ESTIMATION OF EFFICIENCIES OF THE BIVARIATE DERIVATIVES OF SMOOTH POLYNOMIAL KERNELS.

by

Siloko, I. U.; Ikpotokin, O. and Siloko, E. A.[3]

Department of Mathematics and Computer Science, Edo University Iyamho, Nigeria. & Department of Mathematics and Statistics, Ambrose Alli University, Ekpoma, Nigeria. & Department of Mathematics, University of Benin, Benin City, Nigeria.

Corresponding Author: siloko.israel@edouniversity.edu.ng, suzuazor@yahoo.com

Abstract

The bivariate kernel estimator bridges the gap between the univariate kernel and higher dimensional kernels in density estimation. The efficiencies of the univariate kernels have received considerable attention unlike their bivariate counterparts due to the "curse of dimensionality" effect. In this paper, our focus is on the efficiencies of the derivatives of the bivariate kernels of the Uniform, Epanechnikov, Biweight, Triweight, Quadriweight and Gaussian kernels which are members of the beta polynomial family. The bivariate form of these kernel functions were obtain from their univariate counterpart using the product approach. The results obtain shows that the efficiency of the kernels decrease as the powers of the polynomial increases and tends to increase as the derivative order increases.

Key words: *Kernel Density Derivatives, Smooth Polynomial Kernels, Efficiency of Kernels, Asymptotic Mean Integrated Squared Error.*

1. Introduction.

Density estimation is a fundamental data analysis technique in statistics and probability theory. Kernel density estimation is a nonparametric estimation technique with wider applications due to the simplicity of its implementation (Schauer *et al.*, 2013). As a nonparametric estimator, it is mainly for data exploration and visualization purposes but its application has been extended to the machine learning community. In Härdle *et al*. (2004), nonparametric estimation is viewed as the building blocks for different semiparametric estimators where the separability ideology of the independent variables in semiparametric model is in line with the devolution of decision making process in organizations or stages of production in industries in real life situation. Kernel density estimation can also be applied indirectly to other areas of nonparametric estimation such as discriminant analysis, goodness-of-fit testing, hazard rate estimation, bump-hunting, intensity function estimation, and classification with regression estimation. These other areas where kernel density estimation can be applied are contained in Raykar *et al*. (2015).

Kernel derivatives are of significant applications such as locating the local extrema and identification of the point of inflexion of a distribution. Chacon and Duong (2013) investigated the statistical properties of some distributions with kernel density derivative such as location of point of inflexion while in time series analysis, Rondonotti *et al.* (2007) applied kernel density derivative with related data and the results obtained was outstanding. The estimation of the optimal smoothing parameter in kernel density estimation requires the derivative of the unknown probability function and Silverman (1986) suggested that a certain value should be used in the case of the normal kernel. Other data analysis where kernel density derivative can be used include human growth data analysis (Ramsay and Silverman, 2002), investigation of nanoparticles property of data (Charnigo and Srinivasan, 2011) and chemical compositions inferences (De Brabanter *et al.*, 2011). Finally, in parameter

estimation and hypothesis testing, density derivatives has played a significant role, therefore proper estimation of the density derivatives is very important (Sasaki *et al*., 2015).
The limitation of the kernel density estimator is the difficulty of selecting the accurate smoothing parameter. In univariate kernel estimation, the problem of smoothing parameter selection is with less complexity compare with the multivariate setting where there are different forms of smoothing parameterizations. The choice of smoothing parameter is also very important in kernel density derivatives particularly as the order of the derivative to be estimated increases. In Siloko *et al*. (2018), two gradient methods of selecting smoothing parameter in the bivariate kernel density estimation were proposed and the results outperformed the popular cross validation selectors.

