IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 15, Issue 3 Ser. I (May – June 2019), PP 55-62
www.iosrjournals.org

Efficiency of various Bandwidth Selection Methods across Different Kernels

Kimari Florence [(1)], AdemAggrey[(1)], Kiti Leonard [(2)]
[1](Department of Mathematics and Physics, Technical University of Mombasa, Kenya)
[2](Department of Mathematics and Computer Science, Pwani University)
Corresponding Author: Kimari Florence

Abstract: *In statistics, it is common to come across data from a population whose distribution is not known. In such statistics, non-parametric estimation becomes a useful alternative since it does not make assumptions about the distribution of the data. One of themethodsused in non-parametric estimation is the kernel density estimation. Kernel density estimation involves use of kernels to estimate the density of random variables. Kernels are functions that have satisfied the particular properties of a Probability Distribution Function(PDF) that are explained in this study.Popular kernels featured in the research wereGaussian, Epanechnikov, Biweight and Triweight. Theobjective of thestudywas todetermine the most efficient Bandwidth selection method across different kernels. Identifying the optimal Bandwidth among the Bandwidth selection methodswas a problem that this study aimed to address. The methods of Bandwidth selectionused were least squares cross-validation (LSCV), biased cross-validation (BCV), direct plug-in (DPI) and Polasky and Baker plug-in (PBPI). Random samples of size25, 50, 75, 100, 125 and 150 were generated from normal, binomial, Poisson and uniform distributions using R software. Efficiency of eachBandwidth selection methodwas obtained through the Mean Integrated Square Error(MISE).The findings showed that DPI was the most optimal Bandwidth selector method. The results of this research shall be used by other mathematicians in building non-parametric statistical models to address societal needs.*

Date of Submission: 08-05-2018 Date of acceptance: 24-05-2018

I. Introduction

Non-parametric estimation is a data modelling procedure that allows one to make inferences of a population distribution based on a data sample $x_1, x_2, x_3, ..., x_n$ (Eidous, 2010). Parametric methods assume that the sample comes from a known population distribution, but non-parametric estimation methods make no assumptions (Mugdadi&Jetter, 2010). All the interpretations are made based on the sample statistics, in order to come up with a model that best describes the population. In this perspective, non-parametric estimation becomes more superior to parametric estimation. Non-parametric methods can reveal some aspects of data features that could have passed unnoticed if parametric methods or measures of central tendency were used (Wilcox, 2004).The kernel density estimation is a non-parametric method of estimating the pdf of a random variable whose distribution is unknown.

The kernel function is a pdf, hence symmetrical about zero on the horizontal axis. Its first moment is zero and the second moment is greater than zero (Silverman, 1986;Harpole, 2014). The components of the pdf are the kernel, k, and the Bandwidth, h.(Silverman, 1986, Sheather& Jones, 1991;Harpole, 2014).The kernel functions mostly considered are Gaussian, Epanechnikov,Biweight, Tricube and Triweight(Harpole, 2014). Bandwidths that have gained popularity are the cross-validation methods; plug in methods and rules of thumb.

A Bandwidth is a data smoothing parameter, which controls the smoothing of the kernel. It is the one responsible for the appearance of the bumps in a kernel. If the Bandwidth is large, the kernel is smooth in appearance. A large value of h results in a large bias, a low variance and over-smoothing of the curve. This is what causes some of the features to be concealed (Zambom, 2013). A small value of h results in a low bias and an increased variance which results in a spiky curve. The spiky curve is not appealing, though it brings out most of the details in the data. A bias-variance trade-off needs to be considered in the choice of k and h (Silverman, 1986; Bert, 1992).

A Bandwidth chosen using various Bandwidth selection methods helps in the accurate choice of the estimator (Mugdadi&Jetter, 2010). One of the most popular and traditional method is least squares cross validation technique (LSCV) which was proposed by Rudemo (1982) and Bowman (1984). Scott & Terrell (1987) initiated the use of biased cross-validation technique (BCV) to choose the Bandwidth. Silverman (1986) introduced Silverman's rule of thumb (SROT)which was a modification of Normal Rule of Thumb (NROT). Plug-in-methods (PI) and solve the equation methods were suggested by Woodroofe (1970) and later by

Parkand Marron (1990). LSCV is the most studied one but BCV and PI methods have been proved to perform better than LSCV(Park & Marron, 1990).

