Thirteenth International Conference
Financial and Actuarial Mathematics – FAM-2020

Practical Comparison between Reweighted Nadaraya-Watson and Local Kernel Mode Estimators

Yousef Methkal Abd Algani,
Department of Mathematics, Sakhnin College, Israel.
Department of Mathematics,
The Arab Academic college for education in Israel
yosefabdalgani@gmail.com
https://orcid.org/0000-0003-2801-5880

Abstract: The main aim of this research has been to improve and develop the nonparametric kernel estimation of a conditional mode under some regularity conditions. Additionally, the choice of bandwidth depended on n and the location and called local variable bandwidth. Re-Weighted Nadaraya-Watson, $\hat{M}_{RNW}(x)$ and the local variable kernel estimator, $\hat{M}_L(x)$ are compared theoretically and practically. The findings of the comparison indicates that $\hat{M}_L(x)$ is better than $\hat{M}_{RNW}(x)$, and that's refers to the strong effect of the bandwidth h(x) as a smoothing parameter in the local variable estimation. Several simulated and real data applications were used to prove the accuracy of the results obtained by the Reweighed Nadaraya-Watson estimator and Local Variable Kernel estimators

Key words: Practical Comparison - Reweighted Nadaraya-Watson- Local Kernel Mode Estimators

1. Introduction

Kernel estimation refers to a general class of techniques for non–parametric estimation of functions. In comparison to parametric estimators where the estimator has a fixed functional form (structure) and the parameters of this function are the only information we need to store, non-parametric estimators have no fixed structure and depend upon all the data points to reach an estimate (Raid B. Salha1 & Hazem I. El Shekh Ahmed**, 2015).**

A kernel is a weighting function used in non-parametric estimation techniques. Kernel density estimation is absolutely dependent on kernels to estimate random variables' density functions, or in kernel regression to estimate the conditional expectation of a random variable. Kernels are also used in time-series applications. (Nan Yang, et al, 2019)

Kernel density estimation is an alternative statistical intensive method, which involves smoothing the data while saving the overall structure. It is a good method of reconstructing an unknown population from a random sample of data, overcomes the problems of histograms, and has many applications in statistics. Kernel density estimation is considered as a most powerful method that gives accurate results and decreases the expected errors (De Gooijer, J.and Zerom, D. 2003).

There are certain standard notations it would be convenient to mention in the presentation to be undertaken here. If, we have a random sample $X_1, X_2, ..., X_n$ from a continuous univeriate distribution with a probability density function (pdf), that function – which we will be try to estimate, will be denoted here by f. The probability density function may be considered to be one of the most important concepts in statistics. Specifying the probability density function f gives a natural description of the distribution of X, and allows probabilities associated with X to be found from the relation:

$$P(a < x < b) = \int_a^b f(x)\, dx \quad \text{for all } a < b \qquad (1)$$

The symbol $\hat{f}$ will be used to denote the estimator of the probability density function f.

In addition and without loss of generality we will integrate the functions over the real line R, using the symbol $\int f(.)$.

2. The research Problem

This section discusses the development of statistical methods, leading to the kernel density estimation method and A Comparison of Reweighted Nadaraya-Watson and Local Kernel Mode Estimators. To address this problem, the study tries to answer the following questions

1. What is the Kernel estimation?
2. What is the of Reweighted Nadaraya-Watson?
3. What are differences between Reweighted Nadaraya-Watson and Local Kernel Mode Estimators?

3. Kernel Density Estimation

The kernel density estimation involves placing a symmetrical surface over each point, evaluating the distance from the point to a reference location based on a mathematical function, and summing the value of all the surfaces for that reference location. This procedure is repeated for all reference locations (Fan, J. and Q.W. Yao (2005).

