Open access • Journal Article • DOI:10.1037/A0036850

How Bandwidth Selection Algorithms Impact Exploratory Data Analysis Using Kernel Density Estimation — Source link

Jared K. Harpole, Carol M. Woods, Thomas L. Rodebaugh, Cheri A. Levinson ...+1 more authors

Institutions: University of Kansas, Washington University in St. Louis

Published on: 02 Jun 2014 - Psychological Methods (University of Kansas)

Topics: Kernel density estimation, Exploratory data analysis, Estimator, Smoothing and Sample size determination

How Bandwidth Selection Algorithms Impact Exploratory Data Analysis Using Kernel Density Estimation

By

Jared K. Harpole

Submitted to the Department of Psychology and the
Graduate Faculty of the University of Kansas
in partial fulfillment of the requirements for the degree of
Master's of Arts

Carol Woods, Chairperson

Committee members

Pascal DeBoeck

Paul Johnson

Date defended: 02/25/2013

The Thesis Committee for Jared K. Harpole certifies
that this is the approved version of the following thesis :

How Bandwidth Selection Algorithms Impact Exploratory Data Analysis Using Kernel Density Estimation

Carol Woods, Chairperson

Date approved: 02/25/2013

Abstract

Exploratory data analysis (EDA) is important, yet often overlooked in the social and behavioral sciences. Graphical analysis of one's data is central to EDA. A viable method of estimating and graphing the underlying density in EDA is kernel density estimation (KDE). A problem with using KDE involves correctly specifying the bandwidth to portray an accurate representation of the density. The purpose of the present study is to empirically evaluate how the choice of bandwidth in KDE influences recovery of the true density. Simulations were carried out that compared five bandwidth selection methods [Sheather-Jones plug-in (SJDP), Normal rule of thumb (NROT), Silverman's rule of thumb (SROT), Least squares cross-validation (LSCV), and Biased cross-validation (BCV)], using four true density shapes (Standard Normal, Positively Skewed, Bimodal, and Skewed Bimodal), and eight sample sizes (25, 50, 75, 100, 250, 500, 1000, 2000). Results indicated that overall SJDP performed best. However, this was specifically true for samples between 250 and 2,000. For smaller samples (N = 25 to 100), SROT performed best. Thus, either the SJDP or SROT is recommended depending on the sample size.

Acknowledgements

I would like to thank my chairperson Dr. Carol Woods for providing me with the initial motivation and guidance throughout the process. I would also like to express my thanks to my co-chairs Dr. Paul Johnson and Dr. Pascal DeBoeck for assistance. Finally, I wish to acknowledge my wife Rebecca. Without you none of this would be possible nor would any of it be worth it.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Background

1.1 Introduction

Exploratory data analysis (EDA) is important, yet often overlooked in the social and behavioral sciences. The quality of the statistical conclusions depends on the accuracy of the data used in the analyses; in other words, "garbage in garbage out" (Kline, 2008). EDA assists with hypothesis testing by revealing unexpected or misleading patterns in the data (Behrens, 1997). Unfortunately, many psychologists do not utilize EDA in their research.

Central to EDA is graphics, which researchers use to diagnose potential issues in the data and observe trends that can be hidden by summary statistics. When scientists want to analyze and graph the underlying distribution of data, either parametric or non-parametric methods can be employed. A parametric approach assumes that the random sample comes from a known family of distributions such as the Normal, whereas, a nonparametric approach makes no assumption about the distribution underlying the random sample. In practice, a nonparametric approach is preferable because the underlying density is unknown (Mugdadi & Jeter, 2010).

The simplest and most well known nonparametric density estimation technique is the histogram. Histograms are commonly used to illustrate underlying densities due to their ease of implementation and prominence. However, according to Silverman (1986), histograms suffer from

two major problems. First, histograms do not produce a smooth curve of the underlying data. This can cause problems with visual representation of the underlying density as well as computational issues for advanced analyses. Second, data within histograms depend on the endpoints, which can cause a loss of information. For example, it is possible that a point within a specified bin could be closer to points in the adjacent bin versus the bin of origin. An approach that alleviates these two issues is kernel density estimation (Silverman, 1986; Wand & Jones, 1995).

