

A NOTE ON THE CHOICE OF THE SMOOTHING PARAMETER IN THE KERNEL DENSITY ESTIMATE

Daniel F. Froelich

Minnesota State University, Mankato, USA

and

Mezbahur Rahman

Minnesota State University, Mankato, USA

and

BRAC University, Dhaka, Bangladesh

ABSTRACT

Among different density estimation procedures, the kernel density estimation has attracted the most attention. In this paper, the choices for smoothing parameter is discussed when the widely used Gaussian kernel is used in implementing the kernel density estimate. A simulation study is conducted from several mixtures of normal distributions covering a wide range of distributional shapes.

Key words: Bias; Mean Integrated Squared Error; Newton Raphson Method.

1. Introduction

This paper is revisited the exact and approximate mean integrated squares for the Gaussian kernel to estimate the best smoothing parameter. As stated in Silverman (1986), density estimation has become an integral part of non-parametric functional estimation procedures in statistics, where a density function is found from the observed data. In literature, such as, Silverman (1986) and Hart (1997) different methods for density estimation such as histograms, naïve estimator, nearest neighbor method, and orthogonal series estimator are shown, however the one we will take a closer look at is the kernel density estimation method.

The kernel method is based off a kernel, $K(u)$, which is a symmetric and have properties of a density function. This method is a way of using a weighted average based on the importance of the observed data and its closeness to the estimated point to give a better estimate of the observed data's properties, as shown in Hart (1997).

The effectiveness of kernel method can be linked to the choice of smoothing parameter, bandwidth, or window width h depending on the literature. A smoothing parameter can be very large or small

depending on the observed data. A larger smoothing parameter may over smooth the data and cause departure from the true population distribution of the data. On the other hand, a smaller smoothing parameter may capture all the random fluctuations in the data and cause departure from the true population distribution of the data. Hence a compromise is sought. To find the best smoothing parameter we need h that will minimize the true error associated with the data.

According to Silverman (1986), in terms of $\hat{f}(x)$ as an estimator of $f(x)$ the mean integrated square error, MISE, is the most popular method of finding the most accurate estimator, so minimizing the error with respect to h will give the best smoothing parameter. In this paper we will try to not only find the best choice of h from the MISE but also h_a which will be the smoothing parameter estimate by minimizing the approximate MISE, AMISE.

The paper is organized as follows: In section 2 we will look at the kernel density estimation. Here we will explore the properties of kernels and smoothing parameters. In section 3 we will look at a simulation study of MISE and AMISE and the respective smoothing parameters. In section 4 will

look at an application of the methods described in this paper and in section 5 we will make some concluding remarks and will look at any further studies that could be done with this approach.

2. The Choice of h in Kernel Density Estimation

As stated in the introduction the kernel density estimation is a broad field that is extensively studied in statistics. Let X_1, \dots, X_n be a random sample be used to estimate the density function using the kernel density estimation in the form of a continuous density function $\hat{f}(x)$. In this section we will not only discuss the basis of kernel density estimation, the kernel, but we will also discuss the importance of the smoothing parameters to the calculation of the density estimation.

Then kernel estimation method is based off a kernel function, $K(u)$, where $u = \frac{(x-X_i)}{h}$ and the kernel estimate of $f(x)$ can be expressed as the following, as given in Rahman et al. (1996)

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \quad (2.1)$$

According to Hart (1997), a kernel function must satisfy the following conditions which allow for the minimization of the mean integrated squared error.

$$\begin{aligned} \int K(u)du &= 1 \\ \int uK(u)du &= 0 \\ \int u^2K(u)du &= k_2 \neq 0 < \infty \\ \int K^2(u)du &< \infty \end{aligned} \quad (2.2)$$

The first two conditions in (2.2) make it so that the kernel must be a probability density function and have a mean of zero. Satisfying these two properties allows the function to be symmetric and a maximum at zero, which are ideal properties for estimating data that has an unknown distribution. The final two conditions are to allow for the calculation of the MISE and AMISE as we will see in section 3. Silverman (1986) states that we must also have the kernel functions those are continuous and differentiable, so that $\hat{f}(x)$ can also take on these properties. These properties are satisfied by many density functions, so in Table 2.1 we have listed some of the most popular choices as expressed in Silverman (1986).