Briefly the kernel estimator can be viewed as a sum of bumps placed at the observations. The key point to understand how the nonparametric estimators work for the probability density function is to focus first on the definition of this curve (Fan, J., and T. H. Yim, 2004). The limit concept of ratio between probability mass in the neighborhood of a point and the "size" of that neighborhood plays an important role when considering what a simple histogram is doing (Hall, P., J. S. Racine, and Q. Li,2004)

This idea can be easily extended to obtain the moving histogram (or Naïve estimator) and, finally, the kernel density estimator by just giving different weights to the data according to their proximity to the interest point (Salha, R. and Ioannides, D. 2004). A mathematical expression for the well-known Parzen kernel density estimator can be then introduced:

3.1. Properties of the Kernels

In this section we discuss some essential properties of kernels and give examples of common kernel functions.

These properties and functions comprise the cornerstone of derivations related to kernel density estimation.

A kernel is a piecewise continuous function, satisfies the condition,

$$\int K(x)\,dx = 1 \qquad (3.1)$$

Moreover, the kernel is always considering to be symmetric around zero, and need not have bounded support. In most applications K is considered to be a positive unimodal probability density function on R.

3.2. Properties of the Kernel Density Estimators

The properties of the kernel density estimator

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \qquad (3.2)$$

may be listed as follows:

i. The unknown density function $f(x)$ has continuous second derivative $f''(x)$.

ii. The bandwidth $h = h_n$ satisfies $\lim_{n\to\infty} h = 0$, and $\lim_{n\to\infty} nh = \infty$.

iii. The kernel K is abounded probability density function of order 2 and symmetric about the origin.

iv. $\int tK(t)\,dt = 0$ and $\int t^2 K(t)\,dt = k_2 \neq 0$. (K<∞)

v. $E(|Yi \mid \delta \mid Xi = U) \leq M < \infty$ for some $\delta > 2$, in a neighborhood of x

The estimator depends on the bandwidth $h > 0$ which acts as a tuning parameter. For large bandwidth h, the estimate $\hat{f}(x)$ tends to be very slowly varying as a function of x, while small bandwidths will produce a more highly variable function estimate. The positivity of the kernel function $K(u)$ guarantees a positive density estimate $\hat{f}(x)$ and the normalization $\int K(x)\,dx = 1$ implies that $\int \hat{f}(x)\,dx = 1$ which is necessary for $\hat{f}(x)$ to be a density function. Typically, the kernel function $K(u)$ is chosen as a probability density which is symmetric around 0.

3.3. Kernel Estimation of a Conditional PDF

The goal of conditional density estimation is to construct a density estimate $\hat{f}(y|x)$ for a dependent variable y, conditional on a vector of variables x. In this section, we investigate various types of nonparametric kernel methods for estimating the conditional density function $f(y|x)$ of a random variable y given a random vector x.

Assume that $(X_1,Y_1),(X_2,Y_2),.....,(X_n,Y_n)$\in are an independent identically distributed (i i d) sample from real-valued random variables (X,Y) sitting on a given probability space with random variables that share a joint probability density function pdf $f(x,y)$. The marginal probability density function pdf of X_1 is $g(x)=\int f(x,y)dy$, and the conditional probability density function PDF of Y_1, given $X_1=x$, is

$$f(y|x)=\frac{f(x,y)}{g(x)} \tag{3.3}$$

S.t

$g(.) \text{ is positive at } x$

Conditional density estimation was introduced by Rosenblatt (1969). A bias correction was proposed by Hyndman, et al. (1996). Fan and Gijbels (1996) proposed a direct estimator based on local polynomial estimation and Yao (2003). Bandwidth selection rules have been proposed by Bashtannyk and Hyndman (2001), Fan and Yim (2004), and Hall et al. (2004). The related problem of conditional distribution estimation is examined in a paper by Hall et al. (1999).