Kernel density estimation (KDE) uses local averaging to create a smooth curve from a sample of observations. The averaging occurs on the center of each point x within a specified neighborhood of points close to x. The closer the points are to x the more weight they are assigned and the higher the density at x (Wilcox, 2001). Although KDE estimates are not frequently reported in social science research, they have great utility as a tool for identifying outliers and unexpected patterns in the data not apparent from summary statistics (Behrens, 1997; Marmolejo-Ramos & Matsunaga, 2009). For example, Wilcox (2004; 2006) showed how KDE could be used to graphically depict effect sizes between groups and how using KDE can assist in finding group differences hidden by measures of central tendency. Akiskal and Benazzi (2006) used KDE in conjunction with other measures to corroborate their conclusions that major depressive disorder and bipolar II disorder are not distinct but lie on a continuum. Additionally, Osberg and Smeeding (2006) mentioned that the value of KDE lies in presenting a picture that conveys more information than summary statistics.

There are two issues, however, that practitioners must be aware of when utilizing KDE. First, it is necessary to specify a kernel function to estimate the density. There are several common types of kernel estimators used in practice such as Normal, Epanechnikov, biweight, triweight, triangular, and uniform (Scott, 1992). Table 1.1 lists the equations for each kernel. There is no consensus about what type of kernel is best, but several authors note that the choice of the kernel is not particularly important because there is a trivial loss of efficiency in picking one kernel over the other (Scott, 1992; Silverman, 1986; Wand & Jones, 1995). In the present study the Normal kernel will be used to aid in tractability of the calculations of the discrepancy measure.

The second issue involves selecting the bandwidth or smoothing parameter. There are two

Table 1.1: Some Kernels and their Equations

Kernel	Normal	Uniform	Epanechnikov	Biweight	Triweight	Triangle
$K(t)$	$(2\pi)^{-1/2}e^{-\left(\frac{t^2}{2}\right)}$	$\frac{1}{2}$	$\frac{3}{4}\left(1-t^2\right)$	$\frac{15}{16}\left(1-t^2\right)^2$	$\frac{35}{32}\left(1-t^2\right)^3$	$1-\lvert t\rvert$

Note. For all kernels $|t| \leq 1$, 0 otherwise.

commonly used methods of choosing a smoothing parameter in practice, visually and automatically from the data. Visually selecting a bandwidth involves trial and error. This method can be very effective in determining the shape of the underlying distribution but is flawed. A trial and error approach can be time consuming and lacks objectivity (Wand & Jones, 1995). For many density estimation problems in the social and behavioral sciences, visually selecting the bandwidth would not be recommended. Automatic bandwidth selection methods use the data to generate a suitable bandwidth automatically. The goal of data driven bandwidth selection is to alleviate the subjectivity of visually selecting a bandwidth and quickly and accurately select an optimal smoothing parameter. This has been researched extensively in mathematical statistics (Bowman & Azzalini, 1997; Cao et al., 1994; Devroye, 1997; Jones et al., 1996; Park & Marron, 1990; Scott, 1992; Silverman, 1986; Scott & Terrell, 1987; Sheather, 1992; Sheather & Jones, 1991; Wand & Jones, 1995). However, very little research involving the impact of bandwidth selection methods on graphical illustrations exists in the social and behavioral sciences literature. The next sections describes KDE and common bandwidth selection algorithms in more detail.

1.2 Background on Kernel Density Estimation

In KDE it is assumed we are given a sample of n identically and independently distributed (iid) observations $X_1, X_2, \ldots, X_n$ from which a density will be estimated. Let $f(x)$ be the true probability density function (PDF) and $\hat{f}(x;h)$ be the estimated PDF. The kernel density estimate of $f(x)$ is

$$\hat{f}(x;h) = n^{-1}h^{-1}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right), \tag{1.1}$$

where K is the kernel function that satisfies $\int K(y)dy = 1$, $\int yK(y)dy = 0$, and $0 < \int y^2K(y)dy < \infty$, all odd moments are zero ,and h is the bandwidth or smoothing parameter (Silverman, 1986; Wand & Jones, 1995). The $\int y^2K(y)dy$ function is denoted by μ_2 which indicates the second moment of a PDF. When $\mu_2 > 0$ then the kernel is said to be of order two. The unsigned integral symbol $\int$ is taken to be over the real line unless otherwise noted. A more compact way to express Equation 1.1 is by letting $u = x - X_i$ and $K_h(u) = h^{-1}K(u/h)$:

$$\hat{f}(x;h) = n^{-1}\sum_{i=1}^{n} K_h(u_i). \tag{1.2}$$

To ensure that $\hat{f}(x;h)$ is a proper PDF, the kernel K should be chosen to be a unimodal PDF that is symmetric about zero (Mugdadi & Jeter, 2010; Scott, 1992; Silverman, 1986; Wand & Jones, 1995). If the kernel is not chosen in this fashion and/or we let $\mu_2 = 0$ and $\mu_4 > 0$ then μ_4 is the first non-zero even order which is known as a higher-order kernel but is not a proper PDF (see Scott, 1992, pp. 110-114 and Wand & Jones, 1995, pp. 32-33 for details). There are reasons for using higher-order kernels but they are not the focus of this paper and in most practical applications the sample sizes are not feasible to notice a difference (Wand & Jones, 1995, p. 34). For the present study the Normal kernel was chosen for K, defined as

$$K(y) = (2\pi)^{-\frac{1}{2}} exp\left(-\frac{y^2}{2}\right). \tag{1.3}$$

1.3 Kernel Density Error Criteria

Suitable error criteria must be evaluated to examine the performance of various bandwidth selection methods. There are multiple error criteria that could be used such as mean integrated square error (MISE), mean integrated absolute error (MIAE), mean uniform absolute error (MUAE), and mean Hellinger distance (MHD) (Cao et al., 1994; Mugdadi & Jeter, 2010; Wand & Jones, 1995). Most studies, however, have analyzed MISE because it is substantially easier to work with (Jones et al., 1996; Marron & Wand, 1992; Mugdadi & Jeter, 2010; Park & Marron, 1990; Scott & Ter-

rell, 1987; Sheather & Jones, 1991). Additionally, Marron and Wand (1992) identified closed form expressions for the exact MISE of Normal mixture densities with a Normal kernel. For these reasons, the present study used the MISE for the error criterion in evaluating bandwidth selection performance.

A discrepancy measure between $f(x)$ and $\hat{f}(x;h)$ at a specific point is the mean squared error (MSE). The MSE can be expressed as

$$MSE_x\hat{f}(x;h) = \mathbb{E}[\hat{f}(x;h) - f(x)]^2, \tag{1.4}$$

which can be stated in terms of the squared bias and variance as

$$\begin{aligned}
\mathbb{E}[\hat{f}(x;h) - f(x)]^2 &= \mathbb{E}[(\hat{f}(x;h) - \mu_1 + \mu_1 - f(x))^2] \\
&= \mathbb{E}[(\hat{f}(x;h) - f(x))^2] + 2\mathbb{E}[(\hat{f}(x;h) - \mu_1)(\mu_1 - f(x))] + \mathbb{E}[(\mu_1 - f(x))^2] \\
&= [\mu_1 - f(x)]^2 + Var[\hat{f}(x;h)] \\
&= \{Bias[\hat{f}(x;h)]\}^2 + Var[\hat{f}(x;h)], \qquad (1.5)
\end{aligned}$$

where $\mu_1 = \mathbb{E}[\hat{f}(x;h)]$ (here μ_1 indicates the mean and not the second moment of the distribution as mentioned previously) (Wand & Jones, 1995, p. 14). Since Equation 1.5 only calculates the discrepancy at a single point, a loss function is needed to assess the error over the real line. Taking the integral of Equation 1.5 over the real line gives the MISE, which is an average global discrepancy criterion. The MISE is given by

$$MISE_{\hat{f}(x;h)} = \mathbb{E}[\int \{\hat{f}(x;h) - f(x)\}^2 dx], \tag{1.6}$$

which by Fubini's Theorem (see Colley, 2011 pp. 319-320) is equivalent to

$$MISE_{\hat{f}(x;h)} = \int \{\mathbb{E}[\hat{f}(x;h) - f(x)]\}^2 dx + \int Var[\hat{f}(x;h)]dx. \tag{1.7}$$