Table 2.1: Kernel Density Functions

Kernel	Function
Gaussian	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$
Epanechnikov	$K(u) = \frac{3}{4}(1-u^2)I_{(u <1)}$
Uniform	$K(u) = \frac{1}{2}I_{(u <1)}$
Triangular	$K(u) = (1- u)I_{(u <1)}$
Quadratic	$K(u) = \frac{15}{16}(1-u^2)^2I_{(u <1)}$

In this paper we will focus our attention on the commonly used Gaussian kernel. Froelich (2009) showed that Gaussian kernel can be implemented more effectively than the Epanechnikov kernel when data are generated from the mixtures of normal distributions.

If we take the final two properties from (2.2) the Gaussian kernel gives the following constraints needed to minimize MISE and AMISE in solving for h and h_{α} , where h is the smoothing parameter which minimizes the MISE and h_{α} is the smoothing parameter which minimizes the AMISE.

$$\begin{aligned} k_2 &= \int u^2 K(u)du = 1 \\ \int K^2(u)du &= \frac{1}{2\sqrt{\pi}} \end{aligned} \quad (2.3)$$

The effectiveness of the kernel estimation method comes down to the selection of the smoothing parameter. One must be very careful in selecting the ideal h value. If smoothing parameter is chosen that is too large, this can cover up the features of a distribution. However, if a h value is chosen that is too small, this can overemphasize the data variability. Therefore in most cases the more structure a graph has the smaller the h value and the flatter the graph the larger the h value. The best choice of the smoothing parameter hinges on the sample size and the following three factors as described in Hart (1997): The smoothness of the density function; the distribution of the design points; and the amount of variability among the data.

Since the smoothing parameter is a function of n , the sample size will affect the bias of the solution through h . For larger samples, the smoothing parameter is comparatively smaller as the data variation has to be closer to the population variation. That is, for larger samples, the smaller

smoothing parameter will help to reduce the bias and will adjust the weight function to obtain asymptotically unbiased estimates, Silverman (1986). By decreasing h this increases the integrated variance, however increasing h will increase the bias. Therefore in the selection of the smoothing parameter there is balance of systematic and random error that takes place. The way that we will be minimizing the error is by minimizing the exact and approximate mean integrated squared error.

2.1 Using Approximate MISE

As stated in Silverman (1986) mean integrated squared error is the most popularly used measure of accuracy of $\hat{f}(x)$ and can be expressed as

$$MISE(\hat{f}(x)) = \int E \{ \hat{f}(x) - f(x) \}^2 dx \quad (2.5)$$

It has been shown in Parzen (1962) that from the MISE the minimization of the AMISE is defined by

$$\begin{aligned} AMISE(\hat{f}(x)) &= \frac{1}{4} h^4 k_2^2 \int (f^{(2)}(x))^2 dx \\ &+ \frac{1}{nh} \int K^2(u) du \end{aligned} \quad (2.6)$$

From the AMISE it is crucial that we find the optimal value of h so it was also shown in Parzen (1962) that this value is

$$\begin{aligned} h_{\alpha} &= k_2^{-\frac{2}{5}} \left\{ \int K^2(u) du \right\}^{\frac{1}{5}} \\ &* \left\{ \int (f^{(2)}(x))^2 dx \right\}^{-\frac{1}{5}} n^{-\frac{1}{5}} \end{aligned} \quad (2.7)$$

If the density function $f(x)$ of our observed data is known we may solve these approximation expressions by substitution. We have already solved for all terms in (2.6) and (2.7) except $\int (f^{(2)}(x))^2 dx$, which can be obtained by the integration of the second derivative squared of the $f(x)$. In most cases the density function of the observed data will be unknown so $f^{(2)}(x)$ will also be unknown. An estimate is needed for $\hat{f}^{(2)}(x)$ and as Rahman et al. (1996) stated this will be defined as

$$\hat{f}^{(2)}(x) = \frac{1}{nh} \sum_{i=1}^n K^{(2)}\left(\frac{x-x_i}{h}\right) \quad (2.8)$$

Here we will use the Gaussian kernel and a known distribution to find the AMISE and h_{α} . As an example we will show how the AMISE and h_{α} were found when the known distribution was $N(0,1)$. Since the distribution is known we can right the density functions as