Now, we introduce the basic equation of the kernel conditional density estimation (KCDE). In this case, interpolation must happen in both the x and y directions gives us a double kernel estimator, K_1 and K_2. The Kernel function $K(u)$ is assumed to be a Borel symmetric function, and $\{h_n\}$ is a sequence of positive numbers converging to zero. In this study, we use $\hat{f}(x,y)$, $\hat{g}(x)$ and $\hat{f}(y|x)$ as estimators of $f(x,y)$, $g(x)$ and $f(y|x)$ respectively, where

$$\hat{f}(x,y)=\left(nh_n^2\right)^{-1}\sum_{j=1}^{n}K_1\left(\frac{x-X_j}{h_n}\right)K_2\left(\frac{y-Y_j}{h_n}\right)$$

$$\hat{g}(x)=\left(nh_n\right)^{-1}\sum_{j=1}^{n}K\left(\frac{x-X_j}{h_n}\right) \tag{3.4.}$$

and,

$$\hat{f}(y|x)=\frac{\hat{f}(x,y)}{\hat{g}(x)}$$

To construct the proposed conditional density estimator, we first consider two existing kernel-based smoothers of the conditional density, and discuss bandwidth selection for the kernel estimator of conditional density with one explanatory variable. There are several techniques used to investigate some topics of estimation properties. The methods are compared and a practical bandwidth selection strategy which combines the methods is proposed. The methods are compared using two simulation studies and a real data set. The next section deals with methods that intersect and refine the common method used in non-parametric conditional estimation that is widely known as the Nadaraya-Watson (NW) estimator.

4. Comparison of Re-Weighted Nadaraya-Watson and Local Kernel Mode Estimators

According to Cheng, M.-Y., L. Peng, and J.-S. Wu (2007 the estimation of the conditional mode is a matter of both theoretical and practical interest. Suppose that $\left(X_1,Y_1\right),\left(X_2,Y_2\right),\ldots,\left(X_n,Y_n\right)$ are iid random variables distributed as $\left(X,Y\right)$ with joint pdf $f\left(x,y\right)$. The marginal pdf of X is given by $g\left(x\right)=\int_{-\infty}^{\infty} f\left(x,y\right)dy$. Also suppose that the conditional pdf of Y given $X=x$, $f\left(y\,|\,x\right)$ is uniformly continuous in y, and possesses a unique conditional mode $M\left(x\right)$ which is defined as

$$f\left(M\left(x\right)|\,x\right)=\max_{-\infty<y<\infty} f\left(y\,|\,x\right) \qquad (4.1)$$

4.1. Basic Properties and Formulas

Samanta and Tharaneswarar (1990) considered the problem of estimating the conditional mode using the Nadaraya-Watson (NW) estimator of the conditional density function $f_{NW}\left(y\,|\,x\right)$ which is defined as

$$\hat{f}_{NW}\left(y\,|\,x\right)=\frac{\sum_{i=1}^{n} K_h\left(x-X_i\right)K_h\left(y-Y_i\right)}{\sum_{i=1}^{n} K_h\left(x-X_i\right)} \qquad (4.2)$$

where K is a kernel function, $h=h_n$ is a sequence of positive number converges to zero and it is called the bandwidth, and where $K_h\left(.\right)=\frac{1}{h}K\left(\frac{.}{h}\right)$.

Estimating the conditional density function is such a key aspect in the estimation of the conditional mode that the only way of improving the latter estimation is by improving the former.

Since the Nadaraya-Watson estimator suffers from some disadvantages of producing rather large bias and boundary effects, two new estimators of the conditional mode has been proposed by improving the estimation of the Nadaraya-Watson estimator of the conditional density function.

The first estimator is proposed in this research is obtained by using the Re-Weighted Nadaraya-Watson (RNW) estimator as a weighted version of the NW estimator, which combines the better sides of the local linear LL estimators such as bias reduction and no boundary effects and to preserved the property of the NW estimator is always a distribution function.

The RNW conditional density estimator is defined as

$$\hat{f}_{RNW}(y\,|\,x)=\frac{\sum_{i=1}^{n}\tau_i(x)(X_i-x)K_h(y-Y_i)}{\sum_{i=1}^{n}\tau_i(x)K_h(x-X_i)} \qquad (4.3)$$

where $\tau_i(x)$ denote the probability like weights with properties that

$$\left.\begin{array}{c}\tau_i(x)\geq 0,\quad \sum_{i=1}^{n}\tau_i(x)=1,\\ \sum_{i=1}^{n}\tau_i(x)(X_i-x)K_h(x-X_i)=0\end{array}\right\} \qquad (4.4)$$

The role of $\tau_i(x)$ is to adjust the (NW) weights such that the resulting conditional density estimator resembles that from LL estimators.

