in 1964. As the aim of the research the application of interval arithmetic into this method of regression analysis was stated. The title of the paper connects directly to the Nadaraya paper.

The paper is organised as follows: it starts from the short reminder of the Nadaraya-Watson kernel estimator and the brief overview of the interval analysis. Afterwards, the interval approach to kernel regression is presented and followed by the results of experiments on the synthetic data. The paper ends with some final conclusions.

2. KERNEL REGRESSION

Some of the commonly known examples of the nonparametric regression estimators are kernel estimators. This group of methods is developed from the solution of nonparametric estimation of the density function. The Nadaraya-Watson kernel estimator is the simplest kernel regression estimator [6][7]. For the one dimensional case it is given by the equation:

$$\tilde{f}(x) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)}$$

where pairs (x_i, y_i) are known, K is a kernel function and h is so called smoothing parameter. This estimator can be explained as the kind of moving average: the kernel function K is responsible for the shape of weights of averaged values and the smoothing parameter defines the range of input values.

In the Table 1. there are presented most popular kernel functions (I is an indicator function). As we can see only one of them, the Gaussian, has the infinite domain what means that it takes into consideration (estimation of the value at x) all given points, even very distant. Other kernels narrow the neighbourhood of x to the value of the smoothing parameter h.In the onedimensional case only pairs from the training set, which first entries belong to the interval $[x - h, x + h]$ are averaged.

Table 1. Most popular kernel functions

Kernel function	Equation
Uniform	$K(x) = 0.5\, I(-1 < x < 1)$
Triangular	$K(x) = (1 - |x|)\, I(-1 < x < 1)$
Epanechnikov	$K(x) = 0.75(1 - x^2)I(-1 < x < 1)$
Biweight	$K(x) = 0.9375(1 - x^2)^2 I(-1 < x < 1)$
Gaussian	$K(x) = (2\pi)^{-0.5}\exp\left(-x^2/2\right)$

In practice the selection of kernel function generally influences less than the selection of the smoothing parameter. The less complicated method of estimating its value is approximation of the Mean Integrated Square root Error (MISE). Its final results – optimal values of h – can be evaluated as follows:

$$h_0 = 1.06\, \tilde{\sigma} n^{-0.2}$$

where $\tilde{\sigma}$ is the standard deviation of arguments (x), or as follows:

$$h_0 = 1.06 \min\{\, 0.74 \cdot IR, \tilde{\sigma}\} n^{-0.2}$$

where IR is an interquartile range of x. Details of these calculations can be found in [8].More advanced methods of estimating h can be also found in [9][10][11][12][13].

3. INTERVAL COMPUTATIONS

Interval arithmetic is the branch of mathematics where the number is represented as the interval, due to the uncertainty of the measurement that brought the number. As the first use of this kind of number representation the Archimedes approximation of the π can be recalled: Archimedes stated that 223/71 < π < 22/7.

If two interval numbers are considered then it is interesting how the sum, product or other arithmetical operation should be defined, to give the interpretable result. Next subsection brings definitions of most basic interval operations and the following one shows the problem of interval computations.

3.1. Definition of the Interval Arithmetic Operation

If the non-exactness of the number is represented as its lower and upper bound it is necessary to provide new methods of performing calculations on interval numbers. For two interval numbers X and Y their sum must take into consideration all possible values from two intervals as follows:

$$X + Y = \{x + y : x \in X, y \in Y\}$$

This means that sum of two intervals is the set of all possible results of sum of numbers, coming from each particular interval. In the similar way the following simple arithmetic operations may be defined, the difference:

$$X - Y = \{x - y : x \in X, y \in Y\}$$

the product:

$$X \cdot Y = \{x \cdot y : x \in X, y \in Y\}$$

and the quotient:

$$X/Y = \{x/y : x \in X, y \in Y\}$$

The last operation requires to assure that $0 \notin Y$.

All operations become less complicated to perform when we just consider their bounds. Assuming that the interval X is the range $[\underline{X}, \overline{X}](\underline{X} \le \overline{X})$ and interval Y is the range $[\underline{Y}, \overline{Y}](\underline{Y} \le \overline{Y})$ we can write simply that:

$$X + Y = [\underline{X} + \underline{Y}, \overline{X} + \overline{Y}]$$

The similar way of defining the subtraction leads to the following formula:

$$X - Y = [\underline{X} - \overline{Y}, \overline{X} - \underline{Y}]$$

which can be derived from the dependence:

$$X - Y = X + (-Y)$$

Situation becomes a little more complicated when the product of two intervals is taken into consideration. Due to conditions of signs of lower and upper bounds of intervals the bounds of the result of the operation take values as it is presented in the Table 2.

Table 2. Definition of product of two intervals.