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (2.9)$$

From this the second derivative is calculated and defined as

$$f^{(2)}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} (x^2 - 1) \quad (2.10)$$

Then the integration of this expression squared will give us the following

$$\int (f^{(2)}(x))^2 dx = \frac{3}{8\sqrt{\pi}} \quad (2.11)$$

For the Gaussian kernel we can use (2.11) and (2.3) and substitute them into (2.6) and (2.7) we can find AMISE and h_{α} to be

$$AMISE = \frac{3h_{\alpha}^4}{32\sqrt{\pi}} + \frac{1}{2nh_{\alpha}\sqrt{\pi}} \quad (2.12)$$

$$h_{\alpha} = \left(\frac{4}{3n}\right)^{1/5} \quad (2.13)$$

Now we will use the Gaussian kernel with an unknown distribution to find the AMISE and \hat{h}_{α} iteratively. Since the distribution is unknown we

will start with an initial $h = \left(\frac{4}{3n}\right)^{1/5}$, which is the h_{α} of the Gaussian kernel and standard normal density function. As an example we will use a random sample from the standard normal density function. For the Gaussian kernel $\hat{f}^{(2)}(x)$ is calculated by (2.8) as shown in Rahman et al. (1996) to be

$$\begin{aligned} &\int (\hat{f}^{(2)}(x))^2 dx \\ &= \frac{3}{8n^2 h^2 \sqrt{\pi}} \left[h^4 \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{(x_i-x_j)^2}{4h^2}} \right. \\ &\quad - h^2 \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 e^{-\frac{(x_i-x_j)^2}{4h^2}} \\ &\quad \left. + \frac{1}{12} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^4 e^{-\frac{(x_i-x_j)^2}{4h^2}} \right] \end{aligned} \quad (2.16)$$

Once $\hat{f}^{(2)}(x)$ is calculated using the random sample, \hat{h}_α can be calculated by (2.7) and substituting in (2.3) and (2.16). Now that we have \hat{h}_α we will let this be our new h in (2.16) can continue to substitution until \hat{h}_α converges to a solution. Once our approximate smoothing parameter is found we may then calculate the AMISE.

2.2 Using Exact MISE

In section 2.1 we showed how the MISE was defined in (2.5). As done in Rahman et al. (1996) we will let $W_1 = \frac{1}{n} \bar{W} = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$ and along with substitution of the density estimate from equation 2.1 and substituting it into the MISE the exact MISE becomes

$$\begin{aligned} \text{MISE}(\hat{f}(x)) &= \frac{1}{nh^2} \int E(W_1^2) dx + \frac{n-1}{nh^2} \int (E(W_1))^2 dx \\ &\quad - \frac{2}{h} \int E(W_1) f(x) dx + \int (f(x)^2) dx \end{aligned} \quad (2.18)$$

From this expression we can now focus on minimization of MISE. For the substitution method if the distribution of the observed data is known the minimization of the MISE will occur when the first derivative with respect to h is equal to zero and the second derivative with respect to h is positive. Therefore to solve the only term we are missing is h therefore the initial h used will be the h_α from the AMISE. Then we will substitute this h_α into the first and second derivatives. If this was the best h then the first derivative would be equal to zero, however since h_α is just an approximation this is not the case. So by using the following equation similar to Newton Raphson Method the best h was found as

$$h_i = h_{i-1} - \frac{\frac{d\text{MISE}(f(x))}{dh}}{\frac{d^2\text{MISE}(f(x))}{dh^2}} \quad (2.19)$$

As the first derivative limits to zero then h_i limits to the ideal h value for the minimization of the MISE. Finally the MISE can be found by substituting the ideal h back into (2.18).