Taking the expectation of Equation 1.2

$$\begin{aligned}\mathbb{E}[\hat{f}(x;h)] &= \mathbb{E}[n^{-1}\sum_{i=1}^{n} K_h(x-X_i)] \\ &= n^{-1}n\mathbb{E}[K_h(x-X)] \\ &= \mathbb{E}[K_h(x-X)],\end{aligned}$$

using the fact that $\mathbb{E}[g(x)] = \int g(x)f(x)dx$, and the definition of convolution $f * g = \int f(x-y)g(y)dy$,

$$\begin{aligned}\mathbb{E}[K_h(x-X)] &= K_h * g \\ &= \int K_h(x-y)g(y)dy \qquad (1.8)\end{aligned}$$

the bias can be written as $(K_h * f)(x) - f(x)$ and the variance $\{(K_h * f)^2(x)\} + (K_h * f)^2(x)\}$ giving the MISE as

$$MISE_{\hat{f}(x;h)} = n^{-1}\int\{(K_h^2 * f)(x) - (K_h * f)^2(x)\}dx + \int\{(K_h * f)(x) - f(x)\}^2 dx. \qquad (1.9)$$

After some manipulation (see Jeter, 2005, pp. 6-7 and Wand & Jones, 1995, pp. 14-16) the MISE can be expressed as

$$(nh)^{-1}\int K_h^2(x)dx + (1-n^{-1})\int (K_h * f)^2(x)dx - 2\int (K_h * f)(x)f(x)dx + \int f^2(x)dx. \quad (1.10)$$

Note that convolution is used to multiply two functions together over a specified interval to create a new function that blends the two expressions together (see Weisstein, 2013 for more details). Throughout this paper convolution is used for parsimony. Using Equation 1.10 to measure the performance of a bandwidth selection method is straightforward; however, its dependence on the bandwidth h is complex.

To understand the relationship between bandwidth and MISE, an asymptotic approximation to

the MISE is used called the asymptotic mean integrated squared error (AMISE). To motivate the AMISE Taylor series expansions of the bias and variance are carried out to a derivative order of two to ensure that the KDE is a proper density. Using Equation 1.8 and a change of variables by letting $t = (x-y)/h$ the Jacobian of t is h, then $\mathbb{E}[\hat{f}(x;h)] = \int K(t)f(x-ht)dt$. Using a second order Taylor expansion for $f(x-ht)$ gives

$$f(x-ht) = f(x) - htf'(x) + \frac{1}{2}h^2t^2f''(x) + O(h^4)$$

which leads to

$$\begin{aligned}
\mathbb{E}[\hat{f}(x;h)] &= \{f(x) - htf'(x) + \frac{1}{2}h^2t^2f''(x) + O(h^4)\}\int K(t)dt \\
\mathbb{E}[\hat{f}(x;h)] &= f(x) + \frac{1}{2}h^2t^2f''(x)\int K(t)dt + O(h^4) \\
\mathbb{E}[\hat{f}(x;h)] - f(x) &= \frac{1}{2}h^2t^2f''(x)\int K(t)dt + O(h^4) \\
Bias_{\hat{f}(x;h)} &= \frac{1}{2}h^2f''(x)\int t^2K(t)dt + O(h^4) \\
Bias_{\hat{f}(x;h)} &\approx \frac{1}{2}h^2\mu_2(K)f''(x). \qquad (1.11)
\end{aligned}$$

Note that the symbol $O(h^4)$ indicates that there exists a constant $c > 0$ such that as h approaches zero then the higher order terms in the Taylor expansion remain bounded by ch^4 (see Lange, 2010 pp. 39-43 for details. Using Equation 1.9:

$$n^{-1}\{(K_h^2 * f)(x) - (K_h * f)^2(x)\} = (nh)^{-1}\int K^2(t)f(x-th)dt - n^{-1}\int K(t)f(x-ht)dt$$

the variance can be approximated via a first order Taylor expansion (see Wolter, 2007, p. 231 for

reasoning) as

$$
\begin{aligned}
n^{-1}\{(K_h^2 * f)(x) - (K_h * f)^2(x)\} &= (nh)^{-1} \int K^2(t)\{f(x) + ht f'(x) + \cdots\}dt - n^{-1} \int K(t) f(x-ht)dt \\
&= (nh)^{-1} \int K^2(t)\{f(x) + o(1)\}dt - n^{-1}(0) \\
&= \frac{R(K)f(x)}{nh} + o\left(\frac{1}{nh}\right). \qquad (1.12)
\end{aligned}
$$

