Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Applications and Applied Mathematics:
An International Journal
(AAM)

Vol. 15, Issue 2 (December 2020), pp. 875 – 893

The Linear Combination of Kernels in the Estimation Of Cumulative Distribution Functions

[1*] **Abdel-Razzaq Mugdadi and** [2] **Rugayyah Sani**

[1,2]Department of Mathematics and Statistics
Jordan University of Science and Technology
Irbid-JORDAN
[1]aamugdadi@just.edu.jo; [2] rukayya.ibrahim@siu.edu

*Corresponding Author

Received: July 21, 2020; Accepted: October 21, 2020

Abstract

The kernel distribution function estimator method is the most popular nonparametric method to estimate the cumulative distribution function $F(x)$. In this investigation, we propose a new estimator for $F(x)$ based on a linear combination of kernels. The mean integrated squared error, asymptotic mean integrated squared error and the asymptotically optimal bandwidth for the new estimator are derived. Also, based on the plug-in technique in density estimation, we propose a data based method to select the bandwidth for the new estimator. In addition, we evaluate the new estimator using simulations and real life data.

Keywords: Distribution function; kernel smoothing; Plug-in; Bandwidth Selection; Density estimation; Mean square error

MSC 2010 No.: 62G05, 62G07

1. Introduction

In statistics, we are required to derive statistical inferences, descriptions and summaries from a given set of data. The basic properties of this data are conventionally described by the probability density function pdf and its distribution function (cumulative distribution function cdf). In most cases, the pdf or the cdf for a given data are provided and this makes conclusions about the data easier.

However, in practical situations, random variables from these data are required to be studied without the knowledge of the pdf and the cdf. Alternatively, we are given a set of n observations $x_1, x_2, \ldots, x_n$ assumed to be deductions from independent and identically distributed random variables $X_1, X_2, \ldots, X_n$ without the knowledge of the distribution they came from. Here, we are faced with the problem of estimating the unknown pdf and the cdf, the estimation being the first step to be taken in order to understand and represent the behavior of a data.

The goal of estimation of the pdf and cdf can be reached by following either of these two approaches: the parametric approach and the nonparametric approach. The parametric approach involves the assumption that the random variable belongs to a particular family of distribution, for example, Normal family, Gamma family, etc. which has a range of parameters. Thus, several conditions to use the parametric estimate must be investigated and if these conditions are satisfied, the estimate will have more power over the nonparametric estimate. The estimation using this parameterized functions is usually associated with the data interpreting being too rigid due to the restriction by parameters. Thousands of investigators used the parametric approach to estimate the parameters: for example, Shakhatreh et al. (2016) used the parametric approach in the estimation of the beta generalized linear exponential distribution. Ahsanullah and Habibullah (2015) used the parameter approach in some estimations for the exponential distribution, and Yuan (2018) used a parametric method to estimate the the bivariate Weibull distribution, where the author used the generalized moment method for reliability evaluation.

The nonparametric approach is more often than not more accurate than the parametric approach. This is because there exist no parameter restricting the behavior of the random variables and thus the data is left to “speak for itself.” Hence, the estimates of the pdf and cdf follow the behavior of the data. The most well know approach for density estimation is using the kernel method which was proposed in 1956 by Rosenblatt and independently by Parzen in 1962. This estimator is also known as a Rosenblatt-Parzen estimator. Nadaraya (1962) proposed the kernel method to estimate the cumulative distribution function $F(x)$. There are several measures of error to evaluate the estimator for the cumulative distribution function. Azzalini (1981), Swanepoel (1988), Jones (1990), Sarda (1993) and Jones (1990), and Altman and Leger (1995) discussed the exact and asymptotic mean square error. Ahmad and Mugdadi (2006) derived the weighted mean Hellinger distance for the estimator. The kernel type estimator is a well know technique in the nonparametric estimation. Rabhi and Soltani (2016) investigate the kernel estimate of the conditional hazard function. In addition, Wand and Jones (1994) discussed the kernel density estimation regression estimation and Multivariate estimation. Additional details about the bandwidth selection can be found in Wand and Jones (1995) and Mugdadi and Jeter (2010).

In this investigation we study the linear combinations of kernels in the the kernel distribution estimator. The cruel parts of the kernel distribution estimator is to derive one of the measure of errors and to obtain the optimal bandwidth. Thus, our goal in this research paper is to derive the mean integrated square error as the main measure of error in kernel estimators then to obtain the optimal bandwidth for the estimator. In addition, we will modify the the Plug in technique to obtain a data based method based on linear combinations of kernels. More details about the plug-in technique can be found in Wand and Jones (1995). Investigators can use linear combinations of kernels in any kernel estimation such as regression and multivariate estimations. Also, it can be use in other measures of errors such as Hellinger Distance.

2. The Linear Combination of Kernels

Let $X_1, X_2, \dots, X_n$ be a random sample. The kernel density estimator for the probability density function $f(x)$ is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{j=1}^{n} k\left(\frac{x - X_j}{h}\right), \tag{1}$$

where $k(.)$ is the kernel function and h is the bandwidth. In addition, the kernel estimator for the cumulative distribution function $F(x)$ is given by

$$\hat{F}(x) = \frac{1}{n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right), \tag{2}$$

where $K(x) = \int_{-\infty}^{x} k(t)\, dt$. We replace the kernel function, $k(.)$ in the kernel probability density estimator by the linear combination of kernels, $\sum_{i=1}^{p} a_i k_i(.)$, where $p \geq 2$, $\sum_{i=1}^{p} a_i = 1$ and $a_i s$ are positive, to obtain

$$\hat{f}^*(x) = \frac{1}{nh} \sum_{j=1}^{n} \sum_{i=1}^{p} a_i k_i\left(\frac{x - X_j}{h}\right), \tag{3}$$

where for each i, $\sum_{i=1}^{k} a_i = 1$, k_i is a kernel function, symmetric about zero and $\int k(u)du = 1$. Linear combination of kernels have been used in density estimation by Ahmad and Ran (2004). A linear combination of two normal densities have been considered by Savchuk et al. (2010) suggesting the possibility of using a combination of other classes of kernels. Thus, a motivation to discuss the linear combination of kernels in the cumulative distribution function estimation is important.