Case	$\underline{X \cdot Y}$	$\overline{X \cdot Y}$
$0 \le \underline{X}$ and $0 \le \underline{Y}$	$\underline{X} \cdot \underline{Y}$	$\overline{X} \cdot \overline{Y}$
$\underline{X} < 0 < \overline{X}$ and $0 \le \underline{Y}$	$\underline{X} \cdot \overline{Y}$	$\overline{X} \cdot \overline{Y}$
$\overline{X} \le 0$ and $0 \le \underline{Y}$	$\underline{X} \cdot \overline{Y}$	$\overline{X} \cdot \underline{Y}$
$0 \le \underline{X}$ and $\underline{Y} < 0 < \overline{Y}$	$\overline{X} \cdot \underline{Y}$	$\overline{X} \cdot \overline{Y}$
$\overline{X} \le 0$ and $\underline{Y} < 0 < \overline{Y}$	$\underline{X} \cdot \overline{Y}$	$\underline{X} \cdot \underline{Y}$
$0 \le \underline{X}$ and $\overline{Y} \le 0$	$\overline{X} \cdot \underline{Y}$	$\underline{X} \cdot \overline{Y}$

$\underline{X} < 0 < \overline{X}$ and $\overline{Y} \le 0$	$\overline{X} \cdot \underline{Y}$	$\underline{X} \cdot \underline{Y}$
$\underline{X} \le 0$ and $\overline{Y} \le 0$	$\overline{X} \cdot \underline{Y}$	$\underline{X} \cdot \overline{Y}$
$\underline{X} < 0 < \overline{X}$ and $\underline{Y} < 0 < \overline{Y}$	$\min\{\underline{X} \cdot \overline{Y}, \overline{X} \cdot \underline{Y}\}$	$\max\{\underline{X} \cdot \underline{Y}, \overline{X} \cdot \overline{Y}\}$

Definition of division can be obtained from the product of inversion of the second argument, assuming again $0 \notin Y$:

$$1/Y = [1/\overline{Y}, 1/\underline{Y}]$$

and

$$X/Y = X \cdot (1/Y)$$

3.2. Problems of the Interval Arithmetic Operations

One of the problems of interval computations is that we cannot assume two expressions in the real arithmetic to be equivalent in the sense of interval analysis. This will be shown on the following example. Let us consider the following formula in the ordinary (real) arithmetic:

$$1 + \frac{b}{a} = \frac{a + b}{a}$$

Both sides of this formula are equivalent as long as the assumption of $a \ne 0$ is fulfilled. Now let us set $a = [3, 5]$ and $b = [7, 10]$ and calculate both sides with the interval arithmetic:

$$L = 1 + \frac{[7,10]}{[3,5]} = [1,1] + [7,10] \cdot \frac{1}{\left[\frac{1}{5}, \frac{1}{3}\right]} = [1,1] + \left[\frac{7}{5}, \frac{10}{3}\right] = \left[\frac{12}{5}, \frac{13}{3}\right] = [2.4, 4.(3)]$$

$$R = \frac{[3,5] + [7,10]}{[3,5]} = \frac{[10,15]}{[3,5]} = [10,15] \cdot \frac{1}{\left[\frac{1}{5}, \frac{1}{3}\right]} = \left[\frac{10}{5}, \frac{15}{3}\right] = [2,5]$$

$$L \ne R$$

Although both expressions are equivalent in the real arithmetic it occurs that they are not in the interval sense. This difference is caused in general by the phenomenon of *interval dependency*. When we have the interval a in the nominator and the denominator of the fraction has the same value from the same interval, in calculation they are treated as independent. It becomes more apparent when we compare the result of squaring interval number.

From the origin of the idea of interval computation we have the following definition of the square function:

$$f(X) = \{x^2 : x \in X\}$$

This can be expanded as:

$$f(X) = \begin{cases} \left[\underline{X}^2, \overline{X}^2\right] & 0 \le \underline{X} \\ \left[\underline{X}^2, \overline{X}^2\right] & \overline{X} \le 0 \\ \left[0, \max\left\{\underline{X}^2, \overline{X}^2\right\}\right] & \underline{X} < 0 < \overline{X} \end{cases}$$

If we are interested in calculating $[-2, 2]^2$ we obtain the interval $[0, 4]$ but if we expand $[-2, 2]^2$ as $[-2, 2] \cdot [-2, 2]$ it will give an interval $[-4, 4]$. In the first approach an interval is calculated as the

set of all possible squares of values from the given SINGLE input. In the second one – all possible values of product of TWO values from the interval are determined.

4. KERNEL INTERVAL REGRESSION

The first approach of interval kernel regression is to apply the formula from the section 2. of the paper without any transformation. From the other side we know, that formulas that are equal in the ordinary real calculation may give different results when applied for interval numbers. So it is worth to transform the original Nadaraya-Watson kernel regression equation and compare its interval results on the same interval data.

