

# Gram-Charlier A Series Based Extended Rule-of-Thumb for Bandwidth Selection in Univariate Kernel Density Estimation

**Bhaveshkumar C. Dharmani**  
School of Electronics & Electrical Engineering,  
Lovely Professional University

---

## Abstract

Bandwidth parameter estimation in univariate *Kernel Density Estimation* has traditionally two approaches. *Rule(s)-of-Thumb* (ROT) achieve ‘*quick and dirty*’ estimations with some specific assumption for an unknown density. More accurate *solve-the-equation-plug-in* (STEPI) rules have almost no direct assumption for the unknown density but demand high computation. This article derives a balancing third approach. Extending an assumption of Gaussianity for the unknown density to be estimated in *normal reference* ROT (NRROT) to near Gaussianity, and then expressing the density using *Gram-Charlier A* (GCA) series to minimize the *asymptotic mean integrated square error*, it derives GCA series based *Extended* ROT (GCAExROT). The performance analysis using the simulated and the real datasets suggests to replace NRROT by a modified GCAExROT rule achieving a balancing performance by accuracy nearer to STEPI rules at computation nearer to NRROT, specifically at small samples. Thus, the article also motivates to derive many such ExROTs based on *probability density function* approximations through other infinite series expansions.

*Keywords:* kernel density estimation (KDE), kernel bandwidth (smoothing) parameter, Gram-Charlier A (GCA) series, scale measures, cumulant measures, rule(s)-of-thumb (ROT), extended rule(s)-of-thumb (ExROT) .

---

## 1. Introduction

Kernel based nonparametric estimation of continuous *probability density function* (PDF) is widely used in statistics and data analysis. The estimation depends upon the selected kernel function and its spread decided by the smoothing or bandwidth parameter. The selection of kernel has limited impact on optimal PDF estimation, although Epanechnikov kernel is the most optimal (Epanechnikov 1969) if an optimal bandwidth parameter is assumed to be available. On the contrary, the optimal selection of bandwidth parameter avoids either too noisy (‘wiggly’) or too smooth an estimation.

There exists various bandwidth selection rules for *kernel density estimation* (KDE). They differ by the selected and the way estimated criterion to measure accuracy of density esti-

mation. *Asymptotic mean integrated squared error* (AMISE) between the estimated density ( $\hat{f}(\cdot)$ ) and the actual density ( $f(\cdot)$ ) is the most widely used performance criteria, though there are many others (Park and Turlach 1992). Estimation of bandwidth parameter minimizing AMISE requires estimating roughness of the second order derivative of the unknown density to be estimated ( $R(f''(\cdot))$ ), where the roughness of a function is defined as an integration of the squared function. The *rule(s)-of-thumb* (ROT) bandwidth selectors estimate  $R(f''(\cdot))$  assuming a reference distribution for  $f(\cdot)$  (Silverman 1986). The simplest and the most popular *normal reference* ROT (NRROT) assumes  $f(\cdot)$  to be a *normal* density. ROT are not among the most optimal, but are used either as a very fast reasonably good estimator or as a first estimator in multistage bandwidth selectors. As ROT depends upon the scale parameter, either *standard deviation* or *interquartile range*, of a distribution, they are called *scale measures* (Janssen, Marron, Veraverbeke, and Sarle 1995). Then, there are *plug-in* (PI) methods for bandwidth selection in KDE those plug-in kernel based estimate of  $R(f''(\cdot))$ . Estimation of any  $n$ th order derivative of  $f(\cdot)$ , denoted as  $(f^{(n)})$ ,  $n$  integer, in turn requires  $R(f^{(n+2)})$  estimate and a so called *pilot* bandwidth to start with. The *direct* PI methods estimate the *pilot* bandwidth with an assumption of some parametric family, usually Gaussian, for the unknown density (Woodroffe 1970). More accurate *solve-the-equation* PI (STEPI) methods estimate the *pilot* bandwidth as a nonlinear function of the bandwidth of interest (Sheather 1983, 1986a; Park and Marron 1990; Sheather and Jones 1991). Minimization of such a non-linear equation using iterative methods demand high computation of the order of  $O(N^2)$  usually, where  $N$  denotes number of samples. But Raykar and Duraiswami (Raykar and Duraiswami 2006) derived  $\epsilon$ -exact approximation algorithm based STEPI method, referred in this article as  $\epsilon$ STEPI, requiring  $O(N + M)$  order of computation, where  $M$  denotes selected number of evaluation points. There are also available some cross-validation (CV) techniques for bandwidth selection in univariate KDE (Scott and Terrell 1987; Jones, Marron, and Sheather 1996). Usually, they have high estimation variance and demand high computation. Also, very often they find an under-smooth bandwidth parameter (Marron 1996). A brief survey of data driven bandwidth selectors in univariate KDE is provided in (Park and Marron 1990; Jones *et al.* 1996; Park and Turlach 1992; Sheather 1986b; Sheather *et al.* 2004; Heidenreich, Schindler, and Sperlich 2013). There has also been some efforts towards exploring Bayesian approach, as well, localized bandwidth estimation for KDE (Kulasekera and Padgett 2006; Cheng, Gao, and Zhang 2019). Interested users may explore many other monographs and articles discussing the topic on univariate and multivariate KDE in a more generalized framework of non-parametric estimations and kernel smoothing (Silverman 1986; Wand and Jones 1994; Simonoff 1996; Scott 2012; Turlach *et al.* 1993; Izenman 1991; Hwang, Lay, and Lippman 1994; Ćwik and Koronacki 1997). Overall, the ROT or *scale measures* for bandwidth estimation are simple and fast but lack accuracy. And, the STEPI methods are quite accurate but demand high computation. So, a data dependent bandwidth parameter selection technique for KDE that balances both accuracy and computation is needed, and is the goal of this article.

The article achieves this goal by proposing a novel class of *Extended* ROT (ExROT) for bandwidth selection in univariate KDE. The assumption of a specific reference density for  $f(\cdot)$  to estimate  $R(f'')$  in ROT is too restrictive. Intuitively, expressing  $f(\cdot)$  in terms of an infinite series of higher order statistics based on a more generalized assumption to estimate  $R(f'')$  should result into a better bandwidth selection rule. As a verification and demonstration to this intuition, the article derives a bandwidth selection rule by extending the Gaussian assumption for  $f(\cdot)$  in NRROT to near Gaussian assumption. The near Gaussian assumption facilitates use of cumulants based *Gram-Charlier A* (GCA) series as an approximation to  $f(\cdot)$  for the estimation of  $R(f''(\cdot))$ . The extended assumption justifies the nomenclature ExROT for this class of bandwidth estimators and GCAExROT for this specific GCA based ExROT. Also, as they depend upon higher order cumulants unlike the *Scale Measures*, those just depend upon the scale parameter, they are named *Cumulant Measures* for bandwidth selection in KDE.

The empirical analysis considered performance measurement against both the simulated, with varying distributions and varying sample sizes from 50 to 50000, and real data sets using minimum MISE as the performance measure. It proved GCAExROT being better than NRROT for the KDE of unimodal normal mixture densities suggested as test densities by Marron and Wand (1992). But the same estimator proved to be the worst performer compare to the rest for KDE of multimodal normal mixture densities from the same set. The article analysed that the poor performance was due to the misleading estimations of *standard deviation* and *kurtosis* for multimodal densities. To overcome such an overestimation of *standard deviation*, Silverman (Silverman 1986) suggested a modified NRROT, identified as Silverman's ROT (SROT). Along the same lines, the article derives a modified GCAExROT, identified as GCAk4ExROT, rule. The empirical simulations proved GCAk4ExROT better than both the NRROT and SROT at any number of samples, as well, comparable to  $\epsilon$ STEPI at small samples. Importantly, the GCAExROT rules achieve such an improved performance at the computational cost comparable to that of ROT, which is quite less compare to that of  $\epsilon$ STEPI. The success of GCAk4ExROT over NRROT and SROT opens the scope of finding many such extended rules based on other existing infinite series expansions of a specific reference PDF. The rest of the article is organized as under. Section 2 discusses necessary background on KDE and AMISE criteria for bandwidth selection. Section 3 derives the novel GCAExROT rule for KDE of one dimension. It is empirically verified for its performance against varying density types at a sample size of 50000 in Section 4. The reasons for poor performance of GCAExROT in multimodal densities are analyzed and a modified GCAExROT rule is suggested in Section 5. The new rule is verified for its performance at steady state in the same section. Section 6 tests all the proposed bandwidth estimators against varying number of samples, varying distributions at small samples and against real datasets. Finally, Section 7 concludes the article.

## 2. Bandwidth selection criteria and selectors in univariate KDE

Let there be a random variable  $X$  with an unknown PDF  $f(x)$ ,  $x \in \mathbb{R}$ . Given  $N$  realizations  $x_1, x_2, \dots, x_N$  of  $X$ , the kernel density estimate  $\hat{f}(x)$  is given by

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h} \mathcal{K} \left( \frac{x - x_i}{h} \right), \quad (1)$$

where  $\mathcal{K}(x)$  is a kernel function and  $h$  is a bandwidth parameter deciding spread of the kernel. Usually,  $\mathcal{K}(x)$  is a symmetric PDF, i.e. it satisfies the following properties:

$$\mathcal{K}(x) \geq 0, \quad \int_{-\infty}^{\infty} \mathcal{K}(x) dx = 1, \quad \int_{-\infty}^{\infty} x \mathcal{K}(x) dx = \mu_1(\mathcal{K}) = 0, \quad \int_{-\infty}^{\infty} x^2 \mathcal{K}(x) dx = \mu_2(\mathcal{K}) < \infty.$$

The accuracy of a PDF estimation can be quantified by available distance measures between PDFs, such as the mean integrated absolute error, the MISE, the *Kullback-Leibler divergence* and many others. Minimizing a selected distance measure over varying  $h$  obtains an optimal  $h$ . The most widely used criteria MISE or IMSE (Integrated Mean Square Error) is given as

$$\text{MISE}(\hat{f}(x)) = E \left\{ \int_{-\infty}^{\infty} (\hat{f}(x) - f(x))^2 dx \right\} = \int_{-\infty}^{\infty} \text{Bias}^2(\hat{f}(x)) dx + \int_{-\infty}^{\infty} \text{Var}(\hat{f}(x)) dx. \quad (2)$$

The assumptions that the underlying unknown density is sufficiently smooth, i.e., all derivatives exist, the kernel has finite fourth order moment, as well,  $\lim_{N \rightarrow \infty} h = 0$  and  $\lim_{N \rightarrow \infty} Nh = \infty$ , i.e.,  $h$  reduces to 0 at a rate slower than  $1/N$ , obtains (Silverman 1986) an asymptotic large sample approximation of MISE as

$$\text{AMISE}(\hat{f}(x)) = \frac{h^4}{4} \mu_2^2(\mathcal{K}) R(f'') + \frac{1}{Nh} R(\mathcal{K}), \quad (3)$$

where  $R(f'') = \int_{\mathbb{R}} (f''(x))^2 dx$  and  $R(\mathcal{K}) = \int_{\mathbb{R}} \mathcal{K}^2(x) dx$ . In general,  $R(g) = \int g^2(x) dx$  is identified as the roughness of function  $g(x)$ . The detailed derivation of AMISE can be found in Appendix 7. Equation (3) proves that the smaller bandwidth values will reduce the estimation bias but increase the estimation variance and vice-versa. An optimal  $h$  minimizing AMISE( $\hat{f}(x)$ ) in Equation (3) can be given as:

$$\begin{aligned} h_{AMISE} &= \left( \frac{R(\mathcal{K})}{\mu_2^2(\mathcal{K})R(f'')N} \right)^{\frac{1}{5}} \\ &= (CN)^{-1/5}, \quad \text{where } C = \frac{\mu_2^2(\mathcal{K})R(f'')}{R(\mathcal{K})}. \end{aligned} \quad (4)$$

Various rules for bandwidth parameter selection based on minimising AMISE criteria differ in the way  $R(f'')$ , the only unknown parameter in Equation (4), is estimated. With the assumption of Gaussian PDF for  $f(x)$ ,  $R(f'') = \frac{3\sigma^{-5}}{8\sqrt{\pi}}$ , where  $\sigma$  is the *standard deviation* of  $f(x)$ . Also, for a standard Gaussian kernel  $\mu_2(\mathcal{K}) = 1$  and  $R(\mathcal{K}) = \frac{1}{2\sqrt{\pi}}$ . Using them with Equation (4) obtains the simplest NRROT as

$$h_{NR} = 1.0592\sigma N^{-1/5}. \quad (5)$$

For more details on bandwidth selectors, refer (Silverman 1986; Wand and Jones 1994; Simonoff 1996).