The kernel distribution function estimator is consequently given by

$$\hat{F}^*(x) = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{p} a_i K_i\left(\frac{x - X_j}{h}\right), \tag{4}$$

where for each i, $K_i(x) = \int_{-\infty}^{x} k_i(z)\, dz$ and for the same $a_i s$.

3. The MISE and the Optimal Bandwidth

The most common criterion to evaluate the estimator $\hat{F}(x)$ is the Mean Integrated Squared Error (MSE), which is given by

$$MISE(\hat{F}(x)) = E\int (\hat{F}(x) - F(x))^2 W(x)\,dx,$$

where $W(x)$ is a nonnegative weight function. To derive the optimal bandwidth, we will evaluate the asymptotic mean integrated square error for the estimator. We use assumptions synonymous to those of Sarda (1993). The assumptions on the weight function, the bandwidths, the kernel and the distribution function are:

(I) The weight function, W is bounded and has a compact support.

(II) For each i, K_i is absolutely continuous, $\lim_{x\to-\infty} K_i(x) = 0$ and $\lim_{x\to\infty} K_i(x) = 1$.

(III) Given that $k_i = K_i'$, we have $\int xk_i(x)\,dx = 0$ and $\int x^2 k_i\,dx < \infty$.

(IV) For each i,

$$K_i(z) = \begin{cases} 0, & z > 1, \\ \frac{1}{2} + g_i(z), & |z| < 1, \\ 1, & z < -1, \end{cases}$$

where $g_i(z) = -g_i(-z)$ for all z such that $|z| \le 1$.

(V) F is twice differentiable and F and $|f'|$ are bounded on the support of W.

Under conditions $(I) - (V)$, we will prove Lemma 3.1 and Lemma 3.2. In addition, we will use these two lemmas to prove Theorem 3.1.

Lemma 3.1.

$$Var(\hat{F}^*) = \frac{1}{n}\left(F(x)\left(A + 2B - F(x)\right)\right) + \frac{h}{n} f(x)\left(R_i(K)\right)$$
$$+ \frac{2}{n}\left(E_{il}(K) - (A + 2B)\right) + O(h^2).$$

Proof:

Note that

$$Var(\hat{F}^*(x)) = E(\hat{F}^*(x))^2 + (E(\hat{F}^*(x)))^2. \tag{5}$$

To evaluate E($\hat{F}^*(x)$), we have

$$E\left(\sum_{i=1}^{p} a_i K_i\left(\frac{x-X_i}{h}\right)\right) = \int_{-\infty}^{\infty} \sum_{i=1}^{p} a_i K_i\left(\frac{x-y}{h}\right) f(y)\, dy.$$

Using IV,

$$\begin{aligned}\int_{-\infty}^{\infty} \sum_{i=1}^{p} a_i K_i(z) f(x-hz)\, dz &= \sum_{i=1}^{p} a_i \int_{-\infty}^{\infty} K_i(z) f(x-hz)\, dz \\ &= \int_{-\infty}^{-1} 0 + h\sum_{i=1}^{p} a_i \int_{-1}^{1}\left(\frac{1}{2}+g_i(z)\right) f(x-hz)\, dz \\ &+ h\int_{1}^{\infty} f(x-hz)\, dz.\end{aligned}$$

But,

$$\begin{aligned}h\sum_{i=1}^{p} a_i \int_{-1}^{1}\left(\frac{1}{2}+g_i(z)\right) f(x-hz)\, dz &= h\sum_{i=1}^{p} a_i \int_{-1}^{1}\left(\frac{1}{2}+g_i(z)\right) f(x)\, dz \\ &+O(h^2) \\ &= hf(x) + O(h^2),\end{aligned}$$

and

$$h\int_{1}^{\infty} f(x-hz)\, dz = \int_{-\infty}^{x-h} f(y)\, dy = F(x) - hf(x) + O(h^2).$$

Therefore,

$$E(\hat{F}^*(x)) = F(x) - hf(x) + O(h^2) + hf(x) + O(h^2) = F(x) + O(h^2),$$

and

$$(E(\hat{F}^*(x)))^2 = F(x)^2 + O(h^4).$$

But,

$$\begin{aligned}
E(\hat{F}^*(x))^2 =& E\left(\sum_{i=1}^{p} a_i K_i^2\left(\frac{x-X_j}{h}\right)\right)^2\\
=& E[\sum_{i=1}^{p} a_i^2 K_i\left(\frac{x-X_j}{h}\right)^2\\
&+2\sum\sum_{i\neq l} a_i a_l K_i\left(\frac{x-X_j}{h}\right)K_l\left(\frac{x-X_j}{h}\right)\\
=& \sum_{i=1}^{p} a_i^2 E\left(K_i^2\left(\frac{x-X_j}{h}\right)\right)\\
&+2\sum\sum_{i\neq l} a_i a_l E\left(K_i\left(\frac{x-X_j}{h}\right)K_l\left(\frac{x-X_j}{h}\right)\right)\\
=& \sum_{i=1}^{p} a_i^2 \int_{-\infty}^{\infty} K_i^2\left(\frac{x-y}{h}\right) f(y)\,dy\\
&+\ 2\sum\sum_{i\neq l} a_i a_l \int_{-\infty}^{\infty} K_i\left(\frac{x-y}{h}\right)K_l\left(\frac{x-y}{h}\right) f(y)\,dy.
\end{aligned}$$