These Bandwidth selection methods were recommended by researchers instead of subjectively selecting a Bandwidth through trial and error methods. These methods gave optimal values of the smoothing parameter, h, that was tested using ISE, MISE, AMISE or any other error measures. Hence many authors have studied this area of choosing the optimal Bandwidth using various methods of Bandwidth selection (Sheather& Jones, 1991; Wand & Jones, 1995;Chiu, 1996; Eidouset al., 2010). Choosing the most efficient Bandwidth is critical in order to get a good estimate (Silverman, 1986; Sheather, 2004;Harpole, 2014).

II. Material and Methods

Kernel Density Estimators

The true density, $f(x)$, of a random variable X can be written as:

$$f(x) = \lim_{h \to 0} \frac{1}{2h} p(x-h < X < x+h) \quad \text{......1.1}$$

for each value of x.This f(x) is estimated by $\hat{f}(x)$ which is given by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k\left(\frac{x-x_i}{h}\right) \quad \text{......1.2}$$

Wherek is the kernel, h is the Bandwidth and n is the sample size.

By replacing $z = \left(\frac{x-x_i}{h}\right)$, in the equation (1.2), the function becomes:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} k(z) \quad \text{......1.3}$$

This estimate depends on the parameters, k, h and n. Kernels, have the following properties:

$$\int k(z)dz = 1 \quad \text{......1.4}$$

$$\int zk(z)dz = 0 \quad \text{......1.5}$$

$$\int z^2 k(z)dz = \alpha_2 k > 0 \quad \text{......1.6}$$

The kernel is also symmetrical as k(-z) = k(z) (silverman,1986). Some of the kernel weights commonly used in kernel density estimation aregiven by Table 1.

Table no1: Kernels and their respective Density Functions

Kernel	Kernel weight, k(z)	R(k)= $\int k^2(z)dz$	$\alpha_2 k = \int z^2 k(z)dz$
Epanechnikov	$\frac{3}{4}(1-z^2), \quad \lvert z\rvert<1$ 0 , otherwise	$\frac{3}{5}$	$\frac{1}{5}$
Gaussian	$\frac{1}{(2\pi)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} z^2\right)$	$\frac{1}{2\sqrt{\pi}}$	1
Biweight	$\frac{15}{16}(1-z^2)^2, \quad \lvert z\rvert<1$ 0 , otherwise	$\frac{5}{7}$	$\frac{1}{7}$
Triweight	$\frac{35}{32}(1-z^2)^3, \quad \lvert z\rvert<1$ 0 , otherwise	$\frac{350}{429}$	$\frac{1}{9}$

ISE, MISE and AMISE

In non-parametric methods, an error is created when $\hat{f}(x)$ is used to estimate this truedensity, $f(x)$. The error is given by the difference between the true density and the estimate. When the error is squared, and then integrated, ISE is the resulting value. ISE is used in finding some methods of Bandwidth selection shown in equation 1.14. The expected value of ISE is called the mean integrated square error (MISE) as in equation 1.7. Therefore,MISE is defined as follows:

$$MISE = E\int[\hat{f}(x) - f(x)]^2 \quad \ldots\ldots 1.7$$

The expanded equation of MISE gives the integrated square bias and the integrated variance. The bias is calculated as follows:

$$E\,\hat{f}(x) = f(x) + \frac{1}{2}h^2\alpha_2 k f''(x) \quad \ldots\ldots 1.8$$

where the second term is the bias of $f(x)$. The variance is given by:

$$\mathrm{var}\,\hat{f}(x) = \frac{1}{nh} f(x)\int k^2(z)dz \quad \ldots\ldots 1.9$$

Thus, the sum of integral of the square bias and the integral of the variance yields the MISE.

$$\mathrm{MISE} = \int[\frac{1}{2}h^2\alpha_2 k f''(x)]^2 + \frac{1}{nh}\int f(x)k^2(x)dx \quad \ldots\ldots 1.10$$

where $\alpha_2 k$ is the variance of the kernel and f''(x) is the curvature of the density at point x.