2. The Kernel Density Derivatives.
The kernel estimator is a nonparametric density estimator and is a vital tool in statistical data analysis. The univariate kernel estimator is of the form

$$\hat{f}(\mathrm{x}) = \frac{1}{nh_{\mathrm{x}}}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right), \tag{1}$$

where K is the kernel function, $h_{\mathrm{x}} > 0$ is the smoothing parameter and n is the sample size. In most cases particularly in scientific computing and data intensive applications, the data set X_i are observations or measurements obtained from real life. The kernel function is a non-negative function that satisfies the conditions

$$\begin{cases} \int K(\mathrm{x})d\mathrm{x} = 1, \\ \int \mathrm{x}K(\mathrm{x})d\mathrm{x} = 0 \quad \text{and} \\ \int \mathrm{x}^2 K(\mathrm{x})d\mathrm{x} = k_2(K) > 0 \,. \end{cases} \tag{2}$$

The first condition in Equation (2) means that any weighting function must integrate to unity, hence most kernel functions are probability density functions; the second condition simply states that the average of the kernel is zero, while third condition means that the variance of the kernel is not zero (Scott 1992).
The bivariate kernel density estimator occupies a unique position of bridging the univariate kernel density estimator and other higher dimensional kernel estimators. The usefulness of the bivariate kernel density estimator is mainly in its simplicity of presentation of probability density estimates, either as surface plots or contour plots and serving as the bedrock for understanding other higher dimensional kernel estimators. In bivariate kernel density estimation, $\mathrm{x}\,,\mathrm{y}$ is taken to be the two random variables with a joint probability density function $f(\mathrm{x}\,,\mathrm{y})$. The random variables $\mathrm{X}_i, \mathrm{Y}_i,\ i = 1,2, \dots, n$ are the set of observations and n is the sample size. The bivariate kernel density estimate of $f(\mathrm{x}\,,\mathrm{y})$ is of the form

$$\hat{f}(\mathrm{x}\,,\mathrm{y}) = \frac{1}{nh_{\mathrm{x}}h_{\mathrm{y}}}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}, \frac{\mathrm{y}-\mathrm{Y}_i}{h_{\mathrm{y}}}\right), \tag{3}$$

where $h_{\mathrm{x}} > 0$ and $h_{\mathrm{y}} > 0$ are the smoothing parameters in the X and Y axes and $K(\mathrm{x}\,,\mathrm{y})$ is a bivariate kernel function. The bivariate kernel density estimator in Equation (3) can also be written as the product of two univariate kernel as

$$\hat{f}(\mathrm{x}\,,\mathrm{y}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_{\mathrm{x}}} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)\frac{1}{h_{\mathrm{y}}} K\left(\frac{\mathrm{y}-\mathrm{Y}_i}{h_{\mathrm{y}}}\right). \tag{4}$$

The asymptotic mean integrated squared error is the commonest optimality criterion function used in measuring the performance of smoothing parameter in kernel estimation. Applying Taylor's series expansion in Equation (1) will yield the asymptotic mean integrated squared error (AMISE) which is made up of the asymptotic integrated variance and the asymptotic integrated squared bias given by

$$AMISE\big(\hat{f}(\mathrm{x})\,\big) = \frac{R(K)}{nh_{\mathrm{x}}} + \frac{1}{4}\mu_2(\mathrm{K})^2\, h_{\mathrm{x}}^4 R(f_{\mathrm{xx}}), \tag{5}$$

where $R(K)$ is the roughness of the kernel, $\mu_2(\mathrm{K})^2$ is the second moment of the kernel and $R(f_{\mathrm{xx}}) = \int f_{\mathrm{xx}}(\mathrm{x})^2 d\mathrm{x}$ is the roughness of the unknown probability density function (Scott, 1992; Guidoum, 2015). Similarly, the asymptotic mean integrated squared error of the bivariate kernel estimator is of the form

$$AMISE\left(\hat{f}(\mathrm{x}\,,\mathrm{y})\right) = \frac{R(K)^d}{nh_{\mathrm{x}}h_{\mathrm{y}}} + \frac{h_{\mathrm{x}}^4}{4}\mu_2(\mathrm{K})^2 R(f_{\mathrm{xx}}) + \frac{h_{\mathrm{y}}^4}{4}\mu_2(\mathrm{K})^2 R\big(f_{\mathrm{yy}}\big), \tag{6}$$

where $R(K)^d$ is the roughness of the kernel, $\mu_2(\mathrm{K})^2$ is the second moment of the kernel and $R(f_{\mathrm{xx}}) = \int f_{\mathrm{xx}}(\mathrm{x},y)^2 d\mathrm{x}dy$, $R\big(f_{\mathrm{yy}}\big) = \int f_{\mathrm{yy}}(\mathrm{x},y)^2 d\mathrm{x}d\mathrm{y}$ are the roughnesses of the unknown probability density function.