Note that

$$\sum_{i=1}^{p} a_i^2 E\left(K_i^2\left(\frac{x-X_j}{h}\right)\right) = \sum_{i=1}^{p} a_i^2 \int_{-\infty}^{\infty} K_i^2\left(\frac{x-y}{h}\right) f(y)\,dy,$$

and

$$K_i^2(z) = \begin{cases} 0, & y > x+h,\\ \frac{1}{4}+g_i(z)+g_i^2(z), & |z|<1,\\ 1, & y < x-h. \end{cases}$$

Thus,

$$\begin{aligned}
\int_{-\infty}^{\infty}\sum_{i=1}^{p} a_i^2 K_i^2(z) f(x-hz)\,dz =& \sum_{i=1}^{p} a_i^2 \int_{-\infty}^{\infty} K_i^2(z) f(x-hz)\,dz\\
=& \int_{-\infty}^{-1} 0 dz\\
&+h\sum_{i=1}^{p} a_i^2 \int_{-1}^{1}\left(\frac{1}{4}+g_i(z)+g_i^2(z)\right) f(x-hz)\,dz\\
&+h\sum_{i=1}^{p} a_i^2 \int_{1}^{\infty} f(x-hz)\,dz\,,
\end{aligned}$$

and

$$\begin{aligned}
\int_{-1}^{1} K_i^2(z)\,dz =& \int_{-1}^{1}\left(g_i^2(z)+g_i(z)+\frac{1}{2}\right)\,dz\\
=& \int_{-1}^{1} g_i^2(z)\,dz+\frac{1}{2}.
\end{aligned}$$

Therefore,

$$\sum_{i=1}^{p} a_i^2 E(K_i^2(z)) = \sum_{i=1}^{p} a_i^2 F(x) + hf(x) \sum_{i=1}^{p} a_i^2 \left(\int_{-1} 1(K_i^2(z)) - 1 \right) + O(h^2).$$

Let $\sum_{i=1}^{p} a_i^2 = A$. This implies

$$\sum_{i=1}^{p} a_i^2 E(K_i^2(z)) = AF(x) - Ahf(x) + hf(x) \sum_{i=1}^{p} a_i^2 E(K_i^2(z))\, dz + O(h^2).$$

But,

$$\sum\sum\nolimits_{i \neq l} a_i a_l E\left(K_i \left(\frac{x - X_j}{h} \right) K_l \left(\frac{x - X_j}{h} \right) \right)$$
$$= \sum\sum\nolimits_{i \neq l} a_i a_l \int_{-\infty}^{\infty} K_i \left(\frac{x - y}{h} \right) K_l \left(\frac{x - y}{h} \right) f(y) dy.$$

Now,

$$E\left(K_i(z) K_l(z)\right) = \int_{-\infty}^{\infty} K_i(z) K_l(z) f(x - hz)\, dz,$$

where

$$K_i(z) K_l(z) = \begin{cases} 0, & y > x + h, \\ \frac{1}{4} + \frac{1}{2} g_i(z) + \frac{1}{2} g_i(z) + g_i(z) g_l(z), & |z| < 1, \\ 1, & y < x - h. \end{cases}$$

Therefore,

$$\begin{aligned}\int_{-\infty}^{\infty} K_i(z)K_l(z)f(x-hz)\,dz =& \int_{-\infty}^{1} 0\,dz \\ & + \int_{-1}^{1}\left(\frac{1}{4}+\frac{1}{2}g_i(z)+\frac{1}{2}g_l(z)+g_i(z)g_l(z)\right)\times \\ & f(x-hz)\,dz + h\int_{1}^{\infty} f(x-hz)\,dz\times \\ & \int_{-\infty}^{x-h} f(y)\,dy \\ & + h\int_{-1}^{1}\left(\frac{1}{4}+\frac{1}{2}g_i(z)+\frac{1}{2}g_l(z)+g_i(z)g_l(z)\right)\times \\ & f(x-hz)\,dz \\ =& F(x)-hf(x) \\ & + \int_{-1}^{1}\left(\frac{1}{4}+\frac{1}{2}g_i(z)+\frac{1}{2}g_l(z)+g_i(z)g_l(z)\right)\times \\ & f(x)\,dz + O(h^2) \\ =& F(x)-hf(x)+\frac{h}{2}f(x) \\ & + hf(x)\int_{-1}^{1} g_i(z)g_l(z)\,dz + O(h^2) \\ =& F(x)-\frac{h}{2}f(x) \\ & + hf(x)\int_{-1}^{1} g_i(z)g_l(z)\,dz + O(h^2).\end{aligned}$$

Here we note that

$$\int_{-1}^{1} g_i(z)g_l(z)\,dz = \int_{-1}^{1} K_i(z)K_l(z)\,dz - \frac{1}{2}.$$

Therefore,

$$\begin{aligned}\int_{-\infty}^{\infty} K_i(z)K_l(z)f(x-hz)\,dz =& F(x)-\frac{h}{2}f(x)+\left(\int_{-\infty}^{\infty} K_i(z)K_l(z)\,dz-\frac{1}{2}\right) \\ & + O(h^2) \\ =& F(x)-hf(x)+hf(x)\int_{-\infty}^{\infty} K_i(z)K_l(z)\,dz \\ & + O(h^2),\end{aligned}$$

and

$$\begin{aligned}2\sum\sum\nolimits_{i\neq l}a_ia_lE\left(K_i(z)K_l(z)\right) =&2\sum\sum\nolimits_{i\neq l}a_ia_l[F(x)-hf(x)\\&+hf(x)\int_{-\infty}^{\infty}K_i(z)K_l(z)\,dz+O(h^2)]\\=&\left(2\sum\sum\nolimits_{i\neq l}a_ia_lF(x)\right)-\left(2\sum\sum\nolimits_{i\neq l}a_ia_lhf(x)\right)\\&+hf(x)2\sum\sum\nolimits_{i\neq l}a_ia_l\int_{-\infty}^{\infty}K_i(z)K_l(z)\,dz+O(h^2).\end{aligned}$$

