Estimation of Density Using Asymmetric Kernels

Patil S. B.

Department of Statistics, Bhogawati Mahavidyalaya Kurukali

In the nonparametric framework, smooth density estimators are constructed using kernel functions. Generally, kernel functions are symmetric. These estimators exhibit boundary bias and may yield positive outside the support of the true density. To overcome these drawbacks we propose to use asymmetric kernels. We consider a symmetric kernel function (referred to as base kernel function) and modify it to obtain suitable asymmetric kernels to obtain two estimators of the density at 'x'. Based on the simulation study, the proposed density estimators perform better than those based on a standard symmetric kernel and those based on different asymmetric kernels.
Keywords: Kernel Density estimation; Boundary bias; Bandwidth function; Asymmetric kernel.

1. Introduction

In the nonparametric framework, simple estimators for a density function include the Histogram estimator and the Naïve estimator. However, these depend on the choice of the bin width and are not smooth. To obtain smooth density estimators kernel functions are used, generally, kernel functions used are symmetric. Such estimators have boundary bias and could be positive outside the support of the true density. This problem is much savior when the densities have higher concentrations near the boundaries. Jones and Foster (1996), introduced simple nonnegative estimators that are corrected for boundary effects. Yin and Hao (2007) have proposed an adaptive kernel estimator by using beta kernels when the support is known to be the unit interval. Their estimator at x depends on k_x a beta kernel with a modal value at x and the estimator is the average of the values of $k_x(X_i), i = 1,2,.,n$. The selection of bandwidth h is more crucial than the choice of the kernel k. Numerous methods for bandwidth selection have been proposed in the literature (e.g., Silverman (1986); Wand and Jones (1994); Rudemo (1982); Bowman (1984)), but many of these methods encounter computational challenges. As an alternative, Zhang and Wang (2009) proposed a robust normal reference bandwidth that adapts to various types of densities. To mitigate boundary bias, Rattihalli and Patil (2019) used suitable asymmetric kernels to account for the contributions of the observations.

A kernel density estimator $\hat{f}$ at 'x', when using a symmetric kernel, assigns identical weights to observations that are equidistant from 'x'. In the case of distributions like 'Exponential with unknown location', if we use a symmetric kernel, it produces considerable positive mass, even outside of support, where the true density is zero. To address the issue of boundary bias to a significant extent, this article proposes the use of asymmetric kernels (defined in (4)), which are derived from a standard symmetric base kernel. The degree of asymmetry depends on the relative position of 'x', (the point where the density is to be estimated) with reference to the data cloud (collection of observations). When the kernel function used to estimate f(x) depends on x are not the same then the resulting function (given in (5)) is not necessarily a density function and hence it is necessary to use normalizing constant C, so that the estimator is the proper density function. Programs in Matlab have been developed to obtain the estimators (7) and (8) for a given choice of k, base kernel, and h, bandwidth. The performance of the estimators is assessed through a simulation study, where they are compared with competing methods.

In Section 2, we define a class of asymmetric kernels that will be used to derive the estimators and provide the motivation for their selection. The estimators are introduced in Section 3. In Section 4, the performance of the estimators is evaluated through extensive simulations, by considering some models. Finally, the conclusions and scope of the study are presented in section 5.

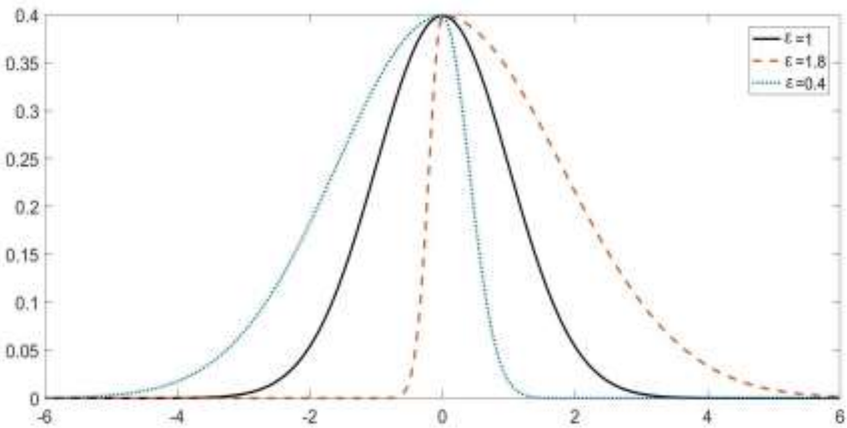


Figure 1: The $k(t, \epsilon)$ functions for $\epsilon = 1, \epsilon = 1.8$ and $\epsilon = 0.4$.