An iterative method was used to find the ideal smoothing parameter for MISE when the

distribution of the observed data was unknown. Since the distribution is unknown we substitute $\hat{f}(x)$ for $f(x)$ in (2.18) to get the following approximation as in Rahman et al. (1996) to be

$$\begin{aligned} \text{MISE}(\hat{f}(x)) &= \frac{1}{nh^2} \int E(W_1^2) dx + \frac{n-1}{nh^2} \int (E(W_1))^2 dx \\ &\quad - \frac{2}{h} \int E(W_1) \hat{f}(x) dx + \int (\hat{f}(x)^2) dx \end{aligned} \quad (2.20)$$

Then the first derivative of (2.20) gives us the following expression

$$\begin{aligned} \frac{d\text{MISE}(\hat{f}(x))}{dh} &= -\frac{2}{nh^3} \int E(W_1^2) dx \\ &\quad - \frac{2(n-1)}{nh^3} \int (E(W_1))^2 dx \\ &\quad + \frac{2}{h^2} \int E(W_1) \hat{f}(x) dx \end{aligned} \quad (2.21)$$

As done in Rahman et al. (1996), (2.21) can be rearranged to solve for \hat{h} , which can be expressed as

$$\hat{h} = \frac{\int E(W_1^2) dx + (n-1) \int (E(W_1))^2 dx}{n \int E(W_1) \hat{f}(x) dx} \quad (2.22)$$

Solving (2.22) iteratively will give the ideal \hat{h} value that can be substituted in to (2.20) to solve for the minimum MISE.

First we will use the Gaussian kernel and a known distribution to find the MISE and h by the substitution method. As an example we will show how the MISE and h were found when the data's density function was known to be $N(0,1)$. To find the MISE for the Gaussian kernel the method as shown in Marron and Wand (1992) was the following

$$\begin{aligned} \text{MISE}(\hat{f}(x)) &= \frac{C_1(r)}{nh} + \left(1 - \frac{1}{n}\right) \\ &\quad * \sum_{s=0}^{r-1} \sum_{s'=0}^{r-1} \frac{(-1)^{s+s'}}{2^{s+s'} s! s'!} U(h; s + s', 2) \\ &\quad - 2 \left(1 - \frac{1}{n}\right) \sum_{s=0}^{r-1} \frac{(-1)^s}{2^s s!} U(h; s, 1) \\ &\quad + U(h; 0, 0) \end{aligned} \quad (2.22)$$

where

$$C_1(r) = \frac{1}{\sqrt{\pi}} \sum_{s=0}^{r-1} \sum_{s'=0}^{r-1} \frac{(2s+2s')!}{2^{2s+2s'} s! s'! (s+s')!} \quad (2.23)$$

and

$$U(h; s, q) = \sum_{i=1}^k \sum_{i'=1}^k w_i w_{i'} h^{2s} * \left((\sigma_i^2 + \sigma_{i'}^2 + qh^2)^{\frac{s}{2}} \right)^{-(s+1)} * \frac{(\mu_i - \mu_{i'})^2}{2(\sigma_i^2 + \sigma_{i'}^2 + qh^2)^{(s+2)}} \quad (2.24)$$

By using this method we found that the MISE of a known standard normal density, $r = 1$, gave us the following

$$\text{MISE}(h) = \frac{1}{2\sqrt{\pi}} * \left(\frac{1}{nh} + \left(1 - \frac{1}{n}\right) (1 + h^2)^{-\frac{1}{2}} - 2\sqrt{2}(2 + h^2)^{-\frac{1}{2}} + 1 \right) \quad (2.25)$$

From this equation the first and second derivatives were found to be

$$\frac{d\text{MISE}(h)}{dh} = \frac{1}{2\sqrt{\pi}} * \left(-\frac{1}{nh^2} - h \left(1 - \frac{1}{n}\right) (1 + h^2)^{-\frac{3}{2}} + h2\sqrt{2}(2 + h^2)^{-\frac{3}{2}} \right) \quad (2.26)$$

$$\frac{d^2\text{MISE}(h)}{dh^2} = \frac{1}{2\sqrt{\pi}} * \left(\frac{2}{nh^3} + \frac{\left(1 - \frac{1}{n}\right)}{(1+h^2)^{\frac{5}{2}}} \left(\frac{3h^2}{(1+h^2)} - 1 \right) + \frac{2\sqrt{2}}{(2+h^2)^{\frac{5}{2}}} \left(\frac{3h^2}{(2+h^2)} - 1 \right) \right) \quad (2.27)$$

Now these expressions can be used to find the best smoothing parameter using (2.19) and then once \hat{h} is found it can be used to find the minimized MISE.