### 3. The Gram-Charlier A series based extended ROT

Bandwidth parameter selection in a one dimensional KDE using AMISE criteria needs the estimation of roughness of second order derivative of the unknown density to be estimated. Instead of Gaussianity, near Gaussianity assumption for an unknown density will allow to approximate it by GCA series expansion. The GCA series expands an unknown PDF as a linear combination of the increasing order differentiations of Gaussian reference PDF, where the coefficients of expansion involve cumulant differences between those of an unknown PDF and a Gaussian reference PDF.

Let there be a random variable  $X$  with PDF  $f(x)$ ,  $x \in \mathbb{R}$ , *characteristic function*  $\mathcal{F}_x$  and *cumulant generating function* (CGF)  $\mathcal{C}(\lambda)$ . By definition,

$$\mathcal{F}_x(\lambda) = \exp(\mathcal{C}(\lambda)) = \exp \left[ \sum_{k=1}^{\infty} c_k \frac{(i\lambda)^k}{k!} \right], \quad (6)$$

where  $c_k$  denotes  $k$ th order cumulant of  $f(x)$ . Let the Gaussian reference PDF  $G(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2 \right]$ , where  $\mu$  is the *mean* parameter and  $\sigma$  is the *standard deviation*, be with  $\gamma_k$  as the  $k$ th order cumulant. Let  $c_k - \gamma_k = \delta_k$  or  $c_k = \gamma_k + \delta_k$ . Also, it is known that the *characteristic function* and the PDF are the Fourier transform (F) of each other, in the sense they are dual. Then, re-writing Equation (6) obtains

$$\begin{aligned} \mathcal{F}_x(\lambda) &= \mathbf{F}(f(x)) = \exp \left[ \sum_{k=1}^{\infty} \delta_k \frac{(i\lambda)^k}{k!} \right] \exp \left[ \sum_{k=1}^{\infty} \gamma_k \frac{(i\lambda)^k}{k!} \right] = \left[ \sum_{k=0}^{\infty} \alpha_k \frac{(i\lambda)^k}{k!} \right] \mathbf{F}(G(x; \mu, \sigma)), \quad (7) \\ \text{where } \sum_{k=0}^{\infty} \alpha_k \frac{(i\lambda)^k}{k!} &= \exp \left[ \sum_{k=1}^{\infty} \delta_k \frac{(i\lambda)^k}{k!} \right]. \quad (8) \end{aligned}$$

It is evident from Equation (8) that the  $\alpha_k$  are related to the cumulant differences  $\delta_k$  in the same manner as the moments are related to the cumulants. Taking inverse Fourier transform ( $\mathbf{F}^{-1}$ ) of the above Equation (7) and using the property  $\mathbf{F}^{-1}((i\lambda)^k \mathbf{F}(G(x))) = (-1)^k D^{(k)}G(x)$ ,

where  $D$  denotes derivative operator with respect to  $x$  and  $D^{(k)}$  denotes  $k$ th order derivative operator, obtains GCA series expansion of  $f(x)$  as

$$f(x) = \sum_{k=0}^{\infty} \alpha_k \frac{(-D)^{(k)}G(x)}{k!}$$

$$\text{or } f(x) = \exp \left[ \sum_{k=0}^{\infty} (c_k - \gamma_k) \frac{(-D)^k G(x; \mu, \sigma)}{k!} \right], \quad (9)$$

where  $G^{(k)}(\cdot)$  denotes  $k$ th order derivative of Gaussian. The derivation is a revised version, using a Gaussian density as a reference density, of the Generalized Gram-Charlier (GGC) series expansion derived in a compact way by Berberan-Santos (Berberan-Santos 2007). There are many other ways to derive univariate GCA series (Hald 2000) or univariate Generalized Gram-Charlier (GGC) series (Schleher 1977; Cohen 1998, 2011).

The  $k$ th order derivative of Gaussian is given by

$$D^{(k)}G(x; \mu, \sigma) = \frac{1}{\sigma^k} (-1)^k H_k \left( \frac{x - \mu}{\sigma} \right) G(x; \mu, \sigma), \quad (10)$$

where  $H_k$  is the  $k$ th order Hermite polynomial, expressed using recursion as

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_{k+1}(x) = xH_k(x) - \frac{d}{dx}H_k(x). \quad (11)$$

Choosing the first and second order cumulants of the unknown PDF to be the same as those of the reference Gaussian PDF, i.e.  $c_1 = \gamma_1 = \mu$  and  $c_2 = \gamma_2 = \sigma^2$ , and taking approximation upto fourth order cumulants obtain

$$f(x) \approx \left[ 1 + \frac{c_3}{3!\sigma^3} H_3 \left( \frac{x - \mu}{\sigma} \right) + \frac{c_4}{4!\sigma^4} H_4 \left( \frac{x - \mu}{\sigma} \right) \right] G(x; \mu, \sigma). \quad (12)$$

Taking derivative twice with respect to  $x$  of the Equation (9) for GCA series yields:

$$f''(x) = \exp \left[ \sum_{k=0}^{\infty} (c_k - \gamma_k) \frac{(-D)^{(k+2)}G(x; \mu, \sigma)}{k!} \right]. \quad (13)$$

Choosing  $c_1 = \gamma_1 = \mu$ ,  $c_2 = \gamma_2 = \sigma^2$ , approximation up to fourth order cumulants and using  $z = \frac{x-\mu}{\sigma}$  obtain

$$f''(x) \approx \left[ \frac{1}{\sigma^2} H_2(z) + \frac{c_3}{3!\sigma^5} H_5(z) + \frac{c_4}{4!\sigma^6} H_6(z) \right] G(x; \mu, \sigma). \quad (14)$$

By definition,

$$R(f'') = \int_{\mathbb{R}} \left[ \frac{1}{\sigma^2} H_2(z) + \frac{c_3}{3!\sigma^5} H_5(z) + \frac{c_4}{4!\sigma^6} H_6(z) \right]^2 G^2(x; \mu, \sigma) dx. \quad (15)$$

The simplification is obtained using the popularly known *Gaussian Integral* and its variants, mention as under:

$$\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}} \quad (16)$$

$$\int_{-\infty}^{\infty} x^k e^{-ax^2} dx = \begin{cases} \frac{(k)!!}{(2a)^{k/2}} \sqrt{\frac{\pi}{a}}, & \text{for } k \text{ even} \\ 0, & \text{for } k \text{ odd,} \end{cases} \quad (17)$$

where  $(k)!! = (k-1)(k-3)(k-5)\cdots 1$  is called the *odd factorial* of  $k$ . Combining these integrals with Equation (11) obtains

$$\begin{aligned}
T_1 &= \int_{-\infty}^{\infty} G^2(z) \left[ \frac{1}{\sigma^2} H_2(z) \right]^2 dx = \frac{1}{\sigma^5} \frac{3}{8\sqrt{\pi}}, \\
T_2 &= \int_{-\infty}^{\infty} G^2(z) \left[ \frac{1}{\sigma^5} H_5(z) \right]^2 dx = \frac{1}{\sigma^{11}} \frac{945}{64\sqrt{\pi}}, \\
T_3 &= \int_{-\infty}^{\infty} G^2(z) \left[ \frac{1}{\sigma^6} H_6(z) \right]^2 dx = \frac{1}{\sigma^{13}} \frac{10395}{128\sqrt{\pi}}, \\
T_4 &= \int_{-\infty}^{\infty} G^2(z) \left[ \frac{2}{\sigma^7} H_2(z) H_5(z) \right] dx = 0, \\
T_5 &= \int_{-\infty}^{\infty} G^2(z) \left[ \frac{2}{\sigma^{11}} H_5(z) H_6(z) \right] dx = 0, \\
T_6 &= \int_{-\infty}^{\infty} G^2(z) \left[ \frac{2}{\sigma^8} H_2(z) H_6(z) \right] dx = \frac{1}{\sigma^9} \frac{105}{16\sqrt{\pi}}.
\end{aligned} \tag{18}$$

Simplifying Equation (15) using Equation (18) brings

$$R(f'') = \frac{1}{\sigma^5} \frac{3}{8\sqrt{\pi}} [1 + 1.0938k_3^2 + 0.3760k_4^2 + 0.7292k_4], \tag{19}$$

where  $k_3 = c_3/\sigma^3$  denotes *skewness* and  $k_4 = c_4/\sigma^4$  denotes *kurtosis* (truly speaking, *excess kurtosis*) of a random variable. Using  $R(f'')$  in Equation (19) into Equation (4) obtains the *Gram-Charlier A* (GCA) series based *Extended Rule-of-Thumb* (GCAExROT) for bandwidth selection in one dimensional KDE using Gaussian kernel as under:

$$\begin{aligned}
h_{\text{GCA}} &= 1.0592\sigma(CN)^{-\frac{1}{5}}, \\
\text{where } C &= 1 + 1.0938k_3^2 + 0.3760k_4^2 + 0.7292k_4.
\end{aligned} \tag{20}$$

The constant  $C$  obtained for some of the special cases, and described as under, better explains GCAExROT:

$$C = \begin{cases} 1, & \text{if both } k_3 = k_4 = 0, \text{ i.e., Gaussian PDF} \\ 1 + 0.3760k_4^2 + 0.7292k_4, & \text{if } k_3 = 0, \text{ i.e., symmetric PDF} \\ 1 + 1.0938c_3^2 + 0.3760c_4^2 + 0.7292c_4, & \text{if } \sigma = 1. \end{cases} \tag{21}$$

As evident, NRROT is one special case of GCAExROT when  $k_3 = k_4 = 0$ .

#### 4. Verification of the GCAExROT bandwidth selector

To verify empirical validity of the derived GCAExROT bandwidth selector, a simple experiment was done. The experiment measures the bandwidth estimator's performance on KDE of simulated sources with varying distributions. There was selected a set of 15 normal mixture densities, used as a test-bed for bandwidth estimators in one dimensional KDE by Marron and Wand (Marron and Wand 1992) and shown in Figure 1. To understand the steady state performance, the number of samples were kept 50000. There were considered 100 simulation trails. GCAExROT performance was compared with those of NRROT, as the representative of ROT methods, and  $\epsilon$ STEPI, as the representative of STEPI methods. The performance comparison was against three parameters – the value of estimated bandwidth, the MISE between the estimated and the actual PDFs and the actual time taken (in Seconds) to estimate the bandwidth. The *Speed-up* achieved by GCAExROT over  $\epsilon$ STEPI is also used for comparison, where the *Speed-up* is defined as the ratio of estimation time by  $\epsilon$ STEPI to estimation time by GCAExROT.