Let $\sum\sum_{i\neq l}a_ia_l = B$. Thus,

$$\begin{aligned}2\sum\sum\nolimits_{i\neq l}a_ia_lE\left(K_i(z)K_l(z)\right) =&2BF(x)-hBf(x)\\&+hf(x)2\sum\sum\nolimits_{i\neq l}a_ia_l\int_{-\infty}^{\infty}K_i(z)K_l(z)\,dz+O(h^2).\end{aligned}$$

Therefore, the variance is

$$\begin{aligned}Var\left(\hat{F}^*(x)\right)=&\frac{1}{n}\left(AF(x)-Ahf(x)+hf(x)\sum_{i=1}^{p}a_i^2E(K_i^2(z))\,dz+O(h^2)\right)\\&+[2BF(x)-hBf(x)+hf(x)2\sum\sum\nolimits_{i\neq l}a_ia_l\int_{-\infty}^{\infty}K_i(z)K_l(z)\,dz\\&+O(h^2)]-\left(F(x)^2+O(h^4)\right)\\=&\frac{1}{n}\left(F(x)\left(A+2B-F(x)\right)\right)+\frac{h}{n}f(x)\left(\sum_{i=1}^{p}a_i^2E(K_i^2(z))\,dz\right)\\&+\frac{2}{n}\left(\sum\sum\nolimits_{i\neq l}a_ia_l\int_{-\infty}^{\infty}K_i(z)K_l(z)\,dz-(A+2B)\right)+O(h^2)\\=&\frac{1}{n}\left(F(x)\left(A+2B-F(x)\right)\right)+\frac{h}{n}f(x)\left(R_i(K)\right)\\&+\frac{2}{n}\left(E_{il}(K)-(A+2B)\right)+O(h^2).\end{aligned}$$

■

Lemma 3.2.

$$Bias^2(\hat{F}^*)=\frac{1}{4}h^4(f'(x))^2(\mu_i(k))^2+O(h^6).$$

Proof:

It is known that

$$Bias(\hat{F}^*(x))=E(\hat{F}^*(x))-F(x).$$

But,

$$E(\hat{F}^*(x)) = \int_{-\infty}^{x-h} f(y)dy + h\sum_{i=1}^{p} a_i \int_{-1}^{1} \left(\frac{1}{2} + g_i(z)\right) f(x - hx\, dz).$$

Here also,

$$\int_{-\infty}^{x-h} f(y)dy = F(x) - hf(x),$$

and

$$\begin{aligned} h\sum_{i=1}^{p} a_i \int_{-1}^{1} \left(\frac{1}{2} + g_i(z)\right) f(x - hz)\, dz =& h\sum_{i=1}^{p} a_i \int_{-1}^{1} \left(\frac{1}{2} + g_i(z)\right) \times \\ & \big(f(x) - hzf'(x) + O(h^2)\, dz\big) \\ =& h\sum_{i=1}^{p} a_i \int_{-1}^{1} K_i(z)[f(x) \\ & - hf'(x) + O(h^3)]dz \\ =& hf(x)\sum_{i=1}^{p} a_i \int_{-1}^{1} K_i(z)\, dz \\ & - h^2 f'(x)\sum_{i=1}^{p} a_i \int_{-1}^{1} a_i K_i(z)\, dz + O(h^3). \end{aligned}$$

Using integration by parts,

$$\int_{-1}^{1} K_i(z)\, dz = zK_i(z)|_{-1}^{1} - \int_{-1}^{1} zK_i(z)\, dz.$$

From condition (IV), $zK_i(z)|_{-1}^{1} = 1$, and $\int_{-1}^{1} zK_i(z)\, dz = \frac{1}{2} - \frac{1}{2}\int_{-1}^{1} z^2 k_i(z)\, dz$. Also,

$$\begin{aligned} \sum_{i=1}^{p} a_i \int_{-1}^{1} \left(\frac{1}{2} + g_i(z)\right) f(x - hz)\, dz =& f(x)\sum_{i=1}^{p} a_i - hf'(x)\sum_{i=1}^{p} a_i \times \\ & \left(\frac{1}{2} - \frac{1}{2}\int_{-1}^{1} z^2 k_i(z)\, dz\right) \\ =& f(x) - \frac{h^2}{2} f'(x) \\ & + \frac{h}{2} f'(x)\sum_{i=1}^{p} a_i \int_{-1}^{1} z^2 k_i(z)\, dz. \end{aligned}$$

Therefore,

$$\begin{aligned} Bias(\hat{F}^*(x)) =& F(x) - hf(x) + hf(x) - \frac{h^2}{2} f'(x) \\ & + \frac{h^2}{2} f'(x)\sum_{i=1}^{p} a_i \int_{-1}^{1} z^2 k_i(z)\, dz - F(x) + O(h^3) \\ =& \frac{1}{2} h^2 f'(x)\sum_{i=1}^{p} a_i \int_{-1}^{1} z^2 k_i(z)\, dz + O(h^3), \end{aligned}$$

and

$$Bias^2(\hat{F}^*(x)) = \frac{1}{4}h^4(f'(x))^2\left(\sum_{i=1}^{p} a_i \int_{-1}^{1} z^2 k_i(z)\,dz\right)^2 + O(h^6)$$
$$= \frac{h^4}{4}(f'(x))^2\,(\mu_i(k))^2 + O(h^6). \qquad \blacksquare$$

Theorem 3.1.