As $n \to \infty$ and $h \to 0$ such that $nh \to \infty$, MISE becomes Asymptotic Mean Integrated Square Error (AMISE) which is defined as

$$AMISE = \frac{1}{4}h^4\alpha_2 k^2 R(f'') + \frac{1}{nh}R(k) \quad \ldots\ldots 1.11$$

WhereR(k)= $\int k^2(x)dx$ and $\alpha_2 k$ is the variance for each kernel as given in Table 1. The unknown $R(f'') = \int (f'')^2 dx$ is a measure of the curvature of the density (Sheather, 2004). By differentiating equation 1.11 with respect to h and equating to zero, the result is h_{AMISE}.

$$h_{AMISE} = \left[\frac{R(K)}{\alpha_2(k)^2 nR(f'')}\right]^{\frac{1}{5}} \quad \ldots\ldots 1.12$$

The MISE andh_{MISE}were calculated for each Bandwidth method using different sample sizes in order to find the relative efficiency of each Bandwidth selector method.Relative efficiency, *eff*(*h*), was calculated as:

$$eff(\hat{h}) = \frac{MISE(h_0)}{MISE(\hat{h})} \quad \ldots\ldots 1.13$$

Each Bandwidth was compared with the LSCV Bandwidth, denoted as h_0, since it was used as the default Bandwidth in R packages. The most optimal Bandwidth was one whose MISE value was smaller than that of the LSCV method, resulting in an efficiency, $eff(\hat{h})$, greater than one.

Bandwidth Selection Methods

The optimal Bandwidth, h_{MISE}, is chosen when MISE is minimized. Since the true density is unknown, ways of calculating the optimal Bandwidth have been proposed by many authors. The methods discussed in this researchcomprised of least squares cross-validation(LSCV), biased cross-validation (BCV), Direct plug-in(DPI) and Polasky and Baker plug-in (PBPI).

Least Squares Cross-Validation

The least squares cross-validation (LSCV), also called unbiased cross-validation, involves the integrated square error, ISE(Rudemo,1982 , Bowman, 1984, Borrajo et al, 2017). The error comes from the difference between the true density, $f(x)$, andthe density estimator, $\hat{f}(x)$. ISE is the result of integrating the square of the error.

$$ISE_h = \int \{ \hat{f}(x) - f(x) \}^2 dx \quad \text{……………………………………………………………}1.14$$

The $f(x)$ can be estimated by an unbiased estimator which involves using $(n-1)$ data points by leaving out the x_i value such that

$$\hat{f}_{-i}(x_i) = \frac{1}{(n-1)h} \sum_{j=1, j\neq i}^{n} k\left(\frac{x - x_j}{h} \right) \quad \text{…………………………………………………..}1.15$$

$$LSCV_h = \int \hat{f}(x)^2 dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}(x_i) \quad \text{……………………………………………………}1.16$$

Minimizing the $LSCV_h$function makes the MISE function minimum so as to get h_{LSCV}(Scott & Terrell,1987; Wand & Jones, 1995; Sheather, 2004).

Biased Cross-Validation

This method is an improvement of LSCV. Unlike LSCV that uses ISE, Biased cross-validation (BCV) involves minimizing the Asymptotic MISE. To do so, the unknown estimate $R(f'')$ in AMISE which depends on the underlying true density is estimated by $\hat{R}(f'')$ (Scott & George, 1987; Jones et al., 1996;Harpoleet al., 2014).In this case, the second derivative of the kernel being used to estimate $f(x)$ is used instead of the unknown second derivative of the underlying true density, $R(f'')$.

$$BCV_h = \frac{1}{nh} R(k) + \frac{1}{4} h^4 \alpha_2 k^2 \hat{R}(f'') \quad \text{…………………………………………………..}1.17$$

The least value that minimizes BCV_hlocally qualifies to be the h_{BCV}(Jones et al, 1996).