The derivative of the univariate kernel estimator is obtained by taking the derivative of the kernel density estimator in Equation (1). If the kernel K is sufficiently differentiable r times, then, the rth density derivative of Equation (1) is given by

$$\hat{f}^{(r)}(\mathrm{x}) = \frac{d^r}{d\mathrm{x}^{\mathrm{r}}}\hat{f}(\mathrm{x}) = \frac{1}{nh_{\mathrm{x}}^{r+1}}\sum_{i=1}^{n} K^{(r)}\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right), \tag{7}$$

where $K^{(r)}$ is the rth derivative of the kernel function and K is a symmetric probability density (Scott, 1992). In order for the estimator in Equation (7) to exist, $K^{(r)}$ must exist and not equal to zero. Again the rth density derivative of Equation (4) is given by

$$\hat{f}^{(r)}(\mathrm{x},\mathrm{y}) = \frac{d^r}{d\mathrm{x}^{\mathrm{r}}d\mathrm{y}^{\mathrm{r}}}\hat{f}(\mathrm{x},\mathrm{y}) = \frac{1}{nh_{\mathrm{x}}^{r+1}h_{\mathrm{y}}^{r+1}}\sum_{i=1}^{n} K^{(r)}\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)K^{(r)}\left(\frac{\mathrm{y}-\mathrm{Y}_i}{h_{\mathrm{y}}}\right). \tag{8}$$

The AMISE of the rth derivative of the bivariate kernel function provided the kernel K can be sufficiently differentiated is of the form

$$AMISE^r\left(\hat{f}^{(r)}(\mathrm{x},\mathrm{y})\right) = \frac{R\big(K^{(r)}\big)^d}{nh_{\mathrm{x}}^{2r+1}h_{\mathrm{y}}^{2r+1}} + \frac{1}{4}h_{\mathrm{x}}^4\mu_2(\mathrm{K})^2 R(f_{\mathrm{xx+r}}) + \frac{1}{4}h_{\mathrm{y}}^4\mu_2(\mathrm{K})^2 R\big(f_{\mathrm{yy+r}}\big). \tag{9}$$

where $R\big(K^{(r)}\big)^d$ is the roughness of the rth derivative of the kernel, $\mu_2(\mathrm{K})^2$ is the second moment of the kernel function and $R(f_{\mathrm{xx+r}})$, $R\big(f_{\mathrm{yy+r}}\big)$ are the rth roughnesses of the unknown probability density function.

3. The Beta Polynomial Kernels and Efficiency of Density Derivatives.

The general beta polynomial kernel family for $p \geq 0$ with $\{t \in [-1, 1]\}$ is of the form

$$K_p(t) = \frac{(2p+1)!!}{2^{p+1}p!}(1-t^2)^p\,, \tag{10}$$

where $p = 0,\ 1,\ 2, \dots, \infty$ and the double factorial $(2p+1)!! = (2p+1)(2p-1)\dots 5.3.1.$ As the value of p increases from 0 to 3, we have the Uniform, Epanechnikov, Biweight and

Triweight kernels which are members of the family (Scott, 1992; Hansen, 2005). The popular normal kernel is the limiting case that is when p tends to infinity and the Uniform kernel is the simplest kernel in this family of kernels while the Epanechnikov kernel is regarded as the optimal kernel with respect to an error criterion, the mean integrated squared error. It should be noted that the kernels with higher values of p and their estimates are smoother and also possess more derivatives.
In kernel density estimation, there are two main techniques of obtaining the multivariate kernels from the univariate case. These two popular approaches are the product approach and the spherical approach (Wand and Jones, 1995). The product kernel estimator allows different smoothing parameter to be used for each dimension unlike the fixed kernel that uses a single smoothing parameter. The order of the smoothing parameter that minimizes the AMISE of the product kernel is same as that of the multivariate fixed kernel and the AMISE is also of the same order as that of the multivariate fixed kernel. The advantage of the multivariate product kernel over other forms is that the product approach is beneficial especially when the scales of the variables to be considered differ. Also, in the case of unimodal densities, the product kernel that permits different amount of smoothing for each dimension has been suggested by many authors (Sain, 2002). The product approach uses the product of the marginal univariate kernels and is of the form

$$K_p^{Product}(t) = B^d \prod_{i=1}^{d} (1 - t_i^2)^p \,, \tag{11}$$

where $B^d = \frac{(2p+1)!}{2^{2p+1}(p!)^2}$ is the normalization constant and d is the dimension of the kernel.
The efficiency of a kernel function which is measured in comparison with the Epanechnikov kernel is of the form