2. Class of asymmetric kernels and motivation:

Let $X_1, X_2, \ldots, X_n$ be ordered observations from a random sample of size n drawn from a distribution with an unknown univariate probability density f. A kernel density estimator for f is,

$$f_u(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{x-X_i}{h}\right) \quad (1)$$

where h is the bandwidth, which controls the smoothness of the estimator and k is the kernel function. The choice of bandwidth h is more crucial than that of the kernel k, which is typically symmetric. A symmetric kernel is adequate for estimating density with unbounded support and it is not so for density with bounded supports, as in this case, the estimator is non-zero outside the support specifically near the boundary of the support. To overcome this problem, one may use different asymmetric kernels depending on points x at which the density is to be estimated and the data set.

We consider a typical class of asymmetric kernels $k(t,\epsilon), 0 < \epsilon < 2$ given by

$$k(t,\epsilon) = \begin{cases} k\left(\frac{t}{\epsilon}\right) & t \geq 0 \\ k\left(\frac{t}{2-\epsilon}\right) & t < 0 \end{cases}$$

The graphs of these densities when k corresponds to normal density and when ϵ = 1, 0.4, and 1.8 are given in Figure 1. Rattihalli and Patil(2019) have recommended the following asymmetric kernel functions from this class, $k_i, i = 1,2,..,n$ to account for the contribution by the observation X_i,

$$k_i(t) = k\left(t, \frac{2(n-i+1)}{n+1}\right) = \begin{cases} k\left(\frac{t(n+1)}{2(n-i+1)}\right) & t \geq 0 \\ k\left(\frac{t(n+1)}{2i}\right) & t < 0 \end{cases}$$

Their estimator is given by

$$\hat{f}(x) = \frac{1}{nh}\left\{\sum_{i=1}^{n} k\left(\frac{(x-X_i)(n+2)}{2h(n-i+1)}\right)[x \geq X_i] + \sum_{i=1}^{n} k\left(\frac{(x-X_i)(n+2)}{2h(i+1)}\right)[x < X_i]\right\} \qquad (2)$$

where $X_1, X_2, \dots, X_n$ are ordered observations. Since for each $i = 1, 2, \dots, n$, $k_i(.)$ is a proper density function, $\hat{f}(x)$ is also a proper density function.

2.1 Asymmetry of the kernel: based on the number of observations on either side of x:

To estimate the density at x we use the kernel $k\,(t,a)$ where a depends on the number of observations on either side of x. Let $X_1, X_2, \dots X_n$ be ordered observations. A natural estimator of the distribution function F is $F_n(x) = \frac{r}{n}$ for $X_r \leq x < X_{r+1}$. To estimate the density at x, $X_r \leq x < X_{r+1}$, one possible choice of a is such that $K\,(0,a) = F_n\,(x)$, which yields $a = \frac{2(n-r)}{n}$. But with this choice of a, density estimator will have the support (X_1, X_n), which is undesirable. To overcome this we presume that apart from n observations there exist two observations X_0 which is sufficiently smaller than X_1 and X_{n+1} which is sufficiently larger than X_n.

Thus for $X_r < x < X_{r+1}$ we have, out of $n + 2$ observations, $r + 1$ observations are smaller than x. Hence to estimate density at x select $k\,(t,a)$ with $a = \frac{2(n-r)}{n}$ and we denote it by k_r. Thus,

$$k_r(t) = \begin{cases} k\left(\frac{t(n+2)}{2(n-r+1)}\right) & t \geq 0 \\ k\left(\frac{t(n+2)}{2r}\right) & t < 0 \end{cases} \quad (3)$$