Let us consider the equivalent formula of Nadaraya-Watson kernel estimator, equivalent in the domain of real numbers.

$$\tilde{f}(x) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)} = \sum_{i=1}^{n} y_i \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x-x_j}{h}\right)}$$

5. EXPERIMENTS

Experiments were performed on several artificial datasets, where the estimated function was given but there was a random noise, with the zero mean values, added to each function value. The standard deviation of the noise and specification of all datasets are presented in the Table 3. The first two sets come from [14] and the rest from [15].

Table 3. Specification of used datasets.

Dataset	x	y	σ
1	$[-\pi;\pi]$	$y(x) = e^{\frac{(x-1)^2}{4}}$	0.15
2	$[0;1]$	$y(x) = 0.3[\sin(5x-3)]$	$\sqrt{1.5}$
3	$[-2;2]$	$y(x) = x + 2e^{-16x^2}$	0.4
4	$[-2;2]$	$y(x) = \sin 2x + 2e^{-16x^2}$	0.3
5	$[-2;2]$	$y(x) = 0.3e^{-4(x+1)^2} + 0.7e^{-16(x-1)^2}$	0.1
6	$[-2;2]$	$y(x) = 0.4x + 1$	0.15

Each dataset contained 101 observations, distributed uniformly in the domain. This 101 pairs of observations were recalculated into 201 pairs, which contained original 101 as pairs of interval numbers with their lower and upper bounds equal and 100 new pairs with typically interval numbers whose lower and upper bounds were defined as follows:

Table 4. Transformation of real pairs into interval.

Domain	Pairs
real	$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_{101}, y_{101})\}$
interval	$\{([x_1, x_2], [y_1, y_2]), ([x_2, x_3], [y_2, y_3]), \dots, ([x_{100}, x_{101}], [y_{100}, y_{101}])\}$

Two versions of Nadaraya-Watson kernel estimator were used: the simple (NW1) and the modified, from the section 4 (NW2). As the h estimator the equation basing on interquartile range was used.

For the purpose of estimators evaluations the prediction of values in original (non-interval) 101 points was taken into consideration. As the regression error Root Mean Squared Error was used, which formula is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2}$$

6. RESULTS

Twelve experiments were performed (six datasets on two versions of Nadaraya-Watson interval kernel estimator). On the Figure 1. their results are presented. Points are the noised datasets, black points marked with × are the result of standard Nadaraya-Watson estimator and red rectangles present interval output of the kernel regression. Also the original dependence is slightly visible as the black line. The interval output of the interval regressor is shown as the vertical bar.
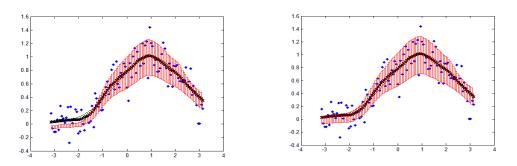

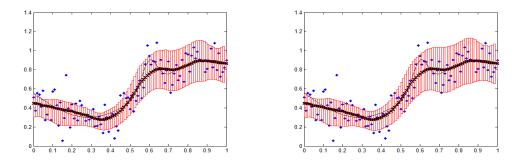Figure 1. NW1 (left) and NW2 (right) results for dataset #1.

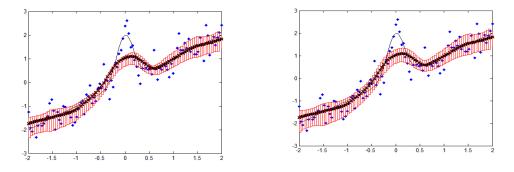Figure 2. NW1 (left) and NW2 (right) results for dataset #2.

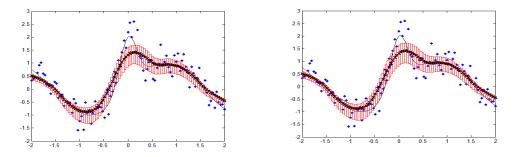Figure 3. NW1 (left) and NW2 (right) results for dataset #3.

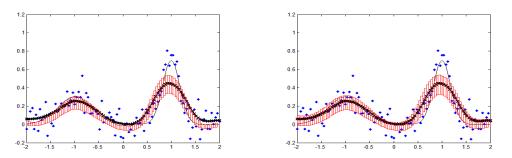Figure 4. NW1 (left) and NW2 (right) results for dataset #4.

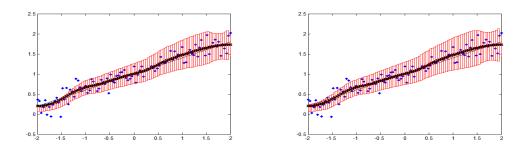Figure 5. NW1 (left) and NW2 (right) results for dataset #5.