$$Eff(K) = \left(\frac{C(K_e)}{C(K)}\right)^{1/4} = \left(\frac{R(K_e)^4 \mu_2(K_e)^2}{R(K)^4 \mu_2(K)^2}\right)^{1/4} , \tag{12}$$

where $C(K) = R(K)^4 \mu_2(K)^2$ is a constant of any given kernel and $C(K_e) = R(K_e)^4 \mu_2(K_e)^2$ is the constant of the Epanechnikov kernel (Silverman, 1986). The Epanechnikov kernel produce the smallest AMISE value in the case of the classical second order kernel and therefore, it is regarded as the optimal kernel with respect to the asymptotic mean integrated squared error.
The efficiency of the kernel derivative also requires the determination of the optimal kernel for its computation. The Epanechnikov kernel cannot be the optimal kernel in kernel density derivatives because its second derivative is a constant meaning that it is not continuously differentiable. Therefore, the efficiency of the rth derivative kernel is given by

$$Eff(K^r) = \left(\frac{C(K_{opt}^r)}{C(K_{r+1})}\right)^{(2r+1)/4} = \left(\frac{R(K_{opt}^r)^{4/(2r+1)} \mu_2(K_{opt})^2}{R(K_{r+1}^r)^{4/(2r+1)} \mu_2(K_{r+1})^2}\right)^{(2r+1)/4} , \tag{13}$$

where $C\left(K_{opt}^r\right) = R(K_{opt}^r)^{4/(2r+1)} \mu_2(K_{opt})^2$ is the optimal kernel for the rth kernel derivative function and $C(K_{r+1}^r) = R(K_{r+1}^r)^{4/(2r+1)} \mu_2(K_{r+1})^2$ is the constant of any given $(r+1)th$ derivative kernel function. Further simplification of Equation (13) resulted in Equation (14) and is of the form

$$Eff(K^r) = \frac{R(K_{opt}^r)}{R(K_{r+1}^r)} \left(\frac{\mu_2(K_{opt})^2}{\mu_2(K_{r+1})^2}\right)^{(2r+1)/4} . \tag{14}$$

The optimal kernel for estimating the rth derivative was shown by Muller (1984) which involves solving for the minimum of $R(K^r)$, subject to the conditions $K_0 = 1$, $K_1 = 0$ and $K_2 < \infty$ and the solution obtained is $p = (r + 1)th$ kernel from the beta polynomial kernels in Equation (10). This implies that when estimating the first derivative $(r = 1)$, the optimal kernel is the Biweight kernel $(p = 2)$ but if we desire to estimate the second derivative $(r = 2)$, then the optimal kernel in this case is the Triweight kernel $(p = 3)$, and the optimality of the kernel functions goes on in that manner. The computation of the efficiency of kernel density derivatives requires the derivative of the rth roughness of the kernel function to be estimated while the second moment of the kernel is not affected irrespective of the derivative order to be estimated.

In computing the efficiency of the kernel derivatives, two very important statistical quantities are the rth roughness of the kernel functions and its second moment as observed in Equation (13) and Equation (14). The rth roughness of a kernel function is given by

$$R(K^r) = \int K^r(t)^2 dt. \tag{15}$$

Also, the second moment of a kernel function is of the form

$$\mu_2(K)^2 = \int t^2 K(t) dt. \tag{16}$$

In computing the statistical properties and efficiencies of the derivatives of the bivariate beta polynomial kernels, we shall specifically consider the Uniform, Epanechnikov, Biweight, Triweight, Quadriweight and Gaussian kernels. The quantities in Equation (15) and Equation (16) are the parameters of interest in the determination of the efficiency of any given kernel function in density estimation.

4. Discussion of Results.

We shall consider the statistical properties of p for which $p = 0,\ 1,\ 2,\ 3,\ 4$ which are the Uniform, Epanechnikov, Biweight, Triweight, Quadriweight kernels and also for $p = \infty$ which is the Gaussian kernel for the bivariate case. The Epanechnikov, Biweight and Triweight are of wide applications because they form the basis when discussing this class of kernels especially the Epanechnikov kernel in the computation of the efficiencies of other kernel functions of this family.