2.2 Asymmetry of the kernel: based on the exact location of x:

With the convention that $X_0 = \infty$, $X_{n+1} = \infty$, the $n + 2$ points $X_0, X_1, \ldots, X_{n+1}$ form $n + 1$ intervals. We consider $\epsilon(x)$, a non-decreasing function on the extended real line which is a piece-wise linear function given by $\epsilon(x) = \frac{1}{n+1}$ for $x \leq X_1$, $\epsilon(x) = \frac{r+\frac{x-X_r}{X_{r+1}-X\_r}}{n+1}$ for $X_r \leq x < X_{r+1}$, $r = 1,2,\ldots,n-1$, and $\epsilon(x) = \frac{n}{n+1}$ for $x > X_n$. Thus $\epsilon(x)$ is continuous, piecewise linear and $\epsilon(X_{r+1}) - \epsilon(X_r) = \frac{1}{n+1}$. The asymmetric kernel to be used to estimate the density function at x is,

$$k_x(t, \epsilon(x)) = \begin{cases} k\left(\frac{t}{2\epsilon(x)}\right) & t \geq 0 \\ k\left(\frac{t}{2(1-\epsilon(x))}\right) & t < 0 \end{cases} \quad (4)$$

If k is a density function symmetric about zero then we note that for each x the function $k_X(t, \epsilon(x))$ is a density function and is symmetric if and only if $\epsilon(x) = 1/2$. In the following section to estimate the density at x we use the kernel (4) and the bandwidths h_{opt} (Silverman (1986) or of Rattihalli and Patil (2019)) and h_{NR} proposed by Zhang and Wang (2009) and based on the simulation we show that proposed estimators based on the asymmetric kernels perform better than those based on Normal and Uniform symmetric kernels.

3 Density estimators using asymmetric kernels:

Let $X_1, X_2, \ldots X_n$ be ordered observations. To estimate the density at x, $X_r \leq x < X_{r+1}$ we use asymmetric kernel k_r defined in (3) depends on x. We note that kernel k used in $\hat{f}_u$ and kernels k_i used in $\hat{f}$ do not depend on x, the point at which density is to be estimated. These estimators are proper density functions. However the nonnegative function,

$$g_1(x) = \frac{1}{nh}\sum_{i=1}^{n} k_r\left(\frac{x-X_i}{h}\right), X_r \leq x < X_{r+1} \quad (5)$$

is not necessarily a density function. Hence we define,

$$\hat{f}_1(x) = \frac{1}{Cnh}\sum_{i=1}^{n} k_r\left(\frac{x-X_i}{h}\right)[X_r \leq x < X_{r+1}] \quad (6)$$

where the normalizing constant C is,

$$C = \sum_{r=0}^{n} \int_{-\infty}^{\infty} g_1(x)[X_r \leq x < X_{r+1}]dx$$

Hence from (3) and (5) we have for $X_r \leq x < X_{r+1}$,

$$\hat{f}_1(x) = \frac{1}{cnh}\left\{\sum_{i=1}^{r} k_r\left(\frac{(x-X_i)(n+2)}{2h(n-r+1)}\right) + \sum_{i=r+1}^{n} k_r\left(\frac{(x-X_i)(n+2)}{2h(r+1)}\right)\right\} \quad (7)$$

Remark 3.1: $\hat{f}_1(x)$ is not continuous at $X_r,\ r = 1, 2, \ldots, n$ and the magnitude of the jump is $\hat{f}_1(X_r+) - \hat{f}_1(X_r-)$.

Remark 3.2: Simulation study reveals that $\hat{f}_1(x)$ is smaller than $\hat{f}(x)$ for x outside the support for any density function and any base symmetric kernel k (.).

The kernel functions to estimate the densities at u, v in a small neighborhood of $X_r, u < X_r < v$ are respectively k_r and k_{r+1} and hence the estimator (7) will not be continuous at X_r. Hence to obtain a smooth estimator, in the following, we use kernels, which in a sense are continuous. To avoid jumps we we use the kernel defined in (4) where ϵ is a continuous function of x. Consider the estimator is given by,

$$\hat{f}_2(x) = \frac{1}{Cnh}\left\{\sum_{i=1}^{r} k_x\left(\frac{(x-X_i)}{\epsilon\prime(x)}\right) + \sum_{i=1}^{n} k_x\left(\frac{(x-X_i)}{\epsilon(x)}\right)\right\} \quad (8)$$

Where $\epsilon(x) = \frac{r+\frac{x-X_r}{X_{r+1}-X\_r}}{n+1}$ and $\epsilon'(x) = 2 - \epsilon(x)$.