Now we will use the Gaussian kernel and a random data set with an unknown distribution to find the MISE and \hat{h} . For the Gaussian iterative method we

will start with an initial $h = \left(\frac{4}{3n}\right)^{\frac{1}{5}}$, which is the h_{α} of the Gaussian kernel and standard normal density function as shown in section 2.1. Again as our unknown data set we will use a random sample

from the standard normal density function as an example. To solve for (2.20) and (2.22) the following equations were calculated as in Rahman et al. (1996),

$$E(W_1) = \int w_1 \hat{f}(y) dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-y}{h}\right)^2} * \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^n e^{-\frac{1}{2}\left(\frac{y-x_i}{h}\right)^2} dy = \frac{1}{2n\sqrt{\pi}} \sum_{i=1}^n e^{-\frac{1}{4h^2}(x-x_i)^2} \quad (2.36)$$

$$E(W_1^2) = \int w_1^2 \hat{f}(y) dy = \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\left(\frac{x-y}{h}\right)^2} * \frac{1}{nh\sqrt{2\pi}} \sum_{i=1}^n e^{-\frac{1}{2}\left(\frac{y-x_i}{h}\right)^2} dy = \frac{1}{2\pi n\sqrt{3}} \sum_{i=1}^n e^{-\frac{1}{3h^2}(x-x_i)^2} \quad (2.37)$$

$$\int_{-\infty}^{\infty} E(W_1^2) dx = \frac{h}{2\sqrt{\pi}} \quad (2.38)$$

$$\int_{-\infty}^{\infty} (E(W_1))^2 dx = \frac{h}{n^2\sqrt{8\pi}} \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{1}{8h^2}(x_j-x_i)^2} \quad (2.39)$$

$$\int_{-\infty}^{\infty} E(W_1) \hat{f}(x) dx = \frac{1}{n^2\sqrt{6\pi}} \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{1}{6h^2}(x_j-x_i)^2} \quad (2.40)$$

$$\int_{-\infty}^{\infty} (\hat{f}(x))^2 dx = \frac{1}{n^2 h\sqrt{4\pi}} \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{1}{4h^2}(x_j-x_i)^2} \quad (2.41)$$

Now (2.38) – (2.40) can be substituted in (2.22) and the ideal \hat{h} value can be found iteratively. While the minimized MISE can be found by using this \hat{h} value along with (2.38) – (2.41) in (2.20).

2.3 Mixtures of Normal Family of Distributions

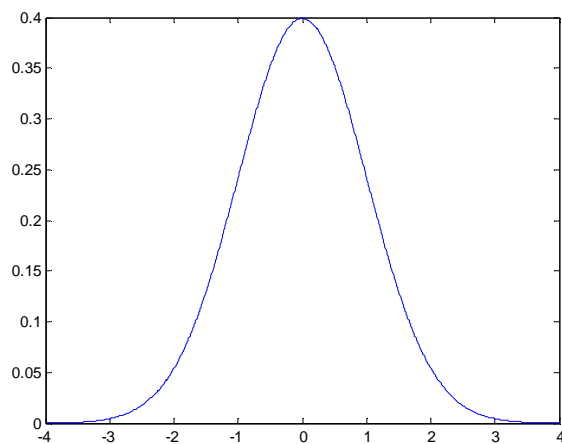
There are many different types of normal family distributions which all have their own properties. We have chosen the following four densities in the following table, which were also used in Rahman et al. (1996).

Table 2.2: Mixtures of Normal Family of Density Functions

Density
#1 – Standard Normal $N(0,1)$
#2 – Skewed Unimodal $\frac{1}{5}N(0,1) + \frac{1}{5}N\left(\frac{1}{2}, \left(\frac{2}{3}\right)^2\right) + \frac{3}{5}N\left(\frac{13}{12}, \left(\frac{5}{9}\right)^2\right)$
#3 – Skewed Bimodal $\frac{3}{4}N(0,1) + \frac{1}{4}N\left(\frac{3}{2}, \left(\frac{1}{3}\right)^2\right)$
#4 – Asymmetric Claw $\frac{1}{2}N(0,1) + \sum_{l=-2}^2 \left(\frac{2^{l-1}}{31}\right) N\left(l + \frac{1}{2}, \left(\frac{2^{-l}}{10}\right)^2\right)$