In Table 4.1 the statistical properties of the two dimensional product kernels of the beta polynomials kernels were computed. The bivariate kernel functions of the Uniform, Epanechnikov, Biweight, Triweight, Quadriweight and Gaussian kernels were obtain using the procedure in Equation (11). The efficiency of the bivariate Epanechnikov kernel is 100 % and the efficiencies of other kernel functions of the beta family are less than 100 %. As evident in Table 4.1, the efficiencies decrease with increase in the values of p .

Table 4.1: Bivariate Kernel Functions with Roughnesses, Moments and

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)$%
$K_0(t) = \frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{9}$	0.864	84.6 %
$K_1(t) = \frac{9}{16}\left((1-t_1^2)(1-t_2^2)\right)^1$	$\frac{9}{25}$	$\frac{1}{25}$	1.000	100 %
$K_2(t) = \frac{225}{256}\left((1-t_1^2)(1-t_2^2)\right)^2$	$\frac{127551}{250000}$	$\frac{2551}{125000}$	0.988	98.8 %
$K_3(t) = \frac{1225}{1024}\left((1-t_1^2)(1-t_2^2)\right)^3$	$\frac{665613}{1000000}$	$\frac{6173}{500000}$	0.974	97.4 %
$K_4(t) = \frac{99225}{65536}\left((1-t_1^2)(1-t_2^2)\right)^4$	$\frac{205677}{250000}$	$\frac{1033}{125000}$	0.963	96.3 %
$K_\emptyset(t) = \frac{1}{2\pi}\, exp\left(-\frac{t_1^2+t_2^2}{2}\right)$	$\frac{1}{4\pi}$	1	0.905	90.5 %

Tables 4.2; 4.3 and 4.4 are the efficiencies of the first to the third derivatives of the bivariate kernels while Table 4.5 shows the efficiencies of all the order of the kernel derivatives considered. In all the cases, the efficiencies of the kernel functions decrease as the values of the powers of p increases but increases as the order of the derivative increases. However, the increase in efficiencies with respect to the derivative order starts decreasing immediately after the optimum value for each particular order.

Table 4.2: Bivariate Roughnesses, Moments and Efficiencies of the First Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)\%$
$K_1(t) = \frac{9}{16}\left((1-t_1^2)(1-t_2^2)\right)^1$	$\frac{9}{10}$	$\frac{1}{25}$	0.868	86.8 %
$K_2(t) = \frac{225}{256}\left((1-t_1^2)(1-t_2^2)\right)^2$	$\frac{153061}{100000}$	$\frac{2551}{125000}$	1.000	100 %
$K_3(t) = \frac{1225}{1024}\left((1-t_1^2)(1-t_2^2)\right)^3$	$\frac{259589}{100000}$	$\frac{6173}{500000}$	0.975	97.5 %
$K_4(t) = \frac{99225}{65536}\left((1-t_1^2)(1-t_2^2)\right)^4$	$\frac{399602}{100000}$	$\frac{1033}{125000}$	0.946	94.6 %
$K_\emptyset(t) = \frac{1}{2\pi}\, exp\left(-\frac{t_1^2+t_2^2}{2}\right)$	$\frac{1}{16\pi}$	1	0.785	78.5 %

Table 4.3: Bivariate Roughnesses, Moments and Efficiencies of the Second Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)\%$
$K_2(t) = \frac{225}{256}\left((1-t_1^2)(1-t_2^2)\right)^2$	$\frac{160714}{10000}$	$\frac{2551}{125000}$	0.836	83.6 %
$K_3(t) = \frac{1225}{1024}\left((1-t_1^2)(1-t_2^2)\right)^3$	$\frac{285548}{10000}$	$\frac{6173}{500000}$	1.000	100 %
$K_4(t) = \frac{99225}{65536}\left((1-t_1^2)(1-t_2^2)\right)^4$	$\frac{539463}{10000}$	$\frac{1033}{125000}$	0.967	96.7 %
$K_\emptyset(t) = \frac{1}{2\pi}\, exp\left(-\frac{t_1^2+t_2^2}{2}\right)$	$\frac{9}{64\pi}$	1	0.875	87.5 %

Table 4.4: Bivariate Roughnesses, Moments and Efficiencies of the Third Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)\%$
$K_3(t) = \frac{1225}{1024}\left((1-t_1^2)(1-t_2^2)\right)^3$	$\frac{642483}{1000}$	$\frac{6173}{500000}$	0.815	81.5 %
$K_4(t) = \frac{99225}{65536}\left((1-t_1^2)(1-t_2^2)\right)^4$	$\frac{116884}{100}$	$\frac{1033}{125000}$	1.000	100 %
$K_\emptyset(t) = \frac{1}{2\pi}\, exp\left(-\frac{t_1^2+t_2^2}{2}\right)$	$\frac{225}{256\pi}$	1	0.892	89.2 %

Table 4.5: Bivariate Efficiencies of Second Order Kernels Derivative.