For more details, see De Gooijer and Zerom (2003).

The second estimator is obtained by using the so-called local variable kernel density estimator, $f_L\left(y\,|\,x\right)$, which depends on a different bandwidth $h\left(x\right)$, for each point x at which $f\left(y\,|\,x\right)$ is estimated. The bandwidth $h(x)$ can be chosen such that $h(x) \propto \frac{1}{\hat{f}(x)}$, where $\hat{f}(x)$ is some pilot estimator of $f(x)$ at the point x. See Wand and Jones (1995, pp. 41). It is defined as

$$\hat{f}(y|x) = \frac{\frac{1}{h(y)}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h(x)}\right) K\left(\frac{y - Y_i}{h(y)}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h(x)}\right)} \qquad (4.5)$$

Two new estimators $\hat{M}_{RNW}$ and $\hat{M}_L$ of the conditional mode function were proposed in

$$f(\hat{M}_{RNW}(x)|x) = \max_{y \in R} \hat{f}_{RNW}(y|x) \qquad (4.6)$$

and,

$$f(\hat{M}_L(x)|x) = \max_{y \in R} \hat{f}_L(y|x) \qquad (4.7)$$

4.2. Theoretical Comparison

Let us donate the Re-Weighted Nadaraya-Watson by , $\hat{M}_{RNW}(x)$ and the local variable kernel estimator by, $\hat{M}_L(x)$. And In this section, we offer a brief theoretical comparison of $\hat{M}_{RNW}(x)$ and $\hat{M}_L(x)$ the performance of the estimator has been tested using the mean squared errors (MSE) and by looking at their asymptotic mean squared errors (MSE's) as $n \to \infty$.

The assumptions underlying these results are standard ones.

Let us supposed that the 3rd derivative of $f(y|x)$ exist, then:+ ∂
$f^{(i,j)}(y|x) = \partial^{i+j} f(y|x)/\partial x^i \partial y^j$

The asymptotic bias and variance of $\hat{M}_{RNW}(x)$ and $\hat{M}_L(x)$ are listed below.

$$MSE = \left(\hat{M}_L(x)\right) = \frac{1}{2}\left[h^2(x) f^{(2,1)}(M(x)|x) + h^2(y) f^{(0,3)}(M(x)|x)\right] \int u^2 K(u) du$$

$$Var\left(\hat{M}_L(x)\right) = \frac{f(x, M(x)) \int\int \left(K(u) K^{(1)}(v)\right)^2 dudv}{nh(x) h^3(y) \left[f^{(0,2)}(x, M(x))\right]^2}$$

$$MSE\left(\hat{M}_{RNW}(x)\right) = \frac{1}{2} h^2 \left[f^{(2,1)}(M(x)|x) + f^{(0,3)}(M(x)|x)\right] \int u^2 K(u) du$$

$$Var\left(\hat{M}_{RNW}(x)\right) = \frac{f(x, M(x)) \int\int \left(K(u) K^{(1)}(v)\right)^2 dudv}{nh^4 \left[f^{(0,2)}(x, M(x))\right]^2}$$

The two asymptotic biases and variances differ by the bandwidth, thus they are the same when $h = h(x) = h(y)$, as will occur when the choice of the local variable bandwidth is

$$h(x) = \frac{h}{\hat{f}(x)} \quad \text{and} \quad \hat{f}(x) = 1.$$

This means as the value of $\hat{f}(x)$ closer to one, the two asymptotic MSE are closer.

In general, however, the MSE of the two estimators are different and depend on the relation between the local variable and constant bandwidth.