Direct Plug-in Method

Direct Plug-in methods (DPI) have been discussed widely (Sheather& Jones 1991; Wand & Jones 1995; Sheather 2004; Harpoleet al, 2014;Varetet al, 2019). The $R(f'')$ in h_{AMISE} is replaced by an estimate by choosing a pilot Bandwidth b to get $R(\hat{f}_b'')$. An initial density estimate, commonly from Gaussian kernel, is used to estimate*h*. This value is plugged into the h_{AMISE} and computed. Then after a series of iterations, which are two or more, the result ish_{DPI} (Zambom& Dias, 2013). The $h_{DPI}$is given by:

$$h_{DPI} = \left[\frac{R(K)}{\alpha_2(k)^2 n R(\hat{f}_b'')} \right]^{\frac{1}{5}} \quad \text{…………………………………………………………….}1.19$$

Polansky and Baker Plug-inMethod

Polansky and Baker(2000) proposed a plug-in method whichuses an initial Bandwidth gin the iterations(Hussein, et al , 2018) .

$$h_{PBPI} = \left(\frac{\eta(k)}{-n\alpha_2^{\,2}(k)\psi_2(g_2)} \right)^{\frac{1}{3}} \quad \text{…………………………………………………………}1.20$$

$$\eta(K) = 2\int_{-\infty}^{+\infty} xK(x)H(x)dx \quad \text{…………………………………………………………....}1.21$$

$$\psi_r(g) = \int_{-\infty}^{+\infty} f^{(r)} f(x)dx \approx \frac{1}{n^2 g^{r+1}} \sum\sum L^{(r)}\left(\frac{x_i - x_g}{g} \right) \quad \text{……………………………...}1.22$$

Simulation and Data Presentation

Simulation is an artificial method of creating data for making inferences instead of using a real data sample.Random samples of size 25, 50, 75, 100, 125 and 150 were be generated from normal, uniform, binomial and Poisson distributions in R software.Normal and Uniform are continuous distributions, while Binomial and Poisson distributions are discrete.

III. Results

Efficiency of BandwidthSelection Methods using Normal Data

From the results of the Bandwidth selectors, DPI method had high efficiency withEpanechnikov, Biweight and Triweight kernels, as in Table 2.However DPI with Gaussian produced very low efficiency. DPI Bandwidths produced wiggly bumps especially using Gaussian kernel as shown in Figure 1. LSCV gave the distribution a smooth curve especially with Gaussian kernel.The most Optimal Bandwidth was the DPI, when Normal data was used. It was followed by PBPI method, then LSCV, and lowest performer was BCV.

Table no2: Efficiency of $\hat{h}$ using Normal data

Normal	sample	LSCV	BCV	DPI	PBPI
Epanechnikov	25	1	0.259323	1.865835	1.195394
	50	1	0.384032	4.372846	1.903808
	75	1	0.320851	4.225239	2.016743
	100	1	0.345750	5.308450	2.327113
	125	1	0.259791	4.067305	2.154894
	150	1	0.205945	3.534605	2.057926
Gaussian	25	1	0.408761	0.324801	0.865374
	50	1	1.010251	0.446627	0.937632
	75	1	1.028547	0.419931	0.912893
	100	1	1.087226	0.502955	0.981441
	125	1	0.995964	0.391021	0.867578
	150	1	0.986089	0.404849	0.822101
Biweight	25	1	0.276188	1.554377	1.131728
	50	1	0.895828	5.512556	2.601693
	75	1	0.939525	6.356082	1.990899
	100	1	1.208417	7.150958	2.188042
	125	1	0.964384	6.559132	2.137574
	150	1	0.790598	5.719086	2.091937
Triweight	25	1	0.344796	3.776865	1.162187
	50	1	0.998316	6.366004	1.704151
	75	1	1.004130	7.594819	1.920563
	100	1	1.228358	8.627121	2.255285
	125	1	1.008789	8.882262	2.183052
	150	1	0.749534	7.007179	1.952000