Figure 6. NW1 (left) and NW2 (right) results for dataset #1.

The qualitative evaluation of three models of kernel regression is presented in the Table 5. For standard Nadraya-Watson regression (column marked as NW) a normal (real) RMSE error is presented, while for other two models their interval error is presented.

Table 5. Comparison of three regression models.

Dataset	NW	NW1	NW2
1	0.17134	[0.13636; 0.21691]	[0.13075; 0.22187]
2	0.12390	[0.089199; 0.16802]	[0.088208; 0.16954]
3	0.45991	[0.39968; 0.54332]	[0.38513; 0.55361]
4	0.41109	[0.35302; 0.4886]	[0.34866; 0.49308]
5	0.12824	[0.11465; 0.14434]	[0.10982; 0.15063]
6	0.14425	[0.095705; 0.20777]	[0.090103; 0.21337]

On Figures 1 to 6 we can see that in most of cases (excluding NW1 model for datasets 1 and 5) the interval version of Nadaraya-Watson kernel estimator covers the results of its original version. It means that the value from the original estimator belongs to the interval returned by the interval estimator.

For both of interval estimators the final regression error also contains the value of the error of the non-interval model. It can be explained very simply with the Figures – as interval outputs are "wider" than the real outputs of the estimator it causes that one of the bounds is closer to the original (input) value and the other is further.

Another interesting remark is that the error of the NW2 model is wider than NW1 and is its superset:

$$RMSE(NW1) \subset RMSE(NW2)$$

7. CONCLUSIONS

This paper presents the new approach on the 50 years old Nadaraya-Watson kernel estimator. The novelty is the combination of the kernel estimator and the interval arithmetic. Due to the phenomenon of interval dependency two versions of this kernel estimator in the interval approach were taken into consideration. Application of any of two modifications gives the opportunity to evaluate the level of the uncertainty of the value estimated with the non-interval analysis.

ACKNOWLEDGEMENTS

REFERENCES

[1] Zadeh, Lofti (1965) "Fuzzy sets",*Information and Control*,Vol. 8, No. 3, pp. 338–353.
[2] Dubois, Didier & Prade, Henri (1978) "Operations on Fuzzy Numbers", *International Journal of Systems Science*, Vol. 9, No. 6, pp. 613-626.
[3] Pawlak, Zdzisław (1991) *Rough Sets: Theoretical Aspects of Reasoning About Data*, Kluwer.
[4] Warmus, Mieczysław (1956) "Calculus of Approximations", *Bulletin de l'Academie Polonaise Des Sciences*, Vol. 4, No. 5, pp. 253-257.

[5] Moore, Ramon (1966) *Interval Analysis*, Prentice – Hall, New Jersey

[6] Nadaraya, Elizbar (1964)"On estimating regression",*Theory of Probability and Its Applications*, Vol. 9, pp.141-142.

[7] Watson, Geoffrey (1964)"Smooth regression analysis",*Sankhya - The Indian Journal of Statistics*, Vol. 26, pp. 359-372.

[8] Silverman Bernard (1986) *Density estimation for statistics and data analysis*. Chapman & Hall

[9] Fan, Jianqing& Gijbels Irene (1992) "Variable bandwidth and local linearregression smoothers", *The Annals of Statistics*, Vol. 20, No. 4, pp.2008–2036.

[10] Gasser, Theo& Kneip, Alois & Kohler, Walther (1991) "A flexible and fast methodfor automatic smoothing",*Journal of American Statistical Association*,Vol. 86, No. 415, pp. 643–652.

[11] Terrell, George (1990) "The maximal smoothing principle in densityestimation",*Journal of American Statistical Association*,Vol. 85, No. 410, pp.470–477.

[12] Terrell, George& Scott, David (1992) "Variable kernel density estimation", *The Annals of Statistics*, Vol. 20, No. 3, pp. 1236–1265.

[13] Turlach, Berwin (1993) "Bandwidth selection in kernel density estimation:a review", C.O.R.E. and Institut de Statistique, UniversiteCatholique de Louvain.

[14] Gajek, Lesław& Kałuszka, Marek (2000)*Wnioskowanie Statystyczne (Statistical Reasining)* (in Polish), Warsaw.

[15] Fan, Jianqing& Gijbels, Irene (1995)"Data-driven bandwidth selection in local polynomial regression:variable bandwidth selection and spatial adaptation", *Journal of the Royal StatisticalAssociation, Series B*, Vol. 57, pp. 371 – 394.

AUTHOR

Marcin Michalak was born in Poland in 1981. He received his M.Sc. Eng. in computer science from the Silesian University of Technology in 2005 and Ph.D. degree in 2009 from the same university. His scientific interests is in machine learning, data mining, rough sets and biclustering. He is an author and coauthor of over 50 scientific papers.