To describe and distinguish the nature of the four estimators we compute the estimators $\hat{f}_u$, $\hat{f}$, $\hat{f}_1$, $\hat{f}_2$ and sketch their graphs in Figure 2 for the data set $\{0.1, 0.5, 1.4, 4.6\}$. It is observed that $\hat{f}_u$ has higher values at the points away from the data cloud, as compared to other estimators. It is to be noted that only the estimator $\hat{f}_1$ has jumps at the observations 9.1, 0.5, 1.4, and 4.6, and the respective magnitudes of jumps are 0.0757, 0.0313, −0.0644, and −0.0001.

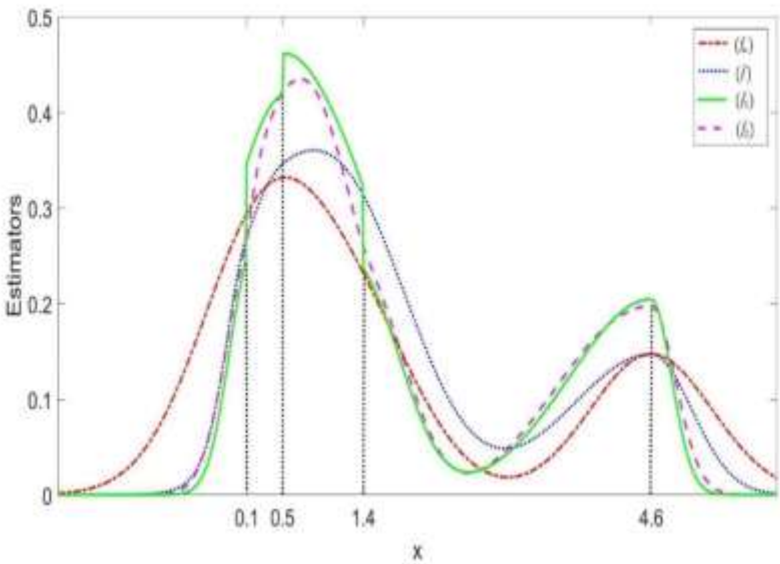


Figure 2: The graphs of the estimators $\hat{f}_u, \hat{f}, \hat{f}_1, \hat{f}_2$ for a typical data set {0.1, 0.5, 1.4, 4.6}

4 Performance Evaluation of the Proposed Estimator:

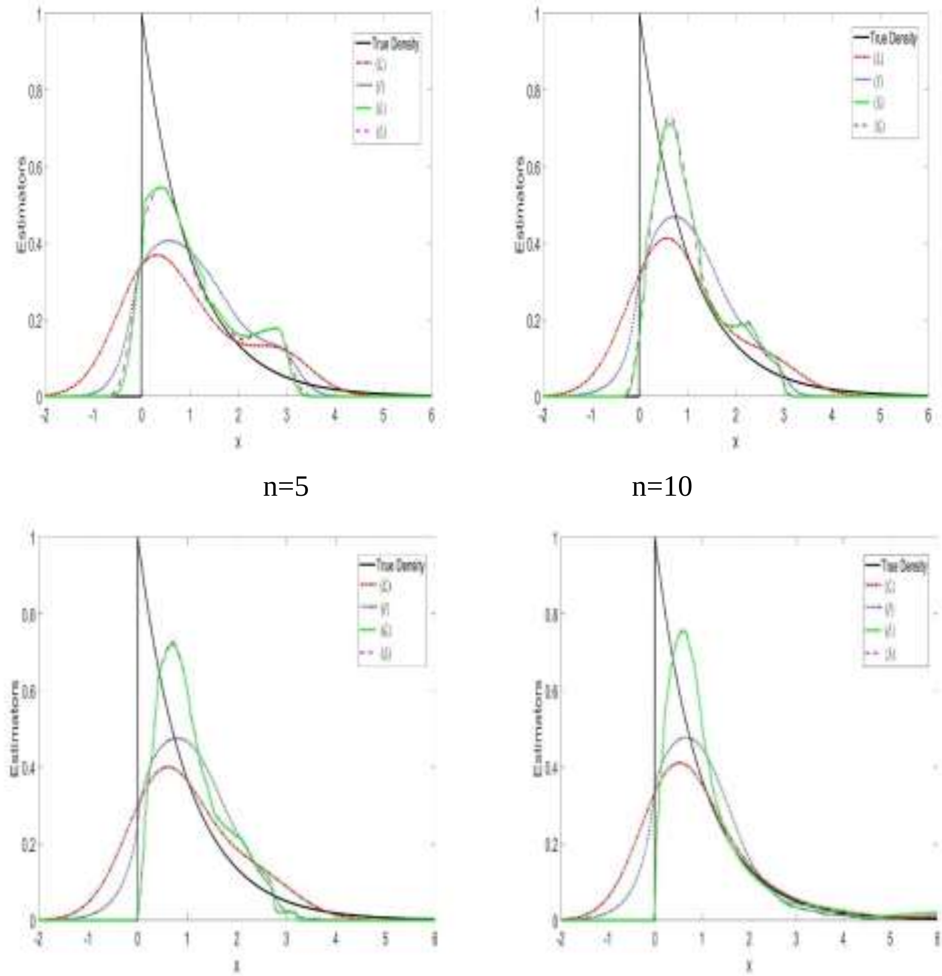
In this section, we evaluate the performance of $\hat{f}_1$, and $\hat{f}_2$ through simulation, comparing them with the usual symmetric kernel density estimator $\hat{f}_u$ of Silverman (1983) and the asymmetric kernel density estimator, $\hat{f}$ of Rattihalli and Patil (2019), using the standard normal density as the base kernel. We generate 1000 samples, each of sizes 5, 10, 30, 50, and 100 from Exponential, Triangular, Normal, Cauchy, $U(-1, 1)$, and Laplace distributions. We note that except at a finite number of points, densities corresponding to these distributions are differentiable any number of times. We consider L_1 and L_2 norms between the true density and the estimator as a comparison measure. The mean of L_1 norms (ML_1) and mean of squares of L_1 norms (MS L_1) are reported in Table-1. Similar results with L_2- norms are reported in Table-2. As expected, for the Exponential, Uniform, and Triangular models, whose supports do not cover the entire real line, we observe that the proposed estimators perform better. In general, $\hat{f}_1$, and $\hat{f}_2$ perform better when the support of the distribution is bounded on one or both sides.