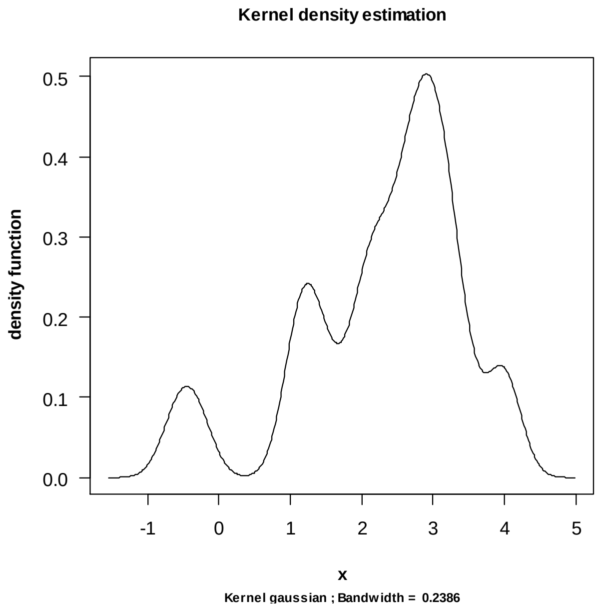
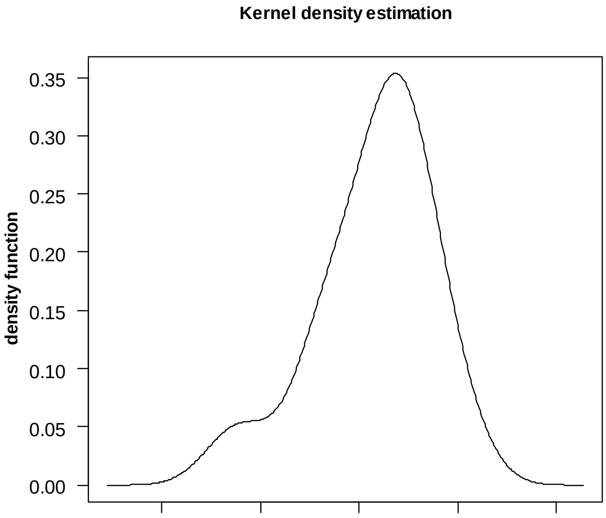


Figure no1:DPI with Gaussian Kernel, usingnormal data n=25, h= 0.236



Figureno 2: Gaussian Kernel, LSCV using normal data n=25, h=0.628

Efficiency of Bandwidth Selection Methodsusing Binomial Data

PBPI was optimal at sample sizes of 50, 100, 125 and 150 while at n=25 and n=75, DPI was optimal when used with Triweight kernel. BCV was optimal with Gaussian kernel in all sample sizes except at n=25 when PBPI was optimal as in Table 3. DPI and PBPI were the most optimal Bandwidth selectors when Biweight kernel was used. Then when Triweight was used, DPI was optimal at sample sizes 25, 75, 100 and 125. PBPI was then optimal at 50 and 150 sample sizes.

The most optimal Bandwidth selector method for the Binomial data was PBPI method. PBPI was followed by DPI, then BCV and lastly LSCV. In research studies, plug-in methods performed better than the cross-validation methods (Sheather& Jones, 1991).

Tableno3: Efficiency of $\hat{h}$ using Binomial data

Binomial	sample	LSCV	BCV	DPI	PBPI
Epanechnikov	25	1	0.271962	2.04173	0.69221
	50	1	0.406971	2.26313	3.37815
	75	1	0.478978	4.76460	3.94058
	100	1	0.439625	4.12934	5.72991
	125	1	0.485359	5.02904	7.73257
	150	1	0.506906	5.46020	9.09538
Gaussian	25	1	1.055414	0.43956	1.31828
	50	1	10.47727	2.09696	3.04026
	75	1	9.618486	2.99271	8.68318
	100	1	8.969581	2.47452	6.47083
	125	1	8.856780	2.46051	5.66450
	150	1	8.412354	2.35722	4.34284
Biweight	25	1	0.672366	6.59465	1.49302
	50	1	0.604064	3.94308	5.70131
	75	1	0.785633	7.92427	3.77270
	100	1	0.607181	7.07686	6.03642
	125	1	0.765770	8.65072	8.84154
	150	1	0.759204	9.36811	12.6826
Triweight	25	1	0.927407	7.43981	4.25357
	50	1	0.867301	5.62751	7.69795
	75	1	1.089805	10.6101	3.42110
	100	1	0.858644	9.87447	5.58596
	125	1	1.182552	12.0975	8.47698
	150	1	1.202683	13.1095	13.8640

Efficiency of Bandwidth Selection Methods using Poisson Data

In Poisson data, when Bandwidth methods were compared with LSCV, PBPI was optimal when Epanechnikov kernel was used. BCV was optimal when Gaussian kernel was used except when n=25, when PBPI had a higher efficiency. DPI was optimal using Biweight and Triweight kernels, except at n=50 when PBPI was optimal.Therefore, the Bandwidth selector with highest efficiency was DPI for Poisson data. DPI was followed by PBPI, then BCV and finally LSCV.