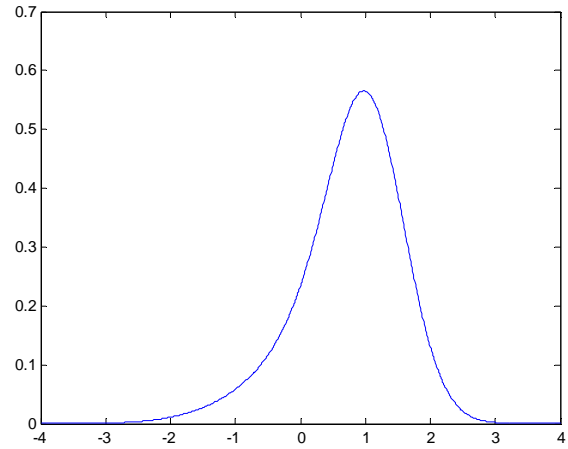
As stated by Marron and Wand (1992) there are many different classes of normal mixtures, which can each create unique problems for curve estimation. Here we only have four distributions, each representing their own class. For more normal mixtures refer to Marron and Wand (1992) for a list of 15 mixtures. The first density is the standard normal distribution that represents a class of unimodal and symmetric densities. A graph of this normal density is shown in the following graph.

Figure 2.3: Standard Normal Density Graph



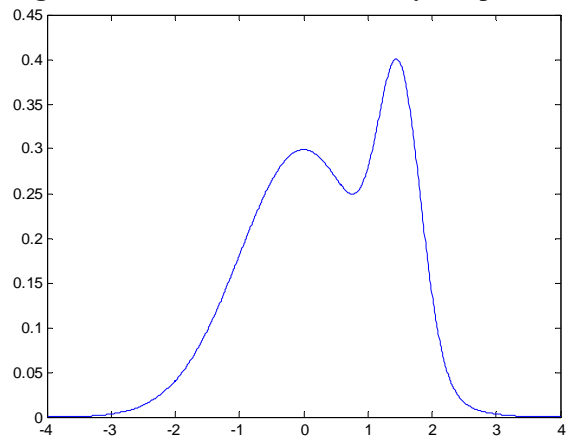
The next density is the skewed unimodal density, which represents the unimodal and asymmetric class of densities and is shown in the following graph.

Figure 2.4: Skewed Unimodal Density Graph



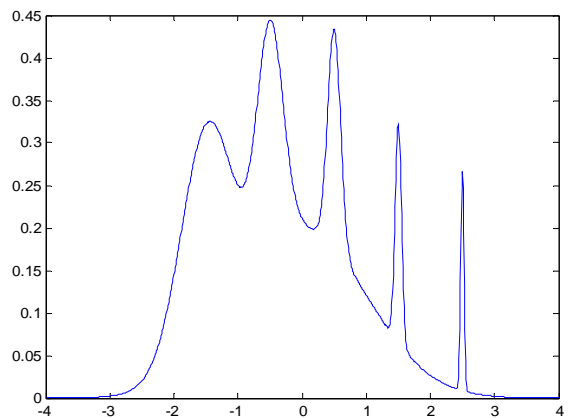
The third distribution is the skewed bimodal density that represents the bimodal density class and is shown in the next graph.

Figure 2.5: Skewed Bimodal Density Graph



The final distribution is the asymmetric claw density is an example from the multimodal and skewed density classes and is shown in the following graph.

Figure 2.6: Asymmetric Claw Density Graph



We will use these four distributions to determine any similarities or differences between the normal families. This variety will give distributions that have different types of structure and parameters, which in turn may help us identify the methods that work best in finding the smoothing parameter and minimizing the MISE for specific families.

3. Simulation Study

In this section we will explore the methods of the minimization of MISE and AMISE, along with finding the ideal smoothing parameter. In the simulation study we explored the substitution method and iterative method for Gaussian kernel. We took sample sizes of 25, 50, 100, 200, and 500 for each of the four normal density distributions in table 2.2.

The following table is the substitution method of the Gaussian kernel as explained in section 2.1 and 2.2.