Based on 1000 samples each of size 5, 10, 50, and 100 from Exponential distribution with mean 1, we sketch the graphs of the averages of the 1000 estimators $\hat{f}_u, \hat{f}, \hat{f}_1, \hat{f}_2$ and the true density function in Fig.3. As desired it is observed that biases of the proposed estimators $\hat{f}_1$, and $\hat{f}_2$ at the points outside the support of the true density are very less. In fact the same was true for almost every generated sample.

To quantify the overestimation, outside the support, the L_1-norm, the areas be- low 0 and under the averages of 1000 estimators for each one of the estimator $\hat{f}_u$, $\hat{f}$, $\hat{f}_1$, $\hat{f}_2$ are computed, for sample sizes n=5, 10 50, and 100 and reported in Table 3. The tabulated values indicate that proposed estimators have very little bias outside the support.

5 Conclusions and Scope:

Data-dependent asymmetric kernels are used to reduce boundary bias. Based on extensive simulation it is observed that proposed estimators reduce the bias and more so outside the support. The performance of estimators is evaluated using different norms and mean-square errors. We have used different scaling to generate asymmetric kernels, however, one can use other methods to generate asymmetric models using symmetric models.



n=5 n=10

n=50 n=100

Figure 3: Graphs of the averages of 1000 estimators based on the samples of sizes 5,10,50, 100 and the true density function.

Table 1: The Mean(ML_1) and the mean of squares($MS\ L_1$) of the 1000 L_1-norms of the estimators.