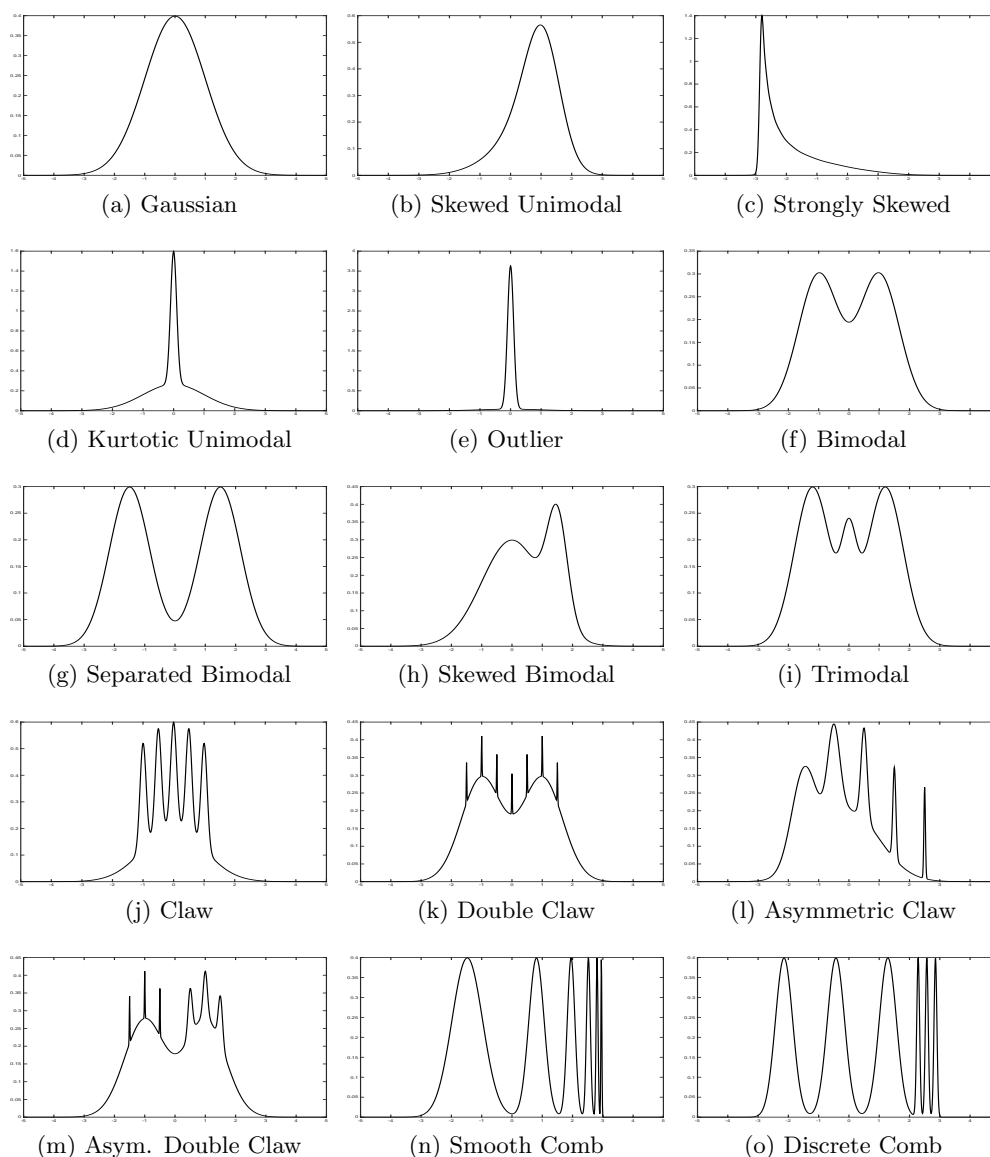


Figure 1: The normal mixture densities proposed by Marron and Wand (Marron and Wand 1992) as a test-bed for performance comparison of various bandwidth selection rules for Kernel Density Estimation (KDE)

Table 1 shows the results as mean of 100 trials. The boldface values denote the best per row, i.e., the best bandwidth estimator giving the least MISE for a density in the first column. The boldface italic values in the table denote a better between NRROT and GCAExROT with the same criteria. The results exhibit  $\epsilon$ STEPI as the best bandwidth estimator at 50000 samples giving the least MISE, except for Gaussian and slightly skewed unimodal distributions. For all unimodal distributions other than the Gaussian, i.e., for skewed, strongly skewed, kurtotic and outlier distributions, GCAExROT has been better than NRROT. Among them, for skewed unimodal distribution it is even better than  $\epsilon$ STEPI. For all multimodal distributions, except the claw distribution, NRROT performance is better than that of GCAExROT. The bandwidth comparison in the table shows that for all distributions, except the Gaussian, smaller the bandwidth better the performance in terms of smaller the MISE. The time comparison in the same table shows that the mean time to estimate the bandwidth parameter using NRROT is of the order of a few tenths of millisecond. This includes the time to estimate *standard deviation* for the rule. The estimation time for GCAExROT is slightly more than that for NRROT and in the range of a few milliseconds at the given 50000 samples. The added extra time includes the time to estimate the third and fourth ordered cumulants. The bandwidth parameter estimation time for  $\epsilon$ STEPI is of the order of a few tens of second. The mean

Table 1: Performance comparison of *normal reference Rule-of-Thumb* (NRROT), Gram-Charlier A series based *extended* ROT (GCAExROT) and  $\epsilon$ -exact approximation algorithm based solve-the-equation plug-in ( $\epsilon$ STEPI) bandwidth selectors for one dimensional Kernel Density Estimation (KDE) against the *normal mixture densities* by Marron and Wand (Marron and Wand 1992) based on the estimated bandwidths, the *mean of the integrated square error* (MISE) between the estimated and the actual densities and the time to estimate the bandwidth were used as the performance measures. There were used 50000 samples. The table entries show the mean of 100 trials. The boldface MISE values denote the smallest and the boldface italic values denote the smaller between those due to NRROT and ExROT corresponding to the density in each row.

PDF	Bandwidth (h)*10			MISE*10 <sup>6</sup>			Time (in Sec)			Mean Speed-up
	NR ROT	GCA Ex ROT	$\epsilon$ STE PI	NR ROT	GCA Ex ROT	$\epsilon$ STE PI	NR ROT *10 <sup>4</sup>	GCA ExROT *10 <sup>3</sup>	$\epsilon$ STE PI	
a	1.217	1.216	1.214	<b>1.403</b>	1.404	1.406	14.059	9.145	7.079	828.1
b	0.993	0.811	0.819	1.801	<b>1.729</b>	1.730	9.154	8.266	10.701	1334.3
c	1.263	0.889	0.200	36.988	<b>26.100</b>	<b>4.038</b>	8.401	8.350	54.205	6871.6
d	0.995	0.807	0.203	30.331	<b>22.731</b>	<b>3.740</b>	8.671	8.408	62.240	7613.2
e	0.401	0.138	0.129	19.221	<b>4.435</b>	<b>4.421</b>	9.355	8.525	78.986	9773.0
f	1.463	1.596	0.970	<b>2.024</b>	2.268	<b>1.588</b>	8.696	8.389	10.246	1261.3
g	1.998	2.137	0.924	<b>3.676</b>	4.140	<b>1.649</b>	8.836	8.421	10.981	1340.7
h	1.332	1.383	0.737	<b>3.093</b>	3.271	<b>1.897</b>	8.336	8.392	15.117	1872.4
i	1.552	1.690	0.788	<b>3.453</b>	3.931	<b>1.796</b>	7.487	8.336	11.518	1438.7
j	1.057	1.052	0.243	23.759	<b>23.637</b>	<b>3.515</b>	12.339	8.613	46.385	5761.9
k	1.457	1.590	0.900	<b>7.798</b>	7.910	<b>7.363</b>	7.630	8.479	10.779	1313.4
l	1.355	1.361	0.322	<b>16.639</b>	16.683	<b>5.282</b>	10.736	9.447	49.793	5489.7
m	1.449	1.576	0.462	<b>11.183</b>	11.681	<b>5.806</b>	9.139	8.970	23.853	2765.2
n	1.999	2.111	0.280	<b>32.276</b>	33.219	<b>8.822</b>	19.356	8.353	33.923	4212.8
o	2.057	2.224	0.230	<b>32.408</b>	33.856	<b>4.637</b>	16.801	10.363	36.925	4463.6

*Speed-up* by GCAExROT over the  $\epsilon$ STEPI is of the order of a few thousands. Throughout the article, the timings are just to provide relative references and are on a machine Intel 2.5 GHz core i5 with 4 GB 1.6 GHz DDR3 RAM and MATLAB R2014b. The operating system for the experiments on a set of normal mixture densities by (Marron and Wand 1992) was MacOS X EI capitan (version 10.11.6) and for the experiments on a set of Generalized Gaussian densities (GGD) was MacOS Mojave (version 10.14.4).

Overall, GCAEXROT has achieved intermediate level accuracy in bandwidth estimations for KDE of non-Gaussian univariate densities but the same proposed bandwidth estimator has performed the worst, even poorer than NRROT, for the multimodal densities in the test. The reasons must be analysed and if possible the bandwidth estimator should be improved. The next session tries to understand how the higher order cumulants affect the bandwidth estimated by GCAExROT and how it can be modified to get the improved performance.

## 5. GCAExROT: understanding and modification

Equation (20) interprets that  $h_{\text{GCA}}$  adjusts  $h_{\text{NR}}$  in Equation (5) for *kurtosis* and *skewness* through the constant  $C^{-1/5}$ , which works as a *scaling coefficient*.

As in Equation (21), given  $k_3 = k_4 = 0$  causes the scaling coefficient  $C = 1$  and GCAExROT is the same as NRROT. The left panel of the Figure 2 plots the *skewness* versus the *scaling coefficient*  $C^{-1/5}$  keeping  $k_4 = 0$  in Equation (20), i. e., considering the effect of *skewness* parameter independent of the kurtosis parameter. The plot shows that the *skewness*, whether positive or negative, scales down  $h_{\text{NR}}$  to give effectively  $h_{\text{GCA}} < h_{\text{NR}}$ .

Similarly, the right panel in Figure 2 shows the plot of *kurtosis* versus *scaling coefficient*  $C^{-1/5}$  keeping  $k_3 = 0$ , i. e., the effect of *kurtosis* independent of the *skewness*. The solid line plot corresponds to GCAExROT and can be understood in four parts. Given the unknown density has  $k_4 > 0$ , i.e., it is superGaussian, implies  $C^{-1/5} < 1$  and  $h_{\text{GCA}} < h_{\text{NR}}$ . Given the unknown density is subGaussian with  $0 > k_4 > -1.9394$ , implies  $C < 1$ ,  $C^{-1/5} > 1$  and  $h_{\text{GCA}} > h_{\text{NR}}$ . Also within this range, for  $0 > k_4 > -1$ ,  $h_{\text{GCA}}$  is increasing but thereafter it



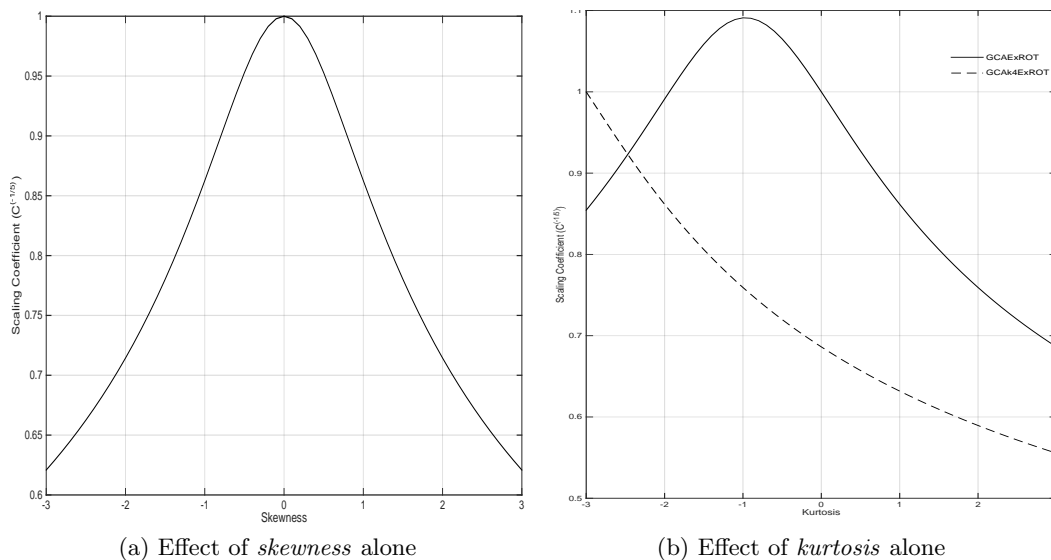


Figure 2: Scaling factor  $C^{-1/5}$  independently due to: (a) Skewness ( $k_3$ ) in GCAExROT and its modification, (b) Kurtosis ( $k_4$ ) in GCAExROT (Solid Line) and GCAk4ExROT (Dashed Line).

is decreasing. Further,  $h_{GCA} < h_{NR}$  for densities with  $k_4 < -1.9394$ , those giving  $C^{-1/5} < 1$ . GCAExROT was expected to be better than NRROT due to its wider assumption of near-Gaussianity on unknown density. The assumption fits well the unimodal normal mixture densities among the test-bed densities and that's why empirically for these densities GCAExROT gave better accuracy than NRROT. This justifies the bandwidth parameter estimated and in turn the scaling provided by GCAExROT over NRROT based on the skewness and the kurtosis parameters. But then why not the similar performance repeated for multimodal distributions? The comparison of estimated bandwidths in Table 1 reveals that the bandwidth parameters estimated by GCAExROT for multimodal densities were larger than corresponding those by NRROT. This confirms that the *kurtosis* estimated was in the range  $0 > k_4 > -1.9394$  as in Figure 2 (b). Examination of the *kurtosis* value of the generated multimodal test densities reconfirmed this, empirically. This implies that the combination of multiple modes in these densities resulted into underestimated *kurtosis* and sub-Gaussianity. Thus, the estimated *kurtosis* failed to represent the local nature of the densities correctly, which consequently, resulted into wrong bandwidth and poor density estimates. Similarly, the *standard deviation* also gets overestimated locally in case of multimodal distributions. Summarizing, the globally estimated cumulants as per their definitions, like *standard deviation* and *kurtosis*, cause the estimated bandwidth to be wrong locally.