Kernel Functions	Derivative Orders (r)			
S/N	0	1	2	3
$K_0(t)$	0.864			
$K_1(t)$	1.000	0.868		
$K_2(t)$	0.988	1.000	0.836	
$K_3(t)$	0.974	0.975	1.000	0.815
$K_4(t)$	0.963	0.946	0.967	1.000
$K_\emptyset(t)$	0.905	0.785	0.875	0.892

5. Conclusion.

This study investigates the efficiencies of the bivariate kernel density derivatives of the beta polynomial family for some powers of p and the limiting case. Of interest are the Uniform, Epanechnikov, Biweight, Triweight, Quadriweight kernels and the Gaussian kernel for the bivariate case. The results presented in Table 4.5 above shows that the efficiencies of the bivariate kernel functions decreases as the power of p increases. However, the efficiencies tend to increase as the derivative order of the bivariate kernels increases but starts decreasing immediately after the optimum kernel for each derivative order. This simply suggests that the estimates of the higher derivatives of the bivariate kernel function will be smoother than those with fewer derivatives.

References

Chacón, J. E. and Duong, T. (2013). Data-Driven Density Derivative Estimation, with Applications to Nonparametric Clustering and Bump Hunting. *Electronic Journal of Statistics*, 7, 499–532.

Charnigo, R. and Srinivasan, C. (2011). Self-Consistent Estimation of Mean Response Functions and Their Derivatives. *Canadian Journal of Statistics*, 39(2), 280–299.

De Brabanter, K., De Brabanter, J. and De Moor, B. (2011). Nonparametric Derivative Estimation, *NAIC, Gent, Belgium.*

Guidoum, A. S. (2015). Kernel Estimator and Bandwidth Selection for Density and its Derivatives. *Department of Probabilities and Statistics, University of Science and Technology, Houari Boumediene, Algeria*.

Hansen, B. E. (2005). Exact Mean Integrated Squared Error of Higher Order Kernel Estimators. *Econometric Theory,* 21, 1031–1057.

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). Nonparametric and Semiparametric Models. *Springer-Verlag Berlin Heidelberg New York.*

Müller, H. G. (1984). Smooth Optimum Estimators of Densities, Regression Curves and Modes. *Annals of Statistics,* 12, 766–774.

Ramsay, J. O. and Silverman, B. W. 2002. Applied Functional Data Analysis. Springer-Verlag.

Raykar, V. C., Duraiswami, R. and Zhao, L. H. (2015). Fast Computation of Kernel Estimators. *Journal of Computational and Graphical Statistics,* 19 (1), 205–220.

Rondonotti, V., Marron, J. S. and Park. C. (2007). SiZer for time series: A New Approach to the Analysis of Trends. *Electronic Journal of Statistics,* 1, 268–289.

Sain, R. S. (2002). Multivariate Locally Adaptive Density Estimation. *Computational Statistics and Data Analysis*, 39, 165–186.

Sasaki, H., Noh, Y. and Sugiyama, M. (2015). Direct Density Derivative Estimation and Its Application in KL-Divergence Approximation. *Appearing in Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS), San Diego, CA, USA. JMLR: W&CP, Volume 38.*

Schauer, K., Duong, T., Gomes-Santos, C. and Goud, B. (2013). Studying Intracellular Trafficking Pathways with Probabilistic Density Maps. *Methods in Cell Biology,* 118, 325–343.

Scott, D. W. (1992). Multivariate Density Estimation. Theory, Practice and Visualisation. *Wiley, New York.*

Siloko, I. U., Ishiekwene, C. C. and Oyegue, F. O. (2018). New Gradient Methods for Bandwidth Selection in Bivariate Kernel Density Estimation. *Mathematics and Statistics,* 6(1), 1–8.

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

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