Model	Estimators	5		10		30		50		100	
		ML_1	$MS\ L_1$	ML_1	$MS\ L_1$	ML_1	$MS\ L_1$	ML_1	$MS\ L_1$	ML_1	$MS\ L_1$
Exp(1)	$\hat{f}_u$	0.6237	0.0191	0.5533	0.0092	0.4737	0.0029	0.4580	0.0020	0.4351	0.0011
	$\hat{f}$	0.6040	0.0275	0.5334	0.0137	0.4607	0.0037	0.4450	0.0025	0.4279	0.0013
	$\hat{f}_1$	0.5580	0.0259	0.4776	0.0104	0.3884	0.0018	0.3568	0.0012	0.2955	0.0006
	$\hat{f}_2$	0.5989	0.0139	0.5040	0.0060	0.4032	0.0015	0.3671	0.0010	0.3015	0.0006
U(-1,1)	$\hat{f}_u$	0.4797	0.0096	0.4471	0.0049	0.4245	0.0017	0.4178	0.0012	0.4165	0.0007
	$\hat{f}$	0.4189	0.0153	0.3578	0.0071	0.3145	0.0017	0.3079	0.0010	0.3007	0.0005
	$\hat{f}_1$	0.4476	0.0096	0.4181	0.0045	0.3963	0.0011	0.3811	0.0006	0.3366	0.0004
	$\hat{f}_2$	0.4807	0.0070	0.4478	0.0034	0.4102	0.0009	0.3898	0.0006	0.3407	0.0004
N(0,1)	$\hat{f}_u$	1.1227	0.0640	1.0756	0.0320	1.0552	0.0112	1.0433	0.0066	1.0409	0.0029
	$\hat{f}$	1.0226	0.0586	0.9267	0.0228	0.8399	0.0043	0.8138	0.0018	0.7970	0.0004

	$\hat{f}_1$	1.1542	0.0272	1.1230	0.0073	1.0701	0.0012	1.0315	0.0005	0.9447	0.0003
	$\hat{f}_2$	1.2073	0.0206	1.1621	0.0060	1.0874	0.0011	1.0424	0.0005	0.9500	0.0003
Cauchy	$\hat{f}_u$	0.5129	0.0302	0.4188	0.0176	0.2747	0.0060	0.2236	0.0037	0.1775	0.0021
	$\hat{f}$	0.5595	0.0341	0.4669	0.0197	0.2748	0.0061	0.2384	0.0040	0.2160	0.0034
	$\hat{f}_1$	0.5537	0.0314	0.4641	0.0165	0.3214	0.0036	0.2755	0.0021	0.2228	0.0011
	$\hat{f}_2$	0.5600	0.0278	0.4721	0.0144	0.3329	0.0033	0.2838	0.0019	0.2270	0.0011
Laplace	$\hat{f}_u$	0.5663	0.0365	0.4231	0.0177	0.2805	0.0063	0.2332	0.0039	0.1906	0.0020
	$\hat{f}$	0.6220	0.0347	0.4765	0.0197	0.3326	0.0068	0.2842	0.0041	0.2467	0.0026
	$\hat{f}_1$	0.6000	0.0312	0.4688	0.0142	0.3475	0.0028	0.3118	0.0011	0.2775	0.0005
	$\hat{f}_2$	0.6017	0.0261	0.4822	0.0120	0.3594	0.0025	0.3211	0.0010	0.2829	0.0004
Triangular	$\hat{f}_u$	0.7459	0.0024	0.6562	0.0019	0.6998	0.0005	0.5585	0.0006	0.3437	0.0010
	$\hat{f}$	0.7086	0.0029	0.5812	0.0014	0.6109	0.0001	0.4271	0.0003	0.1539	0.0007
	$\hat{f}_1$	0.5665	0.0039	0.5469	0.0016	0.6006	0.0007	0.5666	0.0005	0.4901	0.0003
	$\hat{f}_2$	0.5856	0.0027	0.5756	0.0014	0.6138	0.0007	0.5756	0.0005	0.4954	0.0003

Table 2: The Mean(ML2) and the mean of squares(MS L2) of the 1000 L2-norms of the estimators.