### 5.1. Modified GCAExROT proposal

Ideal remedy to get the correct bandwidth through GCAExROT would be to estimate the *kurtosis* and *standard deviation* parameters locally for all the modes in case of KDE of multimodal density. But this will add more complexities and not the goal of a simple ROT and ExROT class of bandwidth estimators. Instead, a correction mechanism to avoid such overestimations or underestimations, validated on empirical and real data, would be a more suitable solution. Historically, to overcome oversmoothing of NRROT due to the overestimation of *standard deviation*, Silverman suggested a correction mechanism. The modified rule, identified as the Silverman's ROT (SROT), simply scales down ROT and given as :

$$h_{SROT} = 0.9\sigma N^{-1/5}. \quad (22)$$

Along the same line, this section derives a modified GCAExROT rule by providing corrections to Equation (20). The overestimation of *standard deviation* can be taken care in the same

way as done for SROT, i.e., the coefficient 1.0592 should be replaced by 0.9. The correction for underestimated kurtosis should take care of two aspects: for a given density having local modes with positive *kurtosis* either the estimated global kurtosis is reduced but still positive or the estimated value is reduced to such an extent that it becomes negative. For example, if the absolute value of estimated *kurtosis* is considered as a correction mechanism, then it will take care of the latter condition but not the former. So, as a remedy and correction to the underestimated *kurtosis*, the article suggests to use the actual *kurtosis* value instead of the *excess kurtosis* value, i.e.,  $k_4 + 3$ . Accordingly, the negatively estimated *kurtosis* will be turned positive and the positively estimated *kurtosis* will be increased. Thus, the remedy takes care of both the aspects of underestimated *kurtosis*. Also, instead of verifying whether the distribution is multimodal or unimodal before the actual application of bandwidth selection rule, the correction is always applied to match the goal of simplification. This defines the modified rule GCAk4ExROT as

$$h_{\text{GCAk4}} = 0.9\sigma(CN)^{-\frac{1}{5}}, \quad (23)$$

$$\text{where } C = 1 + 1.0938k_3^2 + 0.3760(k_4 + 3)^2 + 0.7292(k_4 + 3).$$

The dashed line plot in the right panel of Figure 2 shows *kurtosis* versus *scaling coefficient*  $C^{-1/5}$  in GCAk4ExROT. As can be seen, for Gaussian density, i.e.  $k_4 = 0$ , GCAk4ExROT gives  $C = 6.5716$  and thus scales down  $h_{\text{SROT}}$  by a factor of  $C^{-1/5} = 0.6862$ . This may degrade the performance for Gaussian and slightly near-Gaussian densities. Thus, the new rule should be verified for its performance on the standard test-bed, as well, specifically for densities with *kurtosis* range near that of Gaussian.

## 5.2. Performance of GCAk4ExROT on the standard test-bed

This section verifies the modified GCAExROT bandwidth selection rule, GCAk4ExROT, for its steady state performance at sample size of 50000. The experiment was repeated for the new rules on the set of *normal mixture densities* in Figure 1. Comparison of MISE between the actual and the estimated densities due to the estimated bandwidths by NRROT, GCAExROT, SROT, GCAk4ExROT and  $\epsilon$ STEPI bandwidth selection rules are shown in Table 2. The estimated bandwidths by SROT and GCAk4ExROT are also reported in the same table. As the bandwidths by rest of the bandwidth selectors have already been reported, they are not repeated again. The table entries denote the mean of 100 trials. The boldface values show the least MISE per row, i.e., the best bandwidth estimator for a given density type. The boldface italic values denote the best bandwidth estimator among all except  $\epsilon$ STEPI to provide comparison among the ROT and the ExROT class bandwidth estimators.

It can be observed that GCAk4ExROT has been the second best among all the bandwidth estimators and the best among ROT and ExROT class of bandwidth estimators for the KDE of all the multimodal densities, except the Bimodal density for which it has been the best among all. But this performance of GCAk4ExROT is at the cost of slight loss of performance for Gaussian and slightly skewed unimodal densities compare to GCAExROT, as expected. The results show GCAk4ExROT performance slightly degraded even for the outlier density, but it should be noted that the table entries for MISE are of the order of  $10^{-6}$  and both GCAExROT and GCAk4ExROT have their performances better than both NRROT and SROT. Overall, the experiment here validates that the modified GCAExROT rule has improved performance over GCAExROT for its steady state behaviour at 50000 samples. This motivates to test it further for its performance.

## 5.3. Performance of GCAk4ExROT on densities with varying Kurtosis

This section verifies the GCAk4ExROT for KDE of densities with varying *kurtosis*, as they represent more ideal set of densities and explain the bandwidth corrections over GCAExROT. The densities were generated using symmetrical and unimodal Generalized Gaussian Density (GGD) class of densities with varying the shape parameter  $\beta$ . A density with  $\beta$  value of 2

Table 2: Comparison of mean integrated square error (MISE) between the estimated and actual densities, where estimated densities were using bandwidth selection rules from *normal reference Rule-of-Thumb* (NRROT), Gram-Charlier A series based *extended* ROT (GCAExROT), Silverman’s ROT (SROT), GCAk4ExROT and  $\epsilon$ -exact approximation algorithm based solve-the-equation plug-in ( $\epsilon$ STEPI) bandwidth selectors. The unknown test densities were from the set of *normal mixture densities* by Marron and Wand (Marron and Wand 1992). There were used 50000 samples. The table entries show the mean of 100 trials. The boldface values denote the least MISE per row for a corresponding density. The boldface italic values denote the best among ROT and ExROT class of bandwidth selection rules.

PDF	Bandwidth (h)*10		MISE*10 <sup>6</sup>				
	GCAk4		NRROT	GCA ExROT	SROT	GCAk4 ExROT	$\epsilon$ STEPI
	SROT	ExROT					
Gaussian	0.709	1.214	<b>1.403</b>	1.404	1.450	1.722	1.406
Skewed Unimodal	0.525	0.819	1.801	1.729	<b>1.727</b>	1.996	1.730
Strongly Skewed	0.622	0.200	36.988	26.100	31.751	<b>16.854</b>	<b>4.038</b>
Kurtotic Unimodal	0.517	0.203	30.331	22.731	24.327	<b>11.258</b>	<b>3.740</b>
Outlier	0.112	0.129	19.221	<b>4.435</b>	14.296	4.495	<b>4.421</b>
Bimodal	0.939	0.970	2.024	2.268	1.728	<b>1.584</b>	1.588
Separated Bimodal	1.350	0.924	3.676	4.140	2.794	<b>2.030</b>	<b>1.649</b>
Skewed Bimodal	0.816	0.737	3.093	3.271	2.481	<b>1.925</b>	<b>1.897</b>
Trimodal	1.013	0.788	3.453	3.931	2.722	<b>2.012</b>	<b>1.796</b>
Claw	0.614	0.243	23.759	23.637	19.239	<b>10.888</b>	<b>3.515</b>
Double Claw	0.936	0.900	7.798	7.910	7.627	<b>7.393</b>	<b>7.363</b>
Asymmetric Claw	0.810	0.322	16.639	16.683	14.913	<b>11.559</b>	<b>5.282</b>
Asym. Double Claw	0.931	0.462	11.183	11.681	10.169	<b>8.493</b>	<b>5.806</b>
Smooth Comb	1.338	0.280	32.276	33.219	29.591	<b>25.967</b>	<b>8.822</b>
Discrete Comb	1.367	0.230	32.408	33.856	29.973	<b>27.264</b>	<b>4.637</b>

denote Gaussian distribution,  $\beta < 2$  denote super-Gaussian distribution and  $\beta > 2$  denote sub-Gaussian distribution. Figure 3 (a) shows the super-Gaussian densities, i.e., GGD( $\beta$ ) with  $\beta \leq 2$  and Figure 3 (b) shows the sub-Gaussian densities, i.e., GGD( $\beta$ ) with  $\beta \geq 2$ . The figure also shows corresponding *kurtosis* values obtained as an empirical mean of 100 simulation trials. Simulation results with the sample size of 50000 (large) have been reported

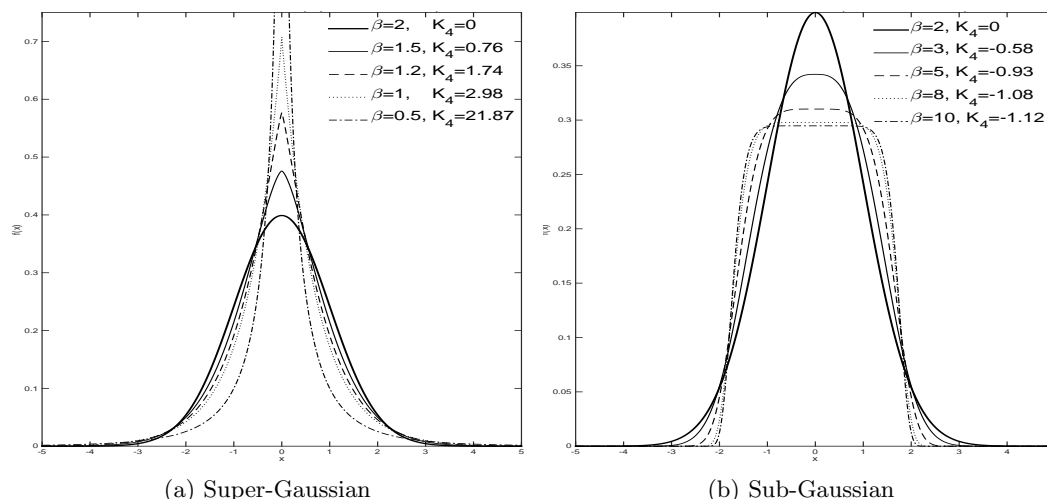


Figure 3: The set of Probability density functions (PDFs) having Generalized Gaussian Distribution (GGD) with varying values of  $\beta$  to have performance comparison of various bandwidth selection rules for Kernel Density Estimation (KDE).

in in Table 3. It shows the *kurtosis* calculated as a mean of all 100 simulation trails for the given density, the mean of the estimated bandwidth parameter ( $h$ ) and the MISE by NRROT, GCAExROT, SROT, GCAk4ExROT and  $\epsilon$ STEPI bandwidth selection algorithms. The boldface values denote the least MISE and in turn the best bandwidth estimator for that particular type of density. The boldface italic values denote a better between NRROT and GCAExROT in their mutual comparison.

Table 3: Performance comparison of *normal reference Rule-of-Thumb* (NRROT), Gram-Charlier A series based *extended* ROT (GCAExROT) and  $\epsilon$ -exact approximation algorithm based solve-the-equation plug-in ( $\epsilon$ STEPI) bandwidth selectors for one dimensional Kernel Density Estimation (KDE) against densities following Generalized Gaussian Distribution (GGD) with specific  $\beta$  values from the set  $\{1, 1.2, 1.5, 2, 5, 8, 10\}$ . Estimated bandwidth, *mean of the integrated square error* (MISE) between the estimated and the actual densities and time to estimate the bandwidth were used as the parameters for comparison. There were used 50000 (large) samples. The table entries show the mean of 100 trials. The boldface values denote the least MISE per row for a corresponding density. The boldface *italic* values denote better between NRROT and GCAExROT.