Under conditions $(I)-(V)$, the MSE for $\hat{F}^*(x)$ is given by:

$$MSE(\hat{F}^*(X)) = \frac{1}{n}F(x)\left((A+2B) - F(x)\right) + \frac{h}{n}f(x)[R_i(K) + 2E_{il}(K)$$
$$-(A+2B)] + \frac{1}{4}h^4 f'(x)^2 \mu_i(k)^2 + O(h^6),$$

where $A = \sum_{i=1}^{p} a_i^2$, $B = \sum\sum_{i\neq l} a_i a_l$, $R_i(K) = \sum_{i=1}^{p} a_i^2 \int_{-1}^{1} K_i(z)^2\,dz$, $E_{il}(K) = \sum\sum_{i\neq l} a_i a_l \int_{-1}^{1} K_i(z)K_l(z)\,dz$, and $\mu_i(k) = \sum_{i=1}^{p} a_i \int_{-1}^{1} z^2 k_i(z)\,dz$

Proof:

The expression for the mean squared error is given by

$$MSE(\hat{F}^*(x)) = Var(\hat{F}^*(x)) + Bias^2(\hat{F}^*(x)).$$

Therefore, using Lemma 3.1 and Lemma 3.2, the MSE is given by:

$$MSE(\hat{F}^*(x)) = \frac{1}{n}F(x)\left((A+2B) - F(x)\right) + \frac{h}{n}f(x)[R_i(K) + 2E_{il}(K)$$
$$-(A+2B)] + \frac{1}{4}h^4 f'(x)^2 \mu_i(k)^2 + O(h^6). \qquad \blacksquare$$

Lemma 3.3.

Under conditions (I) to (V) the $MISE$ for the kernel distribution function estimator using linear combination is given by

$$MISE(\hat{F}^*(x)) = \frac{1}{n}D_1(F) + \frac{h}{n}D_2(F)\left(R_i(K) + 2E_{il}(K) - (A+2B)\right)$$
$$+ \frac{h^4}{4}D_3(F)\,(\mu_i(k))^2 + O(h^6),$$

where $D_1(F) = \int F(x)(A+2B-F(x))f(x)W(x)\,dx$, $D_2(F) = \int f(x)^2 W(x)\,dx$, and $D_3(F) = \int f'(x)^2 f(x)W(x)\,dx$. In addition, A, B, $R_i(K)$, $E_{il}(K)$ and $\mu_i(k)$ are given and defined in Theorem 3.1.

Proof:

$$\begin{aligned}
MISE(\hat{F}^*(x)) &= \int MSE(\hat{F}^*(x))f(x)W(x)\,dx \\
&= \int (\frac{1}{n}F(x)\left((A+2B)-F(x)\right)+\frac{h}{n}f(x)[R_i(K)+2E_{il}(K) \\
&\quad -(A+2B)]+\frac{1}{4}h^4(f'(x))^2\mu_i(k)^2+O(h^6))f(x)W(x)\,dx \\
&= \frac{1}{n}\int F(x)(A+2B-F(x))f(x)W(x)\,dx \\
&\quad +\frac{h}{n}\int f(x)^2W(x)\,dx\,(R_i(K)\ +\ 2E_{il}(K)-(A+2B)) \\
&\quad +\frac{1}{4}h^4\int (f'(x))^2f(x)W(x)\,dx\,\left(\mu_i(k)^2\right)\ +\ O(h^6).
\end{aligned}$$

Hence,

$$\begin{aligned}
MISE(\hat{F}^*(x)) &= \frac{1}{n}D_1(F)+\frac{h}{n}D_2(F)[R_i(K)+2E_{il}(K)- \\
&\quad (A+2B)]+\frac{h^4}{4}D_3(F)\left(\mu_i\right)^2+O(h^6).
\end{aligned}$$

■

Corollary 3.1.

Under the conditions (I) to (V) above, the asymptotic $MISE$ for the kernel distribution function estimator using linear combination is given by:

$$\begin{aligned}
AMISE(\hat{F}^*(x)) &= \frac{1}{n}D_1(F)+\frac{h}{n}D_2(F)[R_i(K)+2E_{il}(K) \\
&\quad -(A+2B)]+\frac{h^4}{4}D_3(F)\left(\mu_i\right)^2,
\end{aligned}$$

and the asymptotically optimal bandwidth is given by:

$$h^*_{AMISE}=\frac{(D_2(F))^{1/3}\left(A+2B-R_i(K)-2E_{il}(K)\right)^{1/3}}{n^{1/3}(D_3(F))^{1/3}\left(\mu_i(k)\right)^{2/3}},$$

where $R_i(K)$, $E_{il}(K)$, $\mu_i(k)$, $D_2(F)$ and $D_3(F)$ are defined in Theorem 3.1 and Lemma 3.3.

Proof:

By deriving the $AMISE(\hat{F}^*(x))$ and having it equal to zero,

$$\frac{1}{n}D_2(F)\left(R_i(K)+2E_{il}(K)-(A+2B)\right)+h^3D_3(F)\left(\mu_i\right)^2=0.$$

Therefore,

$$h^3D_3(F)\left(\mu_i\right)^2=\frac{1}{n}D_2(F)\left(A+2B-R_i(K)-2E_{il}(K)\right),$$

which implies

$$h^3 = \frac{(D_2(F))\,(A + 2B - R_i(K) - 2E_{il}(K))}{n(D_3(F))\,(\mu_i)^2}.$$

Thus,

$$h = \frac{(D_2(F))^{1/3}\,(A + 2B - R_i(K) - 2E_{il}(K))^{1/3}}{n^{1/3}(D_3(F))^{1/3}\,(\mu_i)^{2/3}},$$

which can be denoted by h^*_{AMISE} and is called the asymptotic $MISE$ optimal bandwidth. ■

4. The Plug-in Method for Linear Combination

In all cases of kernel density estimation, the asymptotically optimal bandwidth is derived using either a single kernel or the linear combination of kernels in the estimator, the optimal bandwidth has the terms $D_2(F) = \int f(x)^2 W(x)\,dx$ and $D_3(F) = \int f'(x)^2 f(x) W(x)\,dx$, where both $D_2(F)$ and $D_3(F)$ contain the actual value of the pdf, $f(x)$ and so to estimate the asymptotically optimal bandwidth, we need to estimate the value of f. Thus, it is essential to estimate $D_2(F)$ and $D_3(F)$. One the technique to estimate them is using the plug-in method.