4.3. Practical Comparison

Our study of the practical differences between the local variable kernel (LVKES), $\hat{M}_L(x)$ and the Re-Weighted Nadaraya-Watson estimator (RNW), $\hat{M}_{RNW}(x)$ of the conditional mode is based on applications to real life data and simulation study. The standard Gaussian kernel function is used throughout the applications.

4.3.1. Real Life Data

Consider the Ethanol data, which describes the relationship between the predictor E (Ethanol) and the response NOx (Nitric Oxide) (Source of the data is in S-Plus program). Figure 1 indicates the relationship is not linear. The regression relation estimated by using the two estimators $\hat{M}_L(x)$ and $\hat{M}_{RNW}(x)$. A scatter plot of the data together with the graphs of the two estimations is shown in Figure 1. Figure 1 indicates that $\hat{M}_L(x)$ fits the data better than $\hat{M}_{RNW}(x)$.

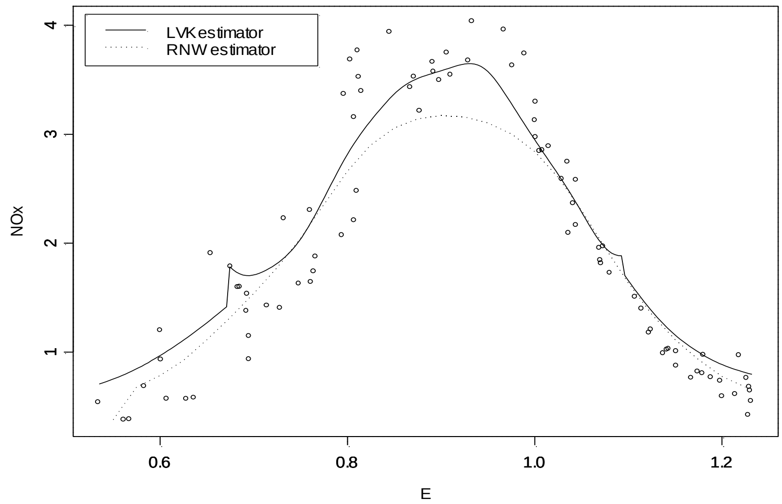

Figure 4.1: (Ethanol Data) Comparison between $\hat{M}_L(x)$ and $\hat{M}_{RNW}(x)$

4.3.2. Simulation Study

A direct comparison between the two estimators is now given using a simulation data. 200 data points were simulated from the model $y = \sin 2\pi(1-x)^2 + xe$, where $e \approx N(0,1)$, and $x \approx uniform[0,1]$. A perfect smooth would recapture the original signals $y = \sin 2\pi(1-x)^2$. A scatter plot of the simulated data together with the graphs of the perfect smooth and the two estimations are shown in Figure 2. Also, the performance of the two estimations is tested using the mean square error (MSE), which is defined as

$$\text{MSE} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}$$

where $\hat{y}$, y stand for the predicted values and the actual values respectively and n is the size of the data. The results of this comparison indicates that MSE($\hat{M}_L(x)$)= 0.0145 and MSE($\hat{M}_{RNW}(x)$)=0.0335. This comparison indicates that the $\hat{M}_L(x)$ estimator is better than $\hat{M}_{RNW}(x)$ estimator.

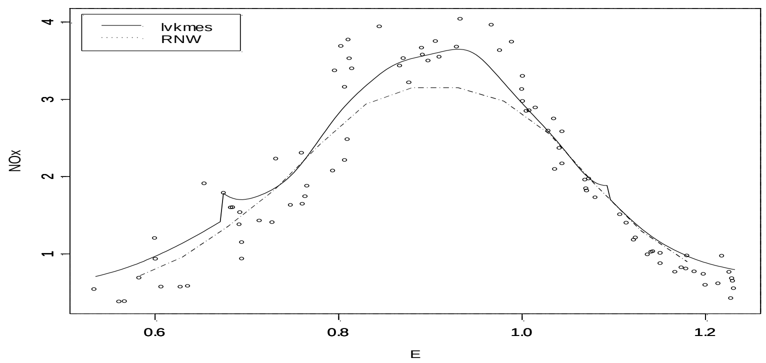

Figure 4.2 (Simulation) Comparison between $\hat{M}_L(x)$ and $\hat{M}_{RNW}(x)$ $\hat{M}_{RNW}(x)$

5. Conclusion and Future Work

The kernel estimator depends on the bandwidth h and the kernel function, we presented the Reweighted Nadaraya Watson estimator as a refinement of Nadaraya Watson estimator, by reweighing the probability weights $\tau_i(x)$.