PDF GGD ( $\beta$ )	Kurt osis	Bandwidth (h)* $10^{-4}$					MISE* $10^5$				
		NR	GCA	SROT	GCA	$\epsilon$ STE	NR	GCA	SROT	GCA	$\epsilon$ STE
		ROT	Ex ROT		k4Ex ROT	PI PI	ROT	Ex ROT		k4Ex ROT	PI PI
1	2.98	5.863	4.035	4.982	2.772	2.748	6.883	<b><i>6.805</i></b>	6.843	<b><i>6.772</i></b>	6.772
1.2	1.74	11.004	8.578	9.350	5.594	6.880	5.208	<b><i>5.169</i></b>	5.179	<b><i>5.146</i></b>	5.150
1.5	0.76	3.565	3.442	3.029	2.036	3.389	<b><i>4.278</i></b>	4.282	4.294	4.318	4.284
2	0	4.831	5.185	4.105	2.981	4.909	6.314	<b><i>6.308</i></b>	6.326	6.340	6.313
5	-0.93	1.843	2.008	1.566	1.200	1.244	8.528	<b><i>8.515</i></b>	8.548	8.571	8.568
8	-1.08	2.346	2.553	1.993	1.533	1.430	8.794	<b><i>8.779</i></b>	8.816	8.842	8.846
10	-1.12	3.580	2.939	3.042	1.872	2.454	3.949	<b><i>3.929</i></b>	3.931	3.929	3.923

The table shows that compare to NRROT, GCAExROT gave less MISE for all GGD( $\beta$ ) set of densities, except the Gaussian, at large sample size of 50000. This reconfirms empirically that the near-Gaussian assumption worked better than the Gaussian assumption for an unknown unimodal density. In turn, this reconfirms the scaling provided by GCAExROT over NRROT, based on *kurtosis*. Comparing the results from all the bandwidth estimators, for Gaussian density NRROT, for super-Gaussian unimodal densities GCAk4ExROT and for sub-Gaussian unimodal densities GCAExROT were the best. This implies that GCAk4ExROT improved over GCAExROT for super-Gaussian but degraded for Gaussian and sub-Gaussian. Though GCAExROT was the best, NRROT was better than all remaining – SROT, GCAk4ExROT and  $\epsilon$ STEPI – bandwidth estimators for sub-Gaussian densities.

Overall, GCAk4ExROT improved over GCAExROT for super-Gaussian and multimodal densities but at the cost of some loss of performance for Gaussian and sub-Gaussian densities. Though the latter part is theoretically important to understand the new rule in its entirety, empirically it may not be of much significance as its performance on the set of *normal mixture densities* developed as a test-bed by Marron and Wand (Marron and Wand 1992) and accepted by the researchers in the field has been satisfactorily improving over GCAExROT. Also, it is a fact that both  $\epsilon$ STEPI and SROT also gave poor performances than NRROT for these Gaussian and sub-Gaussian densities. Accordingly, the next section fully explores the potential of GCAExROT and GCAk4ExROT both for their performances.

## 6. Performance analysis of ExROT bandwidth selectors

Once verified for their performances at steady state with 50000 samples, the detailed performance analysis and empirical usefulness of the derived GCAExROT and GCAk4ExROT have been targeted in this section. They are compared for their performances with NRROT, SROT and  $\epsilon$ STEPI against varying number of samples, against varying distributions at small sample sets and against real data sets.

### 6.1. Performance against varying number of samples

The first experiment was done to compare performances of the same bandwidth estimators against KDE of varying number of samples. The number of samples were varied from 50 to 50000. Figure 4 shows the MISE plotted against the log of number of samples for all 15 *normal mixture densities* using the bandwidth estimators NRROT (Dotted Lines), GCAExROT

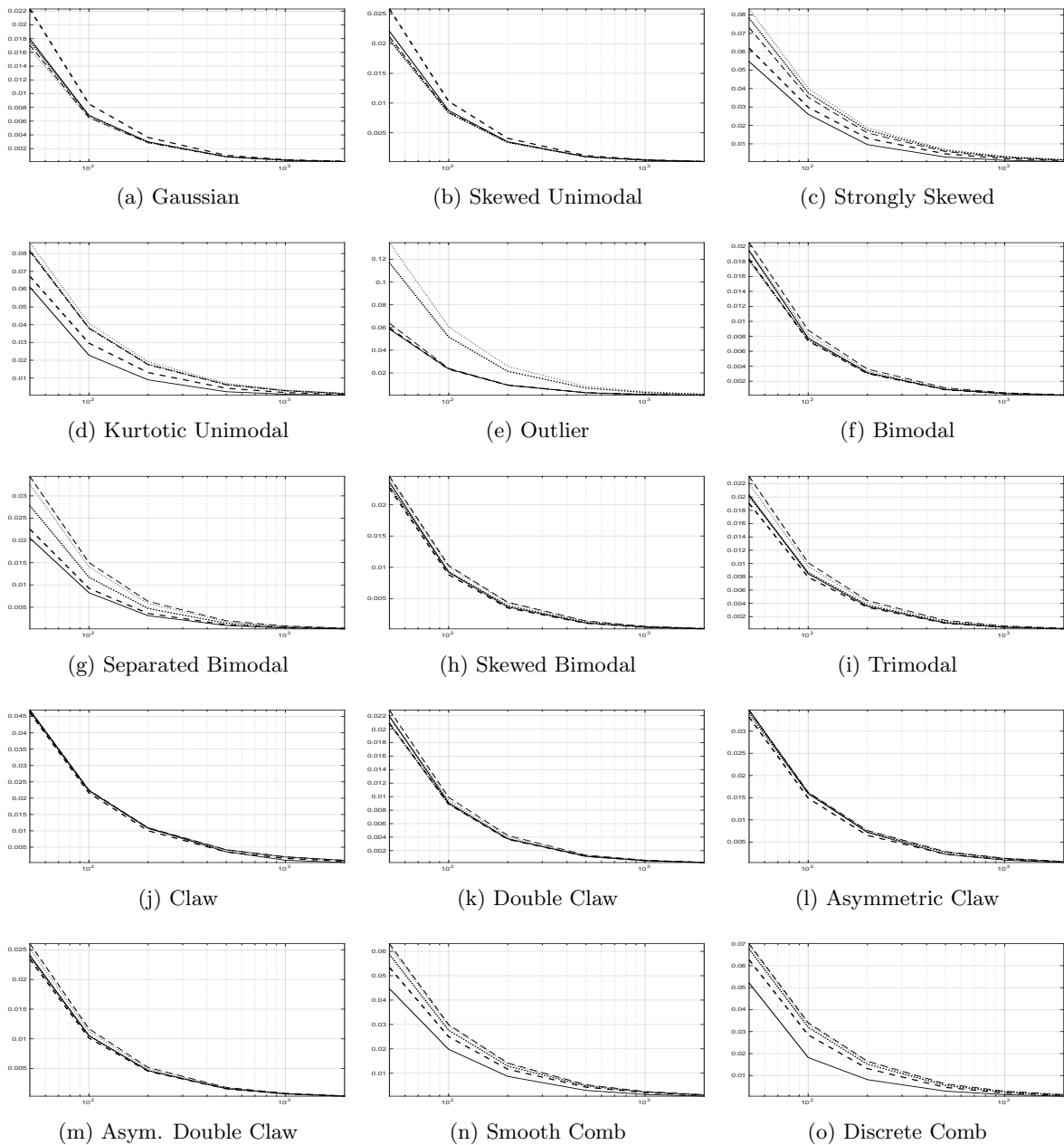


Figure 4: The plots of *mean integrated square error* (MISE) against varying number of samples (from 50 to 2000) using NRROT (Dotted Lines), SROT (Thick Dotted Lines), GCAExROT (Dashed Lines), GCAk4ExROT (Thick Dashed Lines) and  $\epsilon$ STEPI (Solid Lines) bandwidth estimators in Gaussian kernel based estimation of normal mixture density functions in Figure 1. The number of simulation trials for each sample size were 100.

(Dashed Lines), GCAk4ExROT (Thick Dashed Lines) and  $\epsilon$ STEPI (Solid Lines). To achieve visibility, the plots are shown till 2000 samples only. This is also justified as there is almost same linear rate of reduction of MISE with respect to increasing number of samples and there is no much more difference in MISE at large samples for KDE using the given bandwidth estimators. Similar to the results reported at 50000 samples, Figure 4 shows the superiority of NRROT for Gaussian, SROT for skewed Gaussian and  $\epsilon$ STEPI for strongly skewed and kurtotic unimodal densities for any number of samples in terms of achieving minimum mean of the MISE. But un-compare to the results reported at 50000 samples, it shows that at small samples GCAk4ExROT has been the best giving minimum mean of the MISE for outlier density among the unimodal densities and bimodal, skewed bimodal, trimodal, claw, double claw, asymmetrical claw and asymmetrical double claw density types among multimodal densities.

For these densities, at small samples GCAk4ExROT and at large samples  $\epsilon$ STEPI gave the minimum MISE with the transition for the least MISE happening at a few hundreds to a few thousands of samples depending upon the density type. Overall, GCAk4ExROT has been better than both NRROT and SROT for KDE of all except, Gaussian and slightly skewed unimodal (almost Gaussian) density types at any number of samples. It has also been better than  $\epsilon$ STEPI, for KDE of some of the multimodal density types at reduced sample sizes.

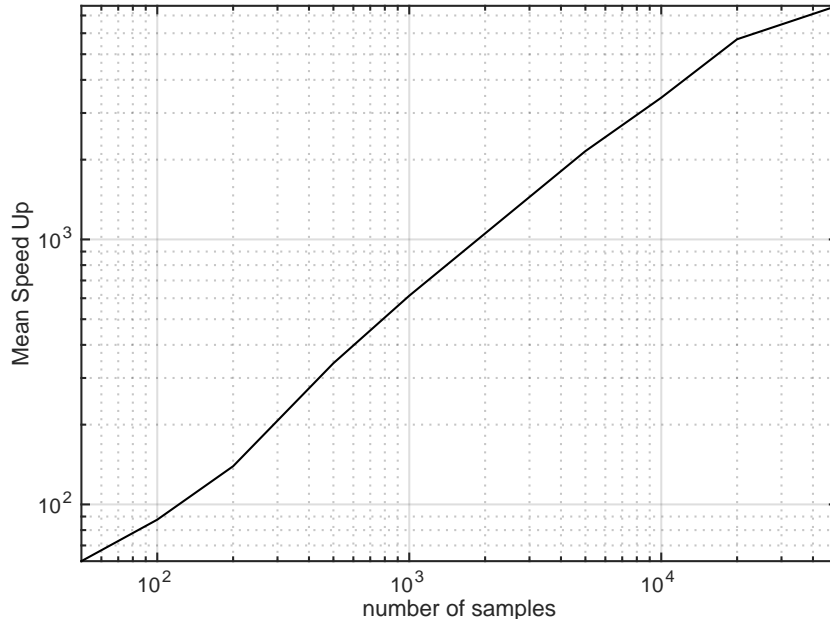


Figure 5: The loglog plot of the Mean Speed-up achieved by GCAk4ExROT over  $\epsilon$ STEPI against increasing number of samples in bandwidth estimation for KDE of a normal mixture density by Marron and Wand (Marron and Wand 1992). There were used Gaussian kernel, Kurtotic unimodal density and 100 simulation trial for each sample set.

Such a performance of GCAk4ExROT is with reduced time complexity compare to that of  $\epsilon$ STEPI. To understand the gain in computational time, the log of the mean *Speed-ups* with respect to the log of number of samples have been plotted in Figure 5. The plot has a slope slightly less than one. This implies that with increasing number of samples the *Speed-up* also increases almost linearly. Overall, the results of this experiment with varying number of samples justify the claim that GCAExROT and GCAk4ExROT provide performance balancing both accuracy and computation, i.e., accuracy better than the ROT and computational cost better than the STEPI bandwidth selectors.