The plug-in method goes back to Woodroofe (1970). Then, it was used in the kernel density estimation by Hall and Marron (1987), Hall and Marron (1991), and Sheather and Jones (1991), and it was used for the cumulative distribution function by Polansky and Baker (2000). However, Altman and Léger (1993) estimated the optimal bandwidth with an estimate for $D_2(F)$ as suggested by Hall and Marron (1987). This estimator is given by

$$\hat{D}_2(F) = \frac{1}{n(n-1)} \sum_{i \neq j} \alpha_v^{-1} k_v \left(\frac{X_i - X_j}{\alpha_v} \right) W(X_i),$$

where α_v is a suitable bandwidth selected by using one of the bandwidth selector methods and also k_v is any suitable kernel, while, the estimator for $D_3(F)$ was given by

$$\hat{D}_3(F) = \frac{1}{n^3 \alpha_b^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} k_b' \left(\frac{X_i - X_j}{\alpha_b} \right) k_b' \left(\frac{X_i - X_k}{\alpha_b} \right),$$

where k_b and α_b are a suitable kernel and its associated bandwidth respectively. The estimators for the linear combination case are given in the following theorem.

Theorem 4.1.

The Plug-in estimators for $D_2(F)$ and $D_3(F)$ using a linear combination of kernels are given by

$$\hat{D}_2^*(F) = \frac{1}{n(n-1)b_1} \sum_{j \neq k} \sum_{i=1}^{p} a_i g_i \left(\frac{X_j - X_k}{b_1} \right) W(X_j),$$

and

$$\hat{D}_3^*(F) = \frac{1}{n^3 b_2^4} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{i=1}^{p} a_i q_i' \left(\frac{X_j - X_k}{b_2} \right) \sum_{i=1}^{p} a_i q_i' \left(\frac{X_j - X_l}{b_2} \right) W(X_j),$$

respectively.

Proof:

Thus, we will estimate $D_2(F)$ using a linear combination of kernels as follow:

$$\begin{aligned}
\hat{D}_2^*(F) &= \int \hat{f}^*(x_j) f(x) W(x)\, dx \\
&= \frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{nh} \sum_{k=1}^{n} \sum_{i=1}^{p} a_i k_i \left(\frac{X_j - X_k}{h} \right) \right) W(X_j) \\
&= \frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{h} \sum_{i=1}^{p} a_i k_i(0) + \frac{1}{(n-1)h} \sum_{j \neq k} \sum_{i=1}^{p} a_i k_i \left(\frac{X_j - X_k}{h} \right) \right) W(X_j) \\
&= \frac{1}{nh} \sum_{i=1}^{p} a_i k_i(0) + \frac{1}{n(n-1)h} \sum_{j \neq k} \sum_{i=1}^{p} a_i k_i \left(\frac{X_j - X_k}{h} \right) W(X_j).
\end{aligned}$$

The terms for which $i = j$, that is the terms with $k_i(0)$ are independent of the data and so can be thought of as bias terms (Hall and Marron, 1987). Therefore,

$$\hat{D}_2^*(F) = \frac{1}{n(n-1)h} \sum_{j \neq k} \sum_{i=1}^{p} a_i k_i \left(\frac{X_j - X_k}{h} \right) W(X_j).$$

For $D_3(F)$, we have

$$\begin{aligned}
D_3(F) &= \int f'(x)^2 f(x) W(x)\, dx \\
&= \int \left(f'(x) \right) \left(f'(x) \right) f(x) W(x)\, dx.
\end{aligned}$$

But

$$\hat{f^*}'(x) = \frac{1}{nh^2} \sum_{j=1}^{n} \sum_{i=1}^{p} a_i k_i' \left(\frac{x - X_j}{h} \right).$$

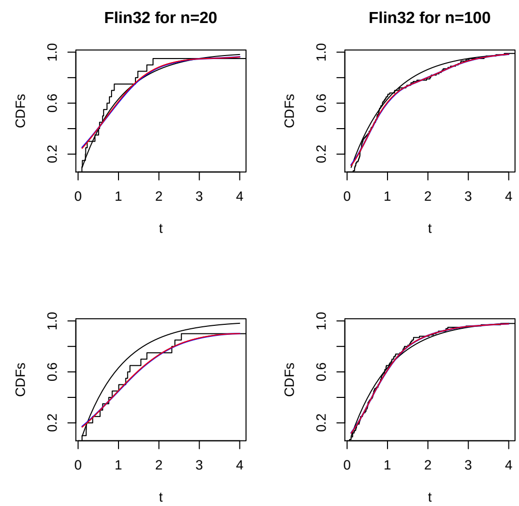

Figure 1. Exp(1)(Black), F_n(step function), $\hat{F}^*$ using Flin(red) and $\hat{F}$(blue)