Tableno 4: Efficiencyof $\hat{h}$ using Poisson data

Poisson	sample	LSCV	BCV	DPI	PBPI
Epanechnikov	25	1	0.12107	0.64039	1.48204
	50	1	0.19485	1.34784	3.16844
	75	1	0.22204	1.61619	3.68055
	100	1	0.25326	2.09454	4.08807
	125	1	0.21645	2.70761	3.86848
	150	1	0.18240	3.78784	4.22832
Gaussian	25	1	0.88434	0.34389	1.03240
	50	1	10.8566	4.05083	10.4289
	75	1	10.3293	3.42805	3.17485
	100	1	9.91906	3.07787	9.55401
	125	1	9.92458	2.80774	8.59837
	150	1	9.67671	2.69459	7.69452
Biweight	25	1	0.24468	2.37968	0.66423
	50	1	0.99994	0.93334	2.52794
	75	1	0.56479	6.20669	1.60456
	100	1	0.60713	7.09975	2.05174
	125	1	0.73729	6.90131	2.67122
	150	1	0.73121	7.54467	3.78404
Triweight	25	1	0.69561	6.29925	1.39409
	50	1	0.75380	6.19863	1.33635
	75	1	0.83594	8.24595	1.54540
	100	1	0.87006	9.75962	1.93622
	125	1	0.90255	9.79133	2.49516
	150	1	1.18413	10.7664	3.50534

Efficiency of Bandwidth Selection Methods using Uniform data

Tableno 5: Efficiency of $(\hat{h})$ using Uniform Data

Uniform	sample	LSCV	BCV	DPI	PBPI
Epanechnikov	25	1	0.28966	2.31579	0.94394
	50	1	0.35695	4.05997	1.49352
	75	1	0.18178	2.45576	0.93219
	100	1	0.12809	1.97297	1.03040
	125	1	0.10892	1.83497	0.87134
	150	1	0.34622	6.23204	2.74453
Gaussian	25	1	0.63203	0.36818	0.97036
	50	1	0.82591	0.46180	1.03161
	75	1	0.84885	0.53945	1.00159
	100	1	0.84885	0.53945	1.00159
	125	1	0.87010	0.48561	0.90171
	150	1	0.96401	0.46875	0.93052
Biweight	25	1	0.46371	3.33573	0.93397
	50	1	0.63401	5.51341	1.44092
	75	1	0.69904	5.75585	2.86744
	100	1	0.23061	2.58779	0.98761
	125	1	0.22959	2.61538	0.86708
	150	1	0.83558	8.21351	2.41573
Triweight	25	1	0.51324	4.18401	0.98702
	50	1	0.69141	6.43117	1.46488
	75	1	0.69711	5.77724	1.57031
	100	1	0.27266	3.14815	1.05107
	125	1	0.29801	3.46881	0.97622
	150	1	0.85990	9.30216	2.24739

When the efficiencies of the Bandwidth selectors were compared, LSCV was followed closely by PBPI when Gaussian kernel was used. DPI was most efficient when using Epanechnikov, Biweight and Triweight kernels. Hence the most optimal Bandwidth method in Uniform data was DPI followed by PBPI, LSCV and finally BCV.

IV. Discussion

The objective of the study was to determine the most efficient Bandwidth selection method across different kernels. In Normal Distribution, DPI Bandwidth selector was the most optimal, especially with Biweight and Triweight kernels.In Binomial distribution, PBPI Bandwidth selector was optimal especially with Epanechnikov kernel.In Poisson distribution, best Bandwidth methodwasDPI. In Uniform distribution, optimal Bandwidth was DPI. DPI had higher efficiencies with Biweight and Triweight kernels compared to its performance with other kernels. The DPI Bandwidths were small; the MISE values were low, resulting in high efficiency. The plug-in methods produced higher efficiencies than the cross-validation methods. In other researches, DPI is said to have a higher convergence rate and a higher consistency than LSCV selector method. The optimal Bandwidth depended on the kernel, the sample size and also on the true density.