The choice of bandwidth depended on n and the location and called local variable bandwidth.

Re-Weighted Nadaraya-Watson, $\hat{M}_{RNW}(x)$ and the local variable kernel estimator, $\hat{M}_L(x)$ are compared theoretically and practically. The comparison indicates that $\hat{M}_L(x)$ is better than $\hat{M}_{RNW}(x)$, and that's refers to the strong effect of the bandwidth h(x) as a smoothing parameter in the local variable estimation.

Several simulated and real data applications were used to prove the accuracy of the results obtained by the Reweighed Nadaraya-Watson estimator and Local Variable Kernel estimators.

In future research of these topics, it may prove fruitful to derive such refined estimators using dependent data, and different types of kernel function formulas.

Also, more work may be done to see if it is possible to obtain a refined estimator by combining the advantages the two previous improvements of the (RNW) and (lvke) estimators in a one new estimator.

References:

1. Cheng, M.-Y., L. Peng, and J.-S. Wu (2007): "Reducing Variance in Univariate Smoothing," The Annals of Statistics, 35(2), 522–542.

2. De Gooijer, J.and Zerom, D. (2003). On Conditional Density Estimation. Statistical Neerlandica, Vol. 57, 159-176.

3. Fan J. and Yao Q. (2003), Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Velage

4. Fan, j. and I. Gijbles (1996), Local Polynomial Modeling and its Applications. London: Chapman and Hall.

5. Fan, J. and Q.W. Yao (2005), Nonlinear Time Series: Nonparametric and Parametric Methods. Springer.

6. Fan, J., and T. H. Yim, (2004). A crossvalidation method for estimating conditional densities. Biometrika, Vol.91, 819-834

7. Hall, P., J. S. Racine, and Q. Li (2004), Cross-validation and the estimation of conditional probability densities. Journal of the American Statistical Association. Vol. 99, (468), 1015-1026

8. Hyndman R. J. and Q. Yao, (2002), Nonparametric Estimation and Symmetry Tests for Conditional Density Functions, *Journal of Nonparametric Statistics*, Vol. 14, Issue 3, 259 – 278

9. Hyndman R.J., D. Bashtannyk, G.K.Grunwald (1996), Estimating and visualizing conditional densities. Journal of Computational and Graphical Statistics, Vol.5,315-336.

10. Nan Yang, et al ,(2019**)** Adaptive Nonparametric Kernel Density Estimation Approach for Joint Probability Density Function Modeling of Multiple Wind Farms**,** Energies 2019, 12, 1356; doi:10.3390/en12071356 www.mdpi.com/journal/energies

11. Raid B. Salha1 & Hazem I. El Shekh Ahmed**, (2015)** Reweighted Nadaraya-Watson Estimator of the Regression Mean**,** International Journal of Statistics and Probability; Vol. 4, No. 1

12. Rosenblatt, M. (1956), Remarks of some nonparametric estimates of a density function, Annals of Mathematical Statistics, Vol. 27, 642 – 669.

13. Salha, R. and Ioannides, D. (2004). Joint asymptotic distribution of the estimated conditional mode at a finite number of distinct points. The national statistical conference, 14-18 Aprile, Lefkada, Greece.

14. Samanta, M., and Thavanesmaran, A. (1990). Nonparametric estimation of the conditional mode. Communications in Statistics. Theory and Methods, Vol. 19, 4515-4524.

15. Wand M. and Jones M., (1995), Kernel Smoothing. Chapman and Hall. ISBN-0:412552701.