## 6.2. Performance against varying distributions at small sample set

As reported before, the performance of various bandwidth estimators do not vary much at steady state, i.e., large number of samples. Also, nowadays small sample performances of the bandwidth estimators are of much more interest. So, here the bandwidth estimators performances over 100 samples and 50 samples have been reported. There were considered 100 simulation trials for each density estimation from the set of *normal mixture densities*, those shown in Figure 1.

Table 4 reports the results for the bandwidth estimated, mean of the MISE and mean speed-up achieved by GCAExROT over  $\epsilon$ STEPI – all against varying distributions at 100 samples. Results at 50 samples have been reported using box-plots of estimated MISE due to the bandwidth estimators in the test and are shown in Figure 6. The labels on the x-axis are sequentially NRROT, GCAExROT, SROT, GCAk4ExROT and  $\epsilon$ STEPI. The mean of MISE reported in Table 4 and the median of MISE in Figure 6, confirm the previous observations of GCAk4ExROT being better compare to both NRROT and SROT for KDE of distributions

other than Gaussian and (slightly) skewed Gaussian, as well, the same algorithm being better than  $\epsilon$ STEPI for KDE of outlier unimodal density and many other multimodal densities at small samples. It should also be noted that overall  $\epsilon$ STEPI is the best giving minimum mean and median of MISE for many of the densities but it also has large estimation variances compare to all other bandwidth estimator as can be visualized in the box-plots in Figure 6. The MISE was in the range of  $10^{-2}$  for 50 samples and varying from  $10^{-2}$  to  $10^{-3}$  for 100 samples. The mean speed up achieved by GCAExROT over  $\epsilon$ STEPI was also increased from few tens at 50 samples to almost hundred at 100 samples.

Table 4: Comparison of mean integrated square error (MISE) between the estimated and actual densities, where estimated densities were using bandwidth selection rules from *normal reference Rule-of-Thumb* (NR-ROT), Gram-Charlier A series based *extended* ROT (GCAExROT), Silverman’s ROT (SROT), GCAk4ExROT and  $\epsilon$ -exact approximation algorithm based solve-the-equation plug-in ( $\epsilon$ STEPI) bandwidth selectors. The unknown test densities were from the set of *normal mixture densities* by Marron and Wand (Marron and Wand 1992). There were used 100 samples. The table entries show the mean of 100 trials. The boldface values denote the least MISE per row for a corresponding density. The boldface italic values denote the least MISE among ROT and ExROT class of bandwidth estimators.

P D F	Bandwidth (h)*10					MISE*10 <sup>2</sup>					Mean Speed -up *10 <sup>-1</sup>
	NR	GCA	S	GCA	$\epsilon$ STE	NR	GCA	S	GCA	$\epsilon$ STE	
	ROT	Ex ROT	ROT	k4Ex ROT	PI	ROT	Ex ROT	ROT	k4Ex ROT	PI	
a	4.172	4.147	3.545	2.440	4.063	<b>0.646</b>	0.653	0.682	0.849	0.681	4.835
b	3.390	2.896	2.880	1.841	2.872	0.837	0.848	<b>0.833</b>	1.029	0.876	5.238
c	4.371	3.174	3.714	2.196	1.657	4.009	3.523	3.764	<b>2.980</b>	<b>2.606</b>	7.348
d	3.391	2.854	2.882	1.800	1.227	4.076	3.793	3.818	<b>2.961</b>	<b>2.279</b>	8.759
e	1.407	0.536	1.195	0.428	0.520	6.070	2.401	5.165	<b>2.325</b>	2.350	9.892
f	5.084	5.508	4.320	3.277	3.963	0.830	0.882	0.755	<b>0.737</b>	0.777	5.572
g	6.887	7.348	5.852	4.657	3.535	1.405	1.502	1.174	<b>0.924</b>	<b>0.826</b>	6.255
h	4.572	4.722	3.885	2.807	3.445	1.009	1.029	0.933	<b>0.883</b>	0.919	5.697
i	5.387	5.832	4.577	3.524	3.983	0.956	1.012	0.865	<b>0.794</b>	0.844	5.686
j	3.662	3.655	3.112	2.157	3.326	2.244	2.244	2.222	<b>2.160</b>	2.236	5.098
k	5.094	5.516	4.329	3.286	3.807	0.948	0.991	0.891	<b>0.890</b>	0.911	5.620
l	4.766	4.758	4.049	2.853	4.313	1.620	1.621	1.586	<b>1.489</b>	1.603	4.982
m	5.040	5.428	4.283	3.235	3.910	1.127	1.170	1.057	<b>1.013</b>	1.050	5.249
n	6.981	7.360	5.931	4.671	2.685	2.942	2.997	2.762	<b>2.499</b>	<b>1.980</b>	6.575
o	7.067	7.632	6.004	4.696	2.227	3.350	3.402	3.195	<b>2.846</b>	<b>1.817</b>	6.627

### 6.3. Performance against real datasets

The bandwidth estimators were also tested on three different real datasets. The Adult database from the UCI machine learning repository (Lichman 2013) contains 32561 instances with 14 attributes per instance. Among them, five *continuous* attributes – age, final weight (fnlwtg), capital gain (cap-gain), capital loss (cap-loss) and hours per week (hrs/wk) – were used for bandwidth and the resultant density estimations. The second dataset used was an ‘Old faithful’ geyser database from R manual, originally from (Azzalini and Bowman 1990). The database contains 272 instances of durations of eruptions and waiting between two consecutive eruptions for the ‘Old faithful’ geyser in Yellowstone National Park, Wyoming, USA. There are many other versions of the same ‘Old faithful’ geyser data and that used by Silverman (Silverman 1986) with 107 observations has also been used here for testing.

As the true densities are not known, the estimators are compared based on the estimated bandwidth values and the visual observation of the estimated densities. Table 5 shows mutual comparison of the bandwidths estimated by the bandwidth estimators in test, as well, the Speed-up achieved by GCAk4ExROT over  $\epsilon$ STEPI. The table entries exhibit that the estimated bandwidths by GCAExROT are smaller than those by NRROT but larger than those by  $\epsilon$ STEPI. Similarly, the bandwidths by GCAk4ExROT are larger than those by  $\epsilon$ STEPI and smaller than those by rest all, including SROT. Among all the estimated bandwidth values, the bandwidth selected by GCAk4ExROT is the nearest to the visually selected best

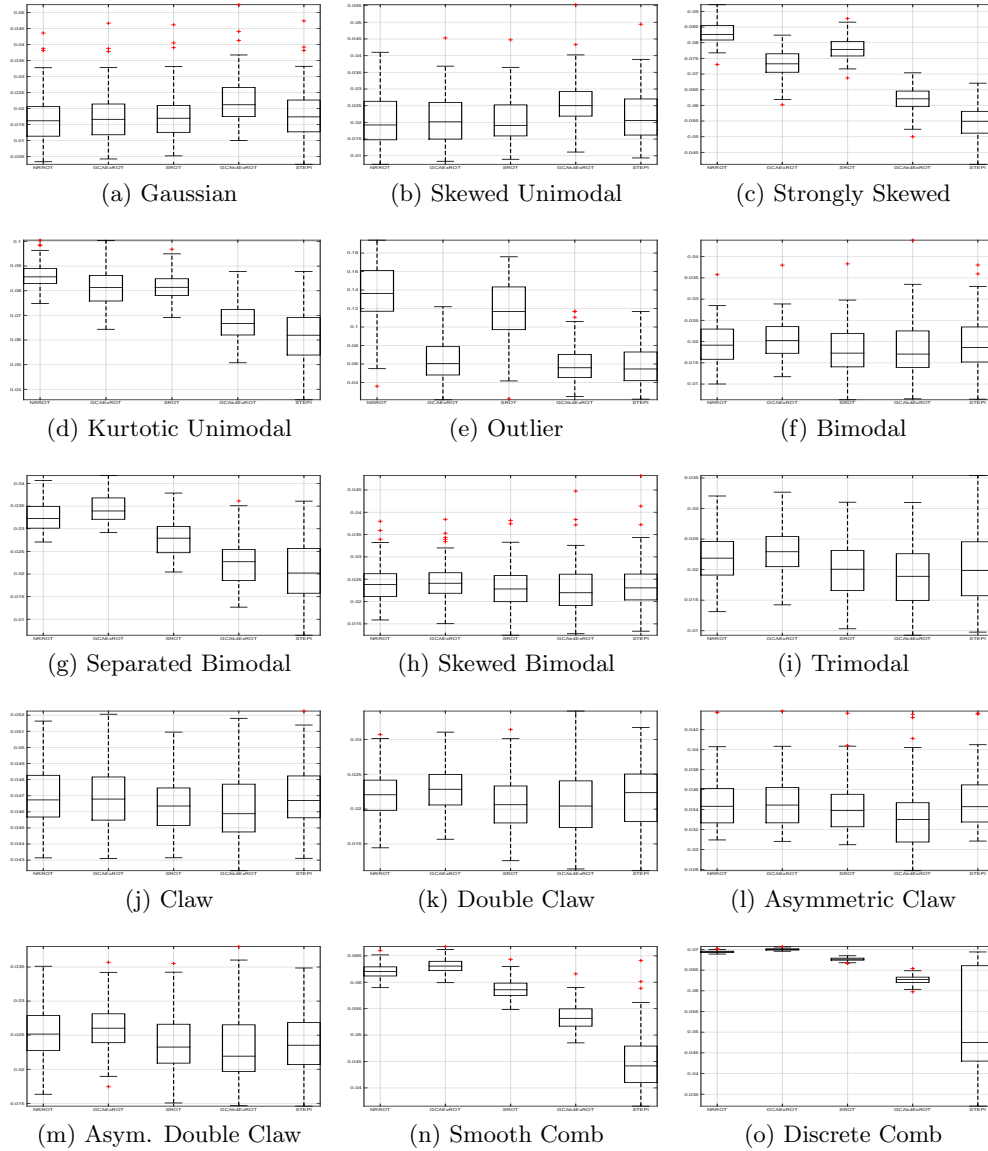


Figure 6: The boxplots of *mean integrated square error* (MISE) using NRROT, GCAExROT, SROT, GCAk4ExROT and  $\epsilon$ STEPI bandwidth estimators for the KDE of normal mixture densities in Figure 1 using sample size of 50 and 100 simulation trials.

Table 5: Performance comparison of various bandwidth selectors for Kernel Density Estimation (KDE) on real data sets. The time calculation is on a machine with features: Intel 2.5 GHz core i5, 4 GB 1.6 GHz DDR3, OS X EI Capitan (version 10.11.6), MATLAB R2010a.

pdf	Bandwidth (h)					Speed -Up
	NR ROT	GCA ExROT	SROT	GCAk4 ExROT	$\epsilon$ STEPI	
<b>Adult database from UCI machine learning repository with 32561 instances</b>						
age	1.808	1.735	1.536	1.059	0.861	72.4
fnlwgt	13992.54	7516.28	11889.432	5631.151	4099.564	5630.7
cap-gain	979.053	157.515	831.899	132.839	2.377	159869.8
cap-loss	53.420	18.607	45.391	15.086	0.123	110507.4
hrs/wk	1.637	1.130	1.391	0.776	0.010	52289.1
<b>'Old faithful' geyser dataset from R manual with 272 observations</b>						
Eruptions	0.394	0.399	0.335	0.267	0.140	26.4
Waitings	4.693	4.851	3.988	3.047	2.497	4.8
<b>'Old faithful' geyser dataset from (Silverman 1986) with 107 observations</b>						
eruptions	0.433	0.427	0.368	0.277	0.181	131.1



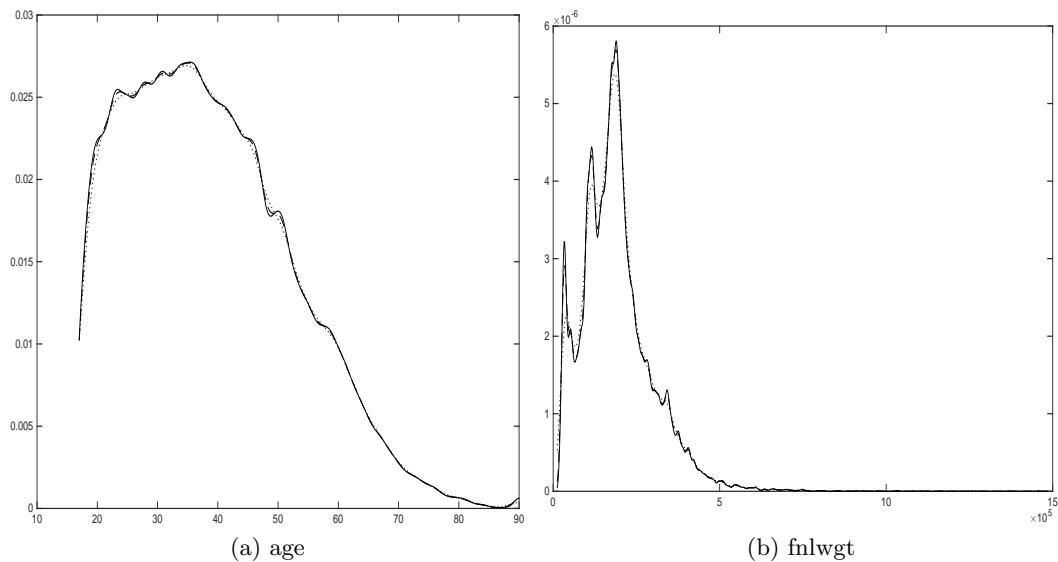


Figure 7: KDE using Gaussian kernel and bandwidths from SROT (Dotted Line), GCAk4ExROT (Dashed Line) and  $\epsilon$ STEPI (Solid Line) for density estimate of two variables (a) age and (b) final weight (fnlwgt) in Adult database from UCI machine learning repository with 32561 instances.