Table 3.1: Exact Gaussian Kernel by Density Substitution

f	n	h	MISE
#1	25	0.6094	0.01373
	50	0.5199	0.00869
	100	0.4455	0.00541
	200	0.3830	0.00332
	500	0.3150	0.00172
#2	25	0.4251	0.02113
	50	0.3591	0.01336
	100	0.3054	0.00830
	200	0.2611	0.00509
	500	0.2136	0.00262
#3	25	0.5549	0.02219
	50	0.4085	0.01508
	100	0.3179	0.00972
	200	0.2572	0.00606
	500	0.2012	0.00314
#4	25	0.6657	0.03587
	50	0.5231	0.03094
	100	0.2016	0.02289
	200	0.1436	0.01597
	500	0.0929	0.00960

Table 3.2: Approximate Gaussian Kernel by Density Substitution

f	n	h_a	AMISE
#1	25	0.5564	0.02535
	50	0.4844	0.01456
	100	0.4217	0.00836
	200	0.3671	0.00480
	500	0.3056	0.00231
#2	25	0.3730	0.03781
	50	0.3247	0.02172
	100	0.2827	0.01247
	200	0.2461	0.00716
	500	0.2049	0.00344
#3	25	0.3248	0.04343
	50	0.2828	0.02494
	100	0.2462	0.01432
	200	0.2143	0.00822
	500	0.1784	0.00395
#4	25	0.0706	0.19983
	50	0.0614	0.11477
	100	0.0535	0.06592
	200	0.0466	0.03789
	500	0.0388	0.01819

In table 3.1 it is observed that MISE, AMISE, h and h_a all decrease as sample size increases as expected. Also the AMISE is greater than the exact MISE, which means that the h_a values are less accurate than the h values. The MISE increases as the amount of structure increases with normal density function.

Next we will examine the iterative method of the Gaussian kernel as explained in sections 2.1 and 2.2. In table 3.2 will be the data collected for the $MISE_{\hat{h}}$, \hat{h} and their standard deviations from the simulations of the Gaussian kernel with random samples of the four density functions.

Table 3.3: Exact Gaussian Kernel Smoothing Parameter by Density Substitution with Standard Deviation

f	n	\hat{h}	$\sigma_{\hat{h}}$
#1	25	0.524054	0.088838
	50	0.401463	0.049727
	100	0.314427	0.028519
	200	0.250957	0.016713
	500	0.191487	0.008181
#2	25	0.432541	0.073874
	50	0.338654	0.044161
	100	0.229979	0.023589
	200	0.180986	0.014252
	500	0.138397	0.006643
#3	25	0.597667	0.082582
	50	0.443332	0.048126
	100	0.333842	0.031732
	200	0.251828	0.023138
	500	0.178267	0.013316
#4	25	0.592336	0.087852
	50	0.453302	0.051224
	100	0.352296	0.028575
	200	0.272101	0.018726
	500	0.183549	0.014232

In this table we observe that as the sample size increases both $MISE_{\hat{h}}$ and \hat{h} decrease as we expected. We also see that the standard deviation decreases as the sample size increases as expected. Next we have the data collected for the $AMISE_{\hat{h}_a}$, \hat{h}_a and their standard deviations from the simulations of the Gaussian kernel with random samples of the four density functions.

Table 3.3: Exact Gaussian Kernel MISE by Density Substitution with Standard Deviation

f	n	$MISE_{\hat{h}}$	$\sigma_{MISE_{\hat{h}}}$
#1	25	0.015789	0.003540
	50	0.011009	0.001871
	100	0.007448	0.000945
	200	0.004870	0.000459
	500	0.002679	0.000185
#2	25	0.019846	0.004465
	50	0.013439	0.002390
	100	0.010472	0.001447
	200	0.006916	0.000728
	500	0.003771	0.000270
#3	25	0.013935	0.002909
	50	0.010483	0.001863
	100	0.007730	0.001313
	200	0.005447	0.000826
	500	0.003219	0.000349
#4	25	0.013873	0.002551
	50	0.009851	0.001553
	100	0.006873	0.000975
	200	0.005124	0.000702
	500	0.003803	0.000534

As we saw with the $MISE_{\hat{h}}$ and \hat{h} , the $AMISE_{\hat{h}_a}$, \hat{h}_a and their standard deviations all decrease as the sample size increase. The iterative method also gives us smaller $MISE_{\hat{h}}$ compared to the $AMISE_{\hat{h}_a}$ and *therefore* a larger \hat{h} compared to \hat{h}_a as we saw with the substitution method. In the approximate iterative method we have greater standard deviations compared to that of the exact iterative method, which we should expect to happen.