Therefore,

$$\begin{aligned}
\hat{D}_3^*(F) &= \int \left(\hat{f^*}'(x)\right)\left(\hat{f^*}'(x)\right) f(x)W(x)\,dx \\
&= \int \left(\frac{1}{nh^2}\sum_{j=1}^{n}\sum_{i=1}^{p} a_i k_i'\left(\frac{x-X_j}{h}\right)\right)\left(\frac{1}{nh^2}\sum_{j=1}^{n}\sum_{i=1}^{p} a_i k_i'\left(\frac{x-X_j}{h}\right)\right) f(x)W(x)dx \\
&= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{nh^2}\sum_{k=1}^{n}\sum_{i=1}^{p} a_i k_i'\left(\frac{X_j-X_k}{h}\right)\right)\left(\frac{1}{nh^2}\sum_{l=1}^{n}\sum_{i=1}^{p} a_i k_i'\left(\frac{X_j-X_l}{h}\right)\right) W(X_j) \\
&= \frac{1}{n^3h^4}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{i=1}^{p} a_i k_i'\left(\frac{X_j-X_k}{h}\right)\sum_{i=1}^{p} a_i k_i'\left(\frac{X_j-X_l}{h}\right) W(X_j).
\end{aligned}$$

Note that the kernels and the bandwidths used here are chosen as desired depending on how well they work in the estimation. That is, they are not necessarily the same as those used in other parts of the optimal bandwidth. ■

5. The Plug-in Method for Linear Combination

5.1. Simulations

In this section, the simulations is performed on twice and independently two independent samples with small ($n = 20$) and large ($n = 100$) samples from exponential distribution with mean one. Figure 1 includes the graphs for the cumulative distribution function $F(x)$ and the estimators using the methods: (1) the empirical, (2) the kernel distribution function $\hat{F}(x)$ using the Epanechnikov kernel, and (3) the linear combination of kernels $\hat{F^*}(x)$ using the kernels Epanenchnikov and Biweight with $a_1 = a_2 = 0.5$. The graphs show that the $\hat{F}$ and $\hat{F^*}$ are smoother than the empirical distribution function. $\hat{F^*}$ can be seen to be closer to $F(x)$ in the graphs with $n = 20$ and they all tend to coincide for $n = 100$.

5.2. Real-Life Data

Using the real world data at White et al. (1960), which is in Table 1, we will estimate the cumulative distribution function. The data were various characteristics of pure beeswax were studied. The beeswax samples were collected from 59 sources and for each sample the characteristics were studied. One of the studied characteristics is the melting point of each sample.

Table 1. The beeswax samples

63.78	63.45	63.58	63.08	63.4	64.52	63.27
63.1	63.34	63.5	63.83	63.63	63.27	63.3
63.83	63.5	63.36	63.86	63.34	63.92	63.88
63.36	63.36	63.51	63.51	63.84	64.27	63.5
63.56	63.39	63.78	63.92	63.92	63.56	63.43
64.21	64.24	64.12	63.92	63.53	63.5	63.3
63.86	63.93	63.43	64.4	63.61	63.03	63.68
63.13	63.41	63.6	63.13	63.69	63.05	62.85
63.31	63.66	63.6				

In Polansky and Baker (1999), the empirical distribution function and the kernel estimate of the distribution of the melting points was studied. We compare among the empirical distribution function, $\hat{F}(x)$ and $\hat{F}^{*}(x)$ using linear combinations of kernels. The weighted function is any function $w(x)$, such that $\int_0^1 w(x)dx = 1$, for simplicity and to have more accurate comparison we used the weighted function that was used by Polansky and Baker (1999), which is $w(x) = 1$, $0 < x < 1$, where other selections can be used and studied, but we will have a different form for the asymptotic optimal bandwidth.

Figures 2, 3 and 4 represent the estimate of the cumulative distribution function for the real life data example using the empirical estimate, the linear combination of kernels and using $\hat{F}(x)$. From these three figures, we noticed that the estimate using linear combination is close to the one with one kernel but it is still closer to the empirical estimator in the three cases, where we can conclude the linear combination may improve the estimation.

6. Conclusion

The kernel distribution estimator is a well known technique to estimate the distribution function $F(x)$, while the easiest technique to estimate $F(x)$ is the empirical estimator. The main disadvantage of the empirical technique that is the estimator is not smooth where the graph is look like

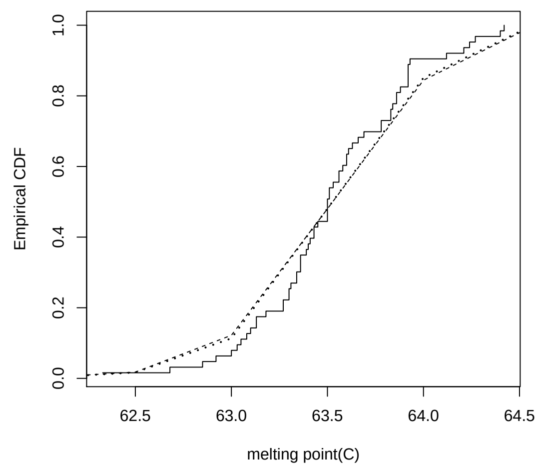

Figure 2. F_n(step function), $\hat{F}^*$ using Flin11(dotted line) and $\hat{F}$(dashed line)