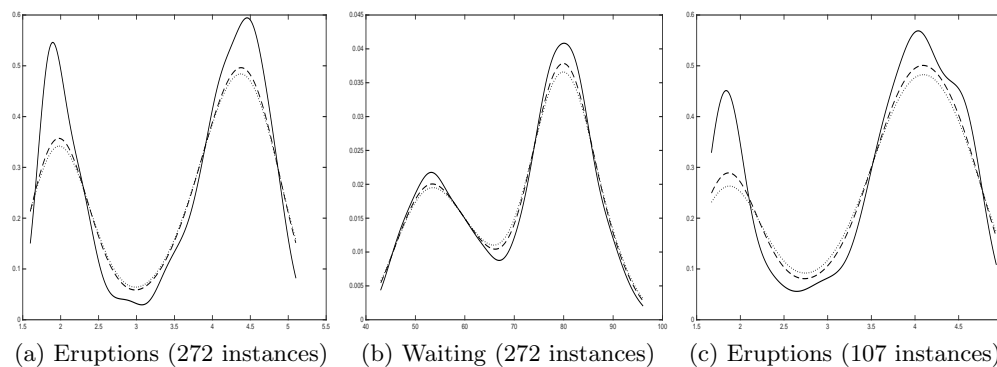


Figure 8: KDE using Gaussian kernel and bandwidths from SROT (Dotted Line), GCAk4ExROT (Dashed Line) and  $\epsilon$ STEPI (Solid Line) for (a) eruptions and (b) waiting in ‘Old Faithful’ geyser database with 272 instances from R manual and originally from (Azzalini and Bowman 1990). Similarly, plot (c) is for eruptions in ‘Old Faithful’ geyser database with 107 observations from (Silverman 1986).

value of  $h = 0.25$  by Silverman (Silverman 1986) for the ‘Old faithful’ geyser dataset with 107 observations. Interested users can verify estimated bandwidth values from many other bandwidth estimators for the same database in (Loader 1999, Section 4). The estimated densities through bandwidths from SROT, GCAk4ExROT and  $\epsilon$ STEPI for some of the variables from Adult database are plotted in Figure 7 and those from ‘Old faithful’ geyser database are plotted in Figure 8. It can be observed that the densities estimated using  $\epsilon$ STEPI as the bandwidth selection rule are quite noisy (‘wiggly’), those estimated using NRROT as the bandwidth selection rule are too smooth and compare to them, the densities estimated through GCAk4ExROT are intermediate, i.e., quite smooth and at the same time without losing any major bumps.

## 7. Conclusion and discussion

Towards bandwidth selection rules for one dimensional KDE, NRROT is the most popular and widely used. It is simple being ‘*quick and dirty*’ and assumes the unknown density to be estimated as Gaussian. The current article extends this assumption to near-Gaussianity and derives GCA series based ExROT, identified as GCAExROT, in Equation (20). The exper-

imental verification of the derived GCAExROT exhibited better performance over NRROT for unimodal densities but poor performance at multimodal distributions. The analysis found that the estimation of the cumulants over multiple modes causes overestimated variance and underestimated *kurtosis* locally for a given mode. Based on the historical evidence of NRROT performance improvement by Silverman deriving SROT, GCAExROT performance has also been improved deriving GCAk4ExROT in (23) by adapting suitable correction mechanisms. Theoretically the extended assumption and empirically the testing on simulated and real datasets proved GCAExROT and the modified rule GCAk4ExROT as ‘*quick and not so dirty*’ bandwidth estimators.

Both the bandwidth selectors derived in this article are simple, quick and with rate of convergence of estimation error against number of samples similar to those of NRROT and SROT. GCAExROT gave better accuracy than both NRROT and SROT for KDE of unimodal densities, except Gaussian, but gave poor accuracy for KDE of multimodal densities. GCAk4ExROT gave better accuracy than both NRROT and SROT for KDE of all, except Gaussian and slightly skewed unimodal, at any given number of samples. At large samples, GCAk4ExROT achieves substantial speed-up compare to the  $\epsilon$ STEPI, and at small samples it achieves accuracy comparable to that of  $\epsilon$ STEPI for some of the multimodal distribution. As such,  $\epsilon$ STEPI gave the smallest bandwidth parameters, resulting into the smallest estimation bias but gave relatively large estimation variance and noisy (‘wiggly’) density estimates. Compare to it, GCAk4ExROT gave small estimation variance and less noisy (‘wiggly’) or more smoother density estimates, but more estimation bias compare to  $\epsilon$ STEPI. Thus, GCAk4ExROT has performance intermediate to those of ROT class and STEPI class bandwidth estimators balancing accuracy and computational cost. So, the article recommends GCAk4ExROT as a replacement for NRROT and SROT. With computations no-longer remaining a major problem nowadays,  $\epsilon$ STEPI may not required to be replaced whenever a precision in estimation is required and sufficient number of samples are available. Still one may select GCAk4ExROT over  $\epsilon$ STEPI in case of small samples with slight loss of accuracy against the gain in computational time, the smoothness in estimation and the reduction in estimation variance. GCAExROT has an performance intermediate to those of ROT class and STEPI class bandwidth estimators for unimodal densities. In case, a prior is available that the density to be estimated is unimodal, GCAExROT instead of GCAk4ExROT can be used to replace NRROT.

The improvement needed for GCAExROT may raise a query whether we can consider the concept of applying *extended* assumption of near-Gaussianity for an unknown density a success or not. It is to be reminded that GCAExROT gave better accuracy for GGD( $\beta$ ) class of unimodal symmetrical densities with varying *kurtosis*. It has also shown success against skewed and other unimodal *normal mixture densities*. This implies that the extended assumption worked very well for KDE of the unimodal densities, which fit better the near-Gaussian assumption. But it is known that the real world densities may not be unimodal and thus may not exactly fit the near-Gaussianity assumption. The modifications needed was to make the near-Gaussian assumption valid in most of the real scenario and that is what it achieved. Thus, the extended assumption is really a success in bringing a class of rules balancing accuracy and computational cost.

The results specifically on real datasets raise a query: why the  $\epsilon$ STEPI rule, following a more complex mechanism to estimate bandwidth, has resulted into such noisy estimates? The answer from (Loader 1999) is that the bandwidth estimators estimate optimal  $h$  that minimizes the AMISE and it is not optimized to achieve the smoothness. AMISE itself is inadequate to address the smoothness concern simultaneously. In spite of this justification, visual observation of smoothness can not be denounced.

The topic of data driven bandwidth estimator for KDE is well established. Till now, the research had been following two major trends. The first trend intends to give simple, ‘*quick and dirty*’ estimate of an unknown density. The intention is achieved by enforcing an assumption of a specific density on the unknown density to be estimated. The second trend intends to

achieve more accurate estimate of an unknown density. The methods following this trend employ an iterative technique to solve a nonlinear equation relating density functionals and bandwidth, and improve the bandwidth parameter estimate iteratively. The methods apply assumption of a specific density on the unknown density to be estimated only during a *pilot* stage. As a whole, they become computationally expensive. The significance of this article can be derived from the fact that it successfully initiates a third trend. The third trend that balances both the previous trends in terms of the approach or the basic assumptions, which in turn also translates into balancing the accuracy and the computations. Instead of either an assumption of a specific density or effectively no such assumption through the multi-stage estimation approach, it assumes the unknown density to be estimated as the member of a generalized group of densities such that it can be expressed as an infinite series expansion through statistics of a known reference density. As these rules not just depend upon the *scale* parameter but also on the higher order cumulants of an unknown density, the class of these rules can also be identified as the ‘*Cumulant Measures*’. With the better results due to wider assumptions and time taken almost of the same order, the rules through the third approach, the ‘*quick and not so dirty*’ rules, can replace the ‘*quick and dirty*’ rules.

By successfully deriving GCAExROT and modifications, this article also serves as a particular demonstration of a new class of bandwidth selection rules based on PDF approximations through infinite series. The PDF approximation through infinite series expansion is a well established area and there exist many such approximations based on various reference PDFs. To give some examples, there are Gram-Charlier series expansions based on the Poisson distribution (Aroian 1937), log-normal distribution (Aroian 1947; Schleher 1977), binomial distribution (Rietz 1927), gamma distribution (Bowers 1966), Wishart distribution (Kollo and von Rosen 1995) and others as reference PDFs. As the results achieved in this article are encouraging, many such rules can be developed and evaluated. ExROT based on *pdf* approximation through infinite series expansion other than GCA series, as well, multivariate generalization of the derived GCAExROT for multivariate KDE are the possible future work.

## Appendix

### AMISE for bandwidth parameter selection in KDE

Given  $N$  realizations of an unknown PDF  $f(x)$ , the kernel density estimate  $\hat{f}(x)$  is given by

$$\hat{f}(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right) = \frac{1}{N} \sum_{i=1}^N K_h(x - x_i). \quad (24)$$

where,  $K(u)$  is the kernel function and  $h$  is the bandwidth parameter. Usually,  $K(u)$  is a symmetric, positive definite and bounded function; mostly a PDF; satisfying the following properties:

$$K(u) \geq 0, \quad \int_{\mathbb{R}^d} K(u) du = 1, \quad \int_{\mathbb{R}^d} u K(u) du = 0, \quad \int_{\mathbb{R}^d} u^2 K(u) du = \mu_2(K) < \infty.$$

The accuracy of a PDF estimation can be quantified by the available distance measures between PDFs, like  $L_1$  norm or mean integrated absolute measure,  $L_2$  norm or mean integrated square error (MISE), Kullback-Leibler divergence and others. The optimal smoothing parameter (the bandwidth)  $h$  is obtained by minimizing the selected distance measure. The bandwidth selection rule based on the most widely used criteria MISE or ISE (Integrated

Square Error) is derived as under (Silverman 1986; Wand and Jones 1994).