4. An Example

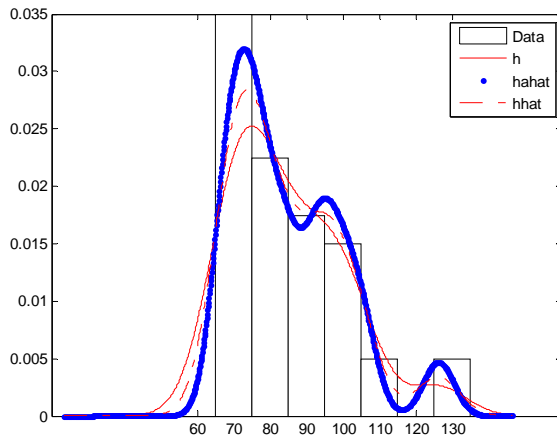
In this chapter we will take a data set and run it through the programs that were made in chapter 3. The data set we will be using is the waist circumference in cm from a set of 40 females from Triola (2008). We then ran the sample through the iterative program, which returned the following results in the following table.

Table 4.1: Table for Application Data Estimations

$K(u)$	Gaussian
\hat{h}	7.800720
\tilde{h}	5.923840
$MISE_{\hat{h}}$	0.001041
\tilde{h}_{α}	4.239083
$AMISE_{\tilde{h}_{\alpha}}$	0.002080

With this table we can see that our exact MISE was less than our AMISE, which is what we wanted. Also each kernel gave us similar answers. Once we have our h values for each kernel we can then find the density function for each smoothing parameter of the observed data. To do this we took 1000 evenly distributed points within 4 standard deviations of the mean and used (2.1), $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$. Once the density function was found it returned the following graphs of the density estimations of the Gaussian kernel when graphed with the data.

Figure 4.1: Gaussian Kernel Density Functions



Even through Table 4.1 depicts that it does not matter which kernel we chose the figures paint a different picture. From the figures it seems as if the Gaussian kernel does a much better job in displaying the observed data. Within the Gaussian figure we can see that the three density functions each do a nice job giving the entire structure of the data. Figure 4.1 is a wonderful illustration of the difference the smoothing parameter makes. As discussed in section 2 the smaller the h value the variability it shows. In figure 4.1 the smallest

smoothing parameter is the \tilde{h}_{α} value, which has the steepest slopes and brings out all the peaks and valleys of the observed data. However the h and \hat{h} values follow through the data with a much smoother and less varying curve. Being a small set of observed data the best choice would be the \hat{h} value, because this also gave us the smaller integrated error. With figure 4.2 the \hat{h} value should be the choice of smoothing parameter due to the fact that it gave the smallest integrated square error and the other two densities have a much smaller smoothing parameter, which is bringing out the data variability in this small data set. This example also shows us that the approximate values are a great place to start, but are not necessarily the best choice.

5. Concluding Remarks and Future Study

Since density estimation is a very important field in statistics there is always new ways to improve the techniques of density estimation. Since there are many kernels and we have only looked at one, we have only scratched the surface. These techniques could be used to analyze all of the other kernels as well. Another option is that with these two kernels different normal family density functions could be analyzed as done in Marron and Wand (1992).

BIBLIOGRAPHY

Silverman, B. W. (1986). “Density Estimation for Statistics and Data Analysis”, Chapman and Hall Ltd., New York.

Hart, J. D. (1997). “Nonparametric Smoothing and Lack-of-Fit Tests”, Springer-Verlag New York, Inc., New York.

Rahman, M., Arnold, B. C., Gokhale, D. V., and Ullah, A. (1996). “Data-based Selection of the Smoothing Parameter in Kernel Density Estimation using Exact and Approximate MISE”, Technical Report #229, Department of Statistics, University of California, Riverside.

Froelich, D. F. (2009). “Mean Integrated Squared Error in the Kernel Density Function Estimation”, Alternative Plan Paper, Minnesota State University, Mankato.

Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode". *Annals of Mathematical Statistics*, 33, 1065-1076.

Marron, J. S. and Wand, M. P. (1992). "Exact Mean Integrated Squared Error", *The Annals*

of Statistics, 20(2), 712-736.

Triola, M. F. (2008). "Essentials of Statistics", Third Edition, Pearson Education, Inc., Boston, MA.