Model	Estimators	5		10		30		50		100	
		ML_2	MS L_2	ML_2	MS L_2	ML_2	MS L_2	ML_2	MS L_2	ML_2	MS L_2
Exp(1)	$\hat{f}_u$	0.1620	0.0043	0.1385	0.0013	0.1211	0.0002	0.1191	0.0001	0.1168	0.0001
	$\hat{f}$	0.1753	0.0110	0.1424	0.0047	0.1138	0.0010	0.1086	0.0006	0.1046	0.0003
	$\hat{f}_1$	0.1911	0.0098	0.1718	0.0040	0.1499	0.0011	0.1364	0.0008	0.1080	0.0004
	$\hat{f}_2$	0.1792	0.0050	0.1642	0.0026	0.1484	0.0010	0.1361	0.0007	0.1083	0.0004
U(-1,1)	$\hat{f}_u$	0.0982	0.0017	0.0802	0.0006	0.0678	0.0001	0.0649	0.0000	0.0631	0.0000
	$\hat{f}$	0.0971	0.0031	0.0715	0.0012	0.0542	0.0002	0.0514	0.0001	0.0488	0.0001
	$\hat{f}_1$	0.1354	0.0033	0.1300	0.0017	0.1262	0.0005	0.1225	0.0003	0.1081	0.0002
	$\hat{f}_2$	0.1492	0.0028	0.1422	0.0016	0.1319	0.0005	0.1261	0.0003	0.1098	0.0002
N(0,1)	$\hat{f}_u$	0.4961	0.0305	0.4629	0.0133	0.4526	0.0043	0.4458	0.0025	0.4443	0.0011
	$\hat{f}$	0.4260	0.0320	0.3556	0.0115	0.3050	0.0026	0.2893	0.0013	0.2803	0.0005
	$\hat{f}_1$	0.5507	0.0230	0.5245	0.0068	0.4973	0.0013	0.4752	0.0006	0.4320	0.0003
	$\hat{f}_2$	0.6046	0.0200	0.5619	0.0062	0.5121	0.0013	0.4842	0.0006	0.4363	0.0003
Cauchy	$\hat{f}_u$	0.0437	0.0010	0.0286	0.0004	0.0135	0.0001	0.0090	0.0000	0.0055	0.0000
	$\hat{f}$	0.0512	0.0013	0.0351	0.0006	0.0129	0.0001	0.0100	0.0000	0.0090	0.0000
	$\hat{f}_1$	0.0493	0.0012	0.0345	0.0005	0.0190	0.0001	0.0138	0.0000	0.0089	0.0000
	$\hat{f}_2$	0.0514	0.0011	0.0362	0.0004	0.0205	0.0001	0.0147	0.0000	0.0093	0.0000
Laplace	$\hat{f}_u$	0.0737	0.0028	0.0391	0.0007	0.0180	0.0001	0.0133	0.0001	0.0102	0.0000
	$\hat{f}$	0.0896	0.0036	0.0504	0.0012	0.0239	0.0002	0.0175	0.0001	0.0132	0.0000
	$\hat{f}_1$	0.0820	0.0028	0.0484	0.0008	0.0268	0.0001	0.0219	0.0000	0.0175	0.0000
	$\hat{f}_2$	0.0831	0.0024	0.0520	0.0008	0.0289	0.0001	0.0233	0.0000	0.0182	0.0000
Triangu lar	$\hat{f}_u$	0.1988	0.0005	0.1624	0.0004	0.1774	0.0001	0.1245	0.0001	0.0559	0.0001
	$\hat{f}$	0.1816	0.0008	0.1303	0.0003	0.1369	0.0000	0.0771	0.0000	0.0132	0.0000
	$\hat{f}_1$	0.2060	0.0021	0.1963	0.0010	0.2221	0.0003	0.1915	0.0002	0.1384	0.0001
	$\hat{f}_2$	0.2229	0.0019	0.2153	0.0010	0.2311	0.0003	0.1973	0.0002	0.1414	0.0001

Table 3: The areas below 0 and the average values of the estimators $\hat{f}_u$, $\hat{f}$, $\hat{f}_1$, $\hat{f}_2$

n	$\hat{f}_u$	$\hat{f}$	$\hat{f}_1$	$\hat{f}_2$
5	0.2454	0.1239	0.0508	0.0610
10	0.1899	0.0898	0.0104	0.0118
50	0.1724	0.0820	0.00013	0.00014
100	0.1720	0.0811	0.00012	0.00013

References

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

2. Jones, M. and Foster, P. (1996). A simple nonnegative boundary correction method for kernel density estimation. Statistica Sinica, pages 1005–1013.
3. Rattihalli, R. and Patil, S. (2019). Data dependent asymmetric kernels for estimating the density function. Sankhya A, pages 1–32.
4. Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics, pages 65–78.
5. Silverman, B. W. (1986). Density estimation for statistics and data analysis, volume 26. CRC press.
6. Wand, M. P. and Jones, M. C. (1994). Kernel smoothing. CRC press.
7. Yin, X.-F. and Hao, Z.-F. (2007). Adaptive kernel density estimation using beta kernel. In 2007 International Conference on Machine Learning and Cybernetics, volume 6, pages 3293–3297. IEEE.
8. Zhang, J. and Wang, X. (2009). Robust normal reference bandwidth for kernel density estimation. Statistica Neerlandica, 63(1):13–23.