$$\begin{aligned}
\text{ISE}(f(x), \hat{f}(x)) &= L_2(f(x), \hat{f}(x)) := \int_{\mathbb{R}^d} (\hat{f}(x) - f(x))^2 dx. \\
\text{MISE}(f(x), \hat{f}(x)) &= E\{\text{ISE}(f(x), \hat{f}(x))\} = E\left\{\int_{\mathbb{R}^d} (\hat{f}(x) - f(x))^2 dx\right\} \\
&= \int_{\mathbb{R}^d} E\{(\hat{f}(x) - f(x))^2\} dx = \int_{\mathbb{R}^d} \text{MSE}(f(x), \hat{f}(x)) = \text{ISE}(f(x), \hat{f}(x)) \\
&= \int_{\mathbb{R}^d} (E\{\hat{f}(x)\} - f(x))^2 + E\{(\hat{f}(x) - E\{\hat{f}(x)\})^2\} dx \\
&= \int_{\mathbb{R}^d} \text{Bias}^2(\hat{f}(x)) dx + \int_{\mathbb{R}^d} \text{Var}(\hat{f}(x)) dx. \tag{25}
\end{aligned}$$

$$\begin{aligned}
\text{Now, } E\{\hat{f}(x)\} &= \frac{1}{Nh} \sum_{i=1}^N E\left\{K\left(\frac{x - x_i}{h}\right)\right\} \\
&= \frac{1}{h} \int_{\mathbb{R}^d} K\left(\frac{x - s}{h}\right) f(s) ds \\
&= \int_{\mathbb{R}^d} K(z) f(x - hz) dz \quad (\because \text{substituting } z = \frac{x-s}{h}).
\end{aligned}$$

Expanding  $f(x - hz)$  using Taylor Series with the assumption of  $f(x)$  being a smooth PDF, i.e., all derivatives exist, we get:

$$\begin{aligned}
E\{\hat{f}(x)\} &= \int_{\mathbb{R}^d} K(z) \left( f(x) - hz f'(x) + \frac{h^2 z^2}{2} f''(x) + O(h^2) \right) dz \\
&= f(x) + \frac{h^2}{2} \mu_2(K) f''(x) + O(h^2). \\
\Rightarrow \text{Bias}(\hat{f}(x)) &\approx \frac{h^2}{2} \mu_2(K) f''(x), \tag{26}
\end{aligned}$$

where  $\mu_2(K) = \int z^2 K^2(z) dz$ , the second order central moment of the kernel.

$$\begin{aligned}
\text{Similarly, } \text{Var}(\hat{f}(x)) &= \text{Var}\left(\frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right)\right) \\
&= \frac{1}{N^2 h^2} \sum_{i=1}^N \text{Var}\left(K\left(\frac{x - x_i}{h}\right)\right) \\
&= \frac{1}{N h^2} \left[ E\left\{K^2\left(\frac{x - x_i}{h}\right)\right\} - E^2\left\{K\left(\frac{x - x_i}{h}\right)\right\} \right] \\
&= \frac{1}{N} \int_{\mathbb{R}^d} \frac{1}{h^2} \left( K^2\left(\frac{x - s}{h}\right) \right) f(s) ds - \frac{1}{N} \left( \frac{1}{h} \int_{\mathbb{R}^d} K\left(\frac{x - s}{h}\right) f(s) ds \right)^2 \\
&= \frac{1}{N h} \int_{\mathbb{R}^d} K^2(z) f(x - hz) dz - \frac{1}{N} \left( f(x) + \text{Bias}(\hat{f}(x)) \right)^2. \tag{27}
\end{aligned}$$

where,  $s$  is the mean of  $x$  and  $z = \frac{x-s}{h}$ . Now, expanding  $f(x - hz)$  using Taylor Series with assumption of  $f(x)$  being a smooth PDF i.e. all derivatives exist

$$\begin{aligned}
\text{Var}(\hat{f}(x)) &= \frac{1}{N h} \int_{\mathbb{R}^d} K^2(z) \left( f(x) - hz f'(x) + \frac{h^2 z^2}{2} f''(x) + O(h^2) \right) dz \\
&\quad - \frac{1}{N} \left( f(x) + \frac{h^2 z^2}{2} f''(x) + O(h^2) \right)^2 \\
\Rightarrow \text{Var}(\hat{f}(x)) &\approx \frac{1}{N h} f(x) \int_{\mathbb{R}^d} K^2(z) dz \quad (\because \text{assuming large } N, \text{ small } h). \tag{28}
\end{aligned}$$

Combining equations (25), (26) and (28)

$$\text{MISE}(\hat{f}(x)) = \frac{h^4}{4}(\mu_2(K))^2 R(f'') + \frac{1}{Nh} R(K) + O(h^4) + O\left(\frac{h}{N}\right),$$

where  $R(f'') = \int (f''(x))^2 dx$  and  $R(K) = \int K^2(z) dz$ . In general,  $R(g) = \int g^2(z) dz$  is identified as the roughness of function  $g(x)$ . An asymptotic large sample approximation AMISE is obtained, assuming  $\lim_{N \rightarrow \infty} h = 0$  and  $\lim_{N \rightarrow \infty} Nh = \infty$  i.e.  $h$  reduces to 0 at a rate slower than  $1/N$ .

$$\text{AMISE}(\hat{f}(x)) = \frac{h^4}{4}(\mu_2(K))^2 R(f'') + \frac{1}{Nh} R(K). \quad (29)$$

The Equation (29) interprets that a small  $h$  increases estimation variance, whereas, a larger  $h$  increases estimation bias. An optimal  $h$  minimizes the total  $\text{AMISE}(\hat{f}(x))$ . So,

$$\begin{aligned} \frac{d}{dh} \text{AMISE}(\hat{f}(x)) &= h^3(\mu_2(K))^2 R(f'') - \frac{1}{Nh^2} R(K) = 0 \\ \Rightarrow h_{\text{AMISE}} &= \left( \frac{R(k)}{\mu_2(K)^2 R(f'') N} \right)^{\frac{1}{5}}. \end{aligned} \quad (30)$$

Thus, the optimal bandwidth parameter depends upon some of the kernel parameters, number of samples and the second derivative of the actual PDF being estimated.

## References

- Aroian LA (1937). "The Type B Gram-Charlier Series." *The Annals of Mathematical Statistics*, **8**(4), 183–192.
- Aroian LA (1947). "The Probability Function of the Product of Two Normally Distributed Variables." *The Annals of Mathematical Statistics*, **18**(2), 265–271.
- Azzalini A, Bowman AW (1990). "A Look at Some Data on the Old Faithful Geyser." *Applied Statistics*, pp. 357–365.
- Berberan-Santos M (2007). "Expressing a Probability Density Function in Terms of another PDF: A Generalized Gram-Charlier Expansion." *Journal of Mathematical Chemistry*, **42**(3), 585–594.
- Bowers NL (1966). "Expansion of Probability Density Functions as a Sum of Gamma Densities with Applications in Risk Theory." *Trans. Soc. Actuaries*, **18**(52), 125–147.
- Cheng T, Gao J, Zhang X (2019). "Nonparametric Localized Bandwidth Selection for Kernel Density Estimation." *Econometric Reviews*, **38**(7), 733–762.
- Cohen L (1998). "Generalization of the Gram-Charlier/Edgeworth Series and Application to Time-Frequency Analysis." *Multidimensional Systems and Signal Processing*, **9**(4), 363–372.
- Cohen L (2011). "On the Generalization of the Edgeworth/Gram-Charlier Series." *Journal of Mathematical Chemistry*, **49**(3), 625–628.
- Ćwik J, Koronacki J (1997). "Multivariate density estimation: A comparative study." *Neural Computing & Applications*, **6**(3), 173–185.
- Epanechnikov VA (1969). "Non Parametric Estimation of a Multivariate Probability Density." *Theory of Probability and Its Applications*, No. 14., pp. 153–158.
- Hald A (2000). "The Early History of the Cumulants and the Gram-Charlier Series." *International Statistical Review*, **68**(2), 137–153. ISSN 1751-5823.
- Heidenreich NB, Schindler A, Sperlich S (2013). "Bandwidth Selection for Kernel Density Estimation: A Review of Fully Automatic Selectors." *AStA Advances in Statistical Analysis*, **97**(4), 403–433.

- Hwang JN, Lay SR, Lippman A (1994). "Nonparametric Multivariate Density Estimation: A Comparative Study." *IEEE Transactions on Signal Processing*, **42**(10), 2795–2810.
- Izenman AJ (1991). "Review Papers: Recent Developments in Nonparametric Density Estimation." *Journal of the American Statistical Association*, **86**(413), 205–224.
- Janssen P, Marron JS, Veraverbeke N, Sarle W (1995). "Scale Measures for Bandwidth Selection." *Journal of Nonparametric Statistics*, **5**(4), 359–380.
- Jones MC, Marron JS, Sheather SJ (1996). "A Brief Survey of Bandwidth Selection for Density Estimation." *Journal of the American Statistical Association*, *91*, pp. 401–407.
- Kollo T, von Rosen D (1995). "Approximating by the Wishart Distribution." *Annals of the Institute of Statistical Mathematics*, **47**(4), 767–783.
- Kulasekera KB, Padgett WJ (2006). "Bayes Bandwidth Selection in Kernel Density Estimation with Censored Data." *Nonparametric Statistics*, **18**(2), 129–143.
- Lichman M (2013). "UCI Machine Learning Repository." URL <http://archive.ics.uci.edu/ml>.
- Loader CR (1999). "Bandwidth Selection: Classical or Plug-in?" *Annals of Statistics*, pp. 415–438.
- Marron JS (1996). "A Personal View of Smoothing and Statistics." In *Statistical Theory and Computational Aspects of Smoothing*, pp. 1–9. Springer.
- Marron JS, Wand MP (1992). "Exact Mean Integrated Squared Error." *The Annals of Statistics*, Vol. *20*, No. *2*, pp. 712–736.
- Park B, Marron JS (1990). "Comparison of Data Driven Bandwidth Selectors." *Journal of the American Statistical Association*, pp. 66–72.
- Park B, Turlach B (1992). "Practical Performance of Several Data Driven Bandwidth Selectors." *CORE Discussion Papers 1992005*, University catholique de Louvain, Center for Operations Research and Econometrics (CORE).
- Raykar VC, Duraiswami R (2006). "Fast Optimal Bandwidth Selection for Kernel Density Estimation." In J Ghosh, D Lambert, D Skillicorn, J Srivastava (eds.), *Proceedings of the sixth SIAM International Conference on Data Mining*, pp. 524–528.
- Rietz HL (1927). *Mathematical Statistics*, volume 3 of *Carus Mathematical Monographs*. 1 edition. Mathematical Association of America.
- Schleher DC (1977). "Generalized Gram-Charlier Series with Application to the Sum of Log-normal Variates (Corresp)." *IEEE Transactions on Information Theory*, **23**(2), 275–280.
- Scott DW (2012). "Multivariate Density Estimation and Visualization." In *Handbook of Computational Statistics*, pp. 549–569. Springer.
- Scott DW, Terrell GR (1987). "Biased and Unbiased Cross-validation in Density Estimation." *Journal of the American Statistical Association*, **82**(400), 1131–1146.
- Sheather S (1983). "A Data-based Algorithm for Choosing the Window Width when Estimating the Density at a Point." *Computational Statistics and Data Analysis*, *1*, pp. 229–238.
- Sheather S (1986a). "An Improved Data-based Algorithm for Choosing the Window Width when Estimating the Density at a Point." *Computational Statistics and Data Analysis*, *4*, pp. 61–65.
- Sheather S (1986b). "The Performance of Six Popular Bandwidth Selection Methods as Same Real Data Sets (with discussion)." *Computational Statistics*, *7*, pp. 225–250.
- Sheather SJ, Jones MC (1991). "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation." *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 683–690.
- Sheather SJ, et al. (2004). "Density Estimation." *Statistical Science*, **19**(4), 588–597.
- Silverman BW (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

- Simonoff J (1996). *Smoothing Methods in Statistics*. Springer Series in Statistics. Springer.
- Turlach BA, *et al.* (1993). *Bandwidth Selection in Kernel Density Estimation: A Review*. Université catholique de Louvain Louvain-la-Neuve.
- Wand P, Jones C (1994). *Kernel Smoothing*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.
- Woodroffe M (1970). "On Choosing a Delta-sequence." *The Annals of Mathematical Statistics*, **41**(5), 1665–1671.

**Affiliation:**

Bhaveshkumar C. Dharmani  
School of Electronics & Electrical Engineering,  
Lovely Professional University, Jalandhar - G. T. Road,  
Phagwara - 144411, Punjab, INDIA  
E-mail: [dharmanic@gmail.com](mailto:dharmanic@gmail.com)