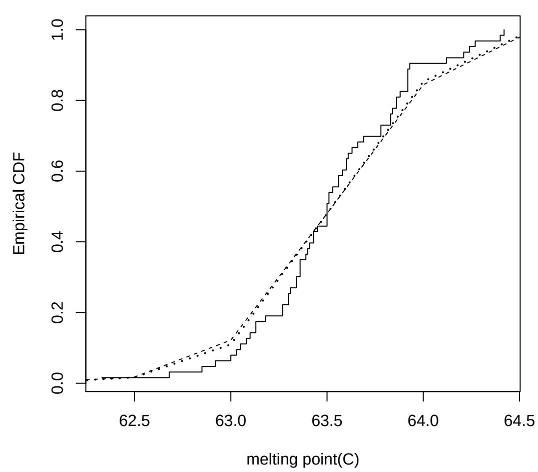

Figure 3. F_n(step function), $\hat{F}^*$ using Flin21(dotted line) and $\hat{F}$(dashed line)

stairs. Thus, it is for sure that the graph based on the empirical estimate is not the correct graph when the data is from a continuous distribution. To evaluate the estimator usually we minimize the error and mainly minimize the mean square error. In this paper we derived the mean integrated square error for the kernel distribution estimator using linear combinations of kernels. Also, we obtained the optimal bandwidth using the linear combinations of kernels. In addition, we proposed the plug in method for the linear combinations of kernels as a data based method to obtain the optimal bandwidth. From the figures we noticed that the method using linear combinations of kernels are closer to the well know empirical distribution function estimator than using a single kernel estimator. In addition, it is clear that the estimator using linear combinations is smoother than the empirical estimator which provides us with an indication that using linear combination of kernels is an effective way in the kernel distribution estimator.

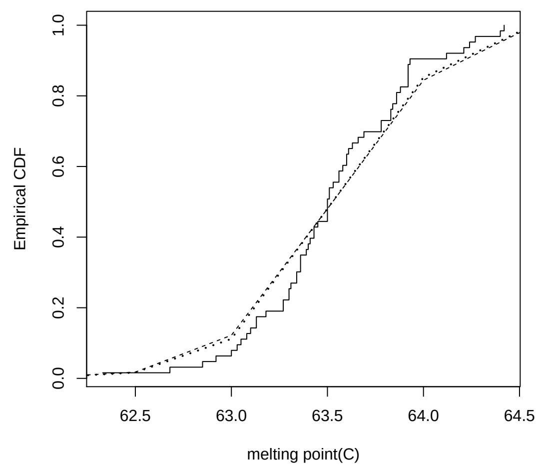

Figure 4. F_n(step function), $\hat{F}^*$ using Flin31(dotted line) and $\hat{F}$(dashed line)

On this investigation we used several regularity conditions on the probability density function, on the kernels and on the weighted function. Even, these conditions are well known and used in the kernel distribution estimator; an investigator may derive the mean integrated square error and study the bandwidth selection with different conditions.

This work can be extended to study other techniques for bandwidth selection such as the least square technique and the Biased in technique. Also, we can study other measures of error such as the mean weighted Hellinger Distance and the Mean Absolute Error. The linear combinations of kernels can be use to extended for the kernel estimator of function of observations.

Acknowledgment:

The authors thank the editor, the associate editor, the associate editor and the referees for their remarks on the earlier version of this article..

REFERENCES

Ahmad, I. A. and Mugdadi, A. R. (2006). Weighted Hellinger distance as an error criterion bandwidth selection in kernel estimation, Journal of Nonparametric Statistics, Vol. 18, No. 2, pp. 215-226.

Ahmad, I. A. and Ran, I. S. (2004). Kernel contrasts: A data-based method of choosing smoothing parameters in nonparametric density estimation, Journal of Nonparametric Statistics, Vol.16, No. 5, pp. 671-707.

Ahsanullah, M. and Habibullah, M. (2015). Estimation of the parameters of Exponential distribution using top-K-lists, Applications and Applied Mathematics: An International Journal, Vol. 10, pp. 71-80

Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function, Journal of Statistical Planning and Inference, Vol. 46, pp. 195–214.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method, Biometrika, Vol. 68, pp. 326-328.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives, Statistics and Probability Letters, Vol. 6, pp. 109-115.

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation, Statistics and Probability Letters, Vol. 9, pp. 129-132.

Mugdadi, A. R. and Jeter, J. (2010). A simulation study for the bandwidth selection in the kernel density estimation based on the exact and the asymptotic MISE, Pakistan Journal of Statistics, Vol. 26, No. 1, pp. 239-265.

Polansky, A.M. and Baker, E. R. (1999). Multistage plug-in bandwidth selection for kernel distribution function estimates, Journal of Statistical Computing Simulation, Vol. 65, pp. 63-80.

Rabhi, A. and Soltani, S. (2016). Exact asymptotic errors of the hazard conditional rate kernel for functional random fields, Applications and Applied Mathematics: An International Journal, Vol. 10, No. 2, pp. 527-558.

Sarda, P. (1993). Smooth parameter selection for smooth distribution functions, Journal of Statistical Planning and Inference, Vol. 35, pp. 65-75.

Savchuk O.Y., Hart, J. D. and Sheather, S. J. (2010). Indirect cross-validation for density estimation, Journal of the American Statistical Association, Vol. 105, No. 489, pp. 415-423.

Shakhatreh, M. K., Yusuf, A. and Mugdadi, A. R. (2016). The beta generalized linear exponential distribution, Statistics, Vol. 50, No. 6, pp. 1346-1362.

Sheather, S.J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimates, J. Roy. Stat. Soc. B, Vol. 53, pp. 683-690.

Swanepoel, J. W. H. (1988). Mean integrated squared error properties and optimal kernels when estimating a distribution function, Communications in Statistics: Theory and Methods, Vol. 17, pp. 3785-3799.

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

White. J., Riethof, M. and Kushnir, I. (1960). Estimation of microcrystalline wax in beeswax, Journal of the Association of Official Analytical Chemists, Vol. 43, pp. 781-790.

Woodroofe, M. (1970). On choosing a delta-sequence, The Annals of Mathematical Statistics, pp. 1665-16671.

Yuan, Fuging. (2018). Paramter estimation for bivariate Weibull distribution using generlized moment method for reliability evaluation, Quality and Relibility Engineering International, Vol. 34, No. 4, pp. 631-640.