V. Conclusion

All in all, DPI Bandwidth selection method was the most optimal method in the study. It was followed by PBPI method. Hence DPI qualified to be a universal method in Bandwidth selection.

References

[1]. Bert, V.E.(1992). Asymptotics for least squares cross-validation Bandwidths in non- smooth cases. The Annals of Statistics,20(3), 1647–1657.

[2]. Borrajo, M. I., González-Manteiga, W. &Martínez-Miranda, M. D. (2017)Bandwidth selection for kernel density estimation with length-biased data, Journal of Non-parametric Statistics, 29(3), 636–668.

[3]. Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71(2), 353–360.

[4]. Chu, C.Y., Henderson, D. J.&Parameter, C. F. (2015).Plug-in Bandwidth selection for kernel density estimation with discrete data.Econometrics,3, 199–214.

[5]. Chiu, S.T.(1996).A comparative review of Bandwidth selection for kernel density estimation. StatisticaSinica,6,129–145.

[6]. Eidous, O. M., Marie, M. A., Ebrahem, M. H. (2010).A comparative study for bandwidth selection in kernel density estimation.Journal of Modern Applied Statistical Methods: 9(1), Article 26.

[7]. Harpole, J. K., Woods,C. M., Rodebaugh, T. L., Levinson, C. A., Lenze,E. J.(2014). HowBandwidth selection algorithms impact exploratory data analysis using kernel density estimation.Psychological Methods, 19(3), 428–443.

[8]. Hussein, K, Nehme, B. and Strauss, O. (2018).Interval Estimation of Value-at-Risk Based on Nonparametric Models. Journal of econometrics,6(4), 47.

[9]. Jones, M.C., Marron, J.S., Sheather S.J. (1996). A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association, 91, 401–407.

[10]. Marron, J. S., Wand, M. P. (1992). Exact mean integrated squared error. The Annals of statistics, 20(2), 712– 736.

[11]. Mugdadi,A. R., Jetter,J. (2010).A simulation study for the Bandwidth selection in the kernel density estimation based on the exact and the Asymptotic MISE.Pak. J. Statist, 26(1), 239–265.

[12]. Park, B.U., Marron, J.S. (1990). Comparison of data-driven Bandwidth selectors.Journal of the American Statistical Association, 85, 66–72.

[13]. Polansky, A.M., Baker, E.R. (2000). Multistage plug-in bandwidth selection for kernel distribution function estimates, Journal of Statistical Computation and Simulation, 65, 63–80.

[14]. Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. Annals of MathematicalStatistics, 3, 832–837.

[15]. Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators.ScandinavianJournal of Statistics, 9, 65–78.

[16]. Scott, D.W. & George, R. T. (1987). Biased and unbiased cross-validation in density estimation. Journal of the American Statistical Association, 82, 1131–1146.

[17]. Scott, D. & Terrell, G. (1987). Biased and unbiased cross-validation in density estimation.Journalof the American Statistical Association, 82(400), 1131–1146.

[18]. Sheather, S.J. (2004) .Density estimation. Statistical Science, 19, 588–597.

[19]. Sheather, S. J. & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical Society, 53, 683–690.

[20]. Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis.Chapman and Hall, London.

[21]. Varet. S., Lacour. C., Massart. P., Rivoirard. V.(2019). Numerical performance of Penalized comparison to Overfitting for multivariate kernel density estimation.Retrieved from https://hal.archives-ouvertes.fr/hal-02002275.

[22]. Wand, M.P , Jones, M.C.(1995).Kernel Smoothing. Chapman and Hall, London.

[23]. Wilcox, R. (2004). Kernel density estimators: An approach to understanding how groups differ. Understanding Statistics, 3(4), 333–348.

[24]. Woodroofe, M. (1970). On choosing a delta-sequence. The Annals of Mathematical Statistics, 41,1665–1671.

[25]. Zambom, A. Z., Dias, R. (2013). A review of kernel density estimation with applications to econometrics.IntEconom Rev (IER), 5(1), 20–42.

Kimari Florence. "Efficiency of various Bandwidth Selection Methods across Different Kernels." IOSR Journal of Mathematics (IOSR-JM) 15.3 (2019): 55-62.