Where $R(g) = \int g(z)^2 dz$ is parsimonious notation for any square integrable function. Note also that $o\left(\frac{1}{nh}\right)$ means that $\lim_{x\to\infty} \frac{o(x)}{x} = 0$. (see Lange, 2010 pp. 39-43 for details. Combining the square of Equation 1.11 and Equation 1.12 gives the AMISE as

$$
AMISE = \frac{1}{4} h^4 \mu_2(K)^2 R(f'') + (nh)^{-1} R(K). \qquad (1.13)
$$

Notice that the $f(x)$ term in Equation 1.12 drops out of Equation 1.13 because $f(x)$ is a probability density and integrates to 1 when integrating with respect to x. The term $R(f'')$ from Equation 1.13 appears from squaring and integrating Equation 1.11 which causes $\hat{f}''(x)$ to become $\int f''(x)^2 dx$ leading to $R(f'')$.

The tradeoff between bias and variance is illustrated by the terms in Equation 1.13. The first term represents the squared bias and the second term represents the variance. If the smoothing parameter h is chosen to minimize the bias, then the resulting density will have a large variance and vice versa. The only parameter that is unknown in Equation 1.13 is $R(f'')$ which is a measure of the roughness or curvature of the density. The larger $R(f'')$ is, the larger the AMISE is, and vice versa (Sheather, 2004). In the next chapter, five bandwidth selection algorithms are introduced that will be the focus of this study.

Chapter 2

Kernel Density Bandwidth Selection Methods

2.1 Normal Rule of Thumb

The Normal rule of thumb (NROT) popularized by Silverman (1986) is well known and implemented in most major software packages. It involves differentiating Equation 1.13 with respect to h and then setting the derivative equal to zero and solving for h. When this calculation is performed, the resulting equation is

$$h_{AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')}\right]^{1/5} n^{-1/5}. \tag{2.1}$$

The only unknown value in Equation 2.1 is $R(f'')$ which must be estimated. NROT estimates f'' by using a reference density. The standard choice is to let $f'' = \Phi_{\sigma^2}$, the $[N(0,\sigma^2)]$ density. When a Normal kernel is used and $f'' = \Phi_{\sigma^2}$ is placed in Equation 2.1, the optimal bandwidth obtained is

$$h_{NROT} = 1.06\sigma n^{-1/5} \tag{2.2}$$

(Silverman, 1986, p. 45; Wand & Jones, 1995, p. 60). The idea behind the NROT is that if the random sample is Normally distributed then this selection method should accurately predict the

optimal bandwidth. If the random sample is not Normally distributed then the NROT has been shown to have a tendency to oversmooth densities (Cao et al., 1994; Jones et al., 1996; Scott, 1992; Silverman, 1986; Wand & Jones, 1995).

2.2 Silverman's Rule of Thumb

Silverman (1986, pp. 47-48) recommended reducing the 1.06 factor in Equation 2.2 to .90 to avoid missing bimodality and cope better with various non-normal unimodal densities. Furthermore, he recommended using the smaller of two scale estimates, the sample standard deviation and the sample interquartile range (IQR) divided by 1.34. Thus, this additional estimate is known as Silverman's (1986) rule of thumb (SROT) and is defined as:

$$h_{SROT} = .90An^{-1/5}, \tag{2.3}$$

where $A = \min\{\hat{\sigma}, IQR/1.34\}$. Silverman (1986, pp. 47-48) conducted small simulations studies with a sample size of 100 and found that SROT performed well on densities with skewness or bimodality. Jones et al. (1996) also tested SROT for sample sizes of 100 and 1000 on 15 densities and found that this selector had a high degree of bias for densities with many features.

2.3 Least Squares Cross-Validation

One of the most well-known bandwidth selection methods is least squares cross-validation (LSCV) or unbiased cross-validation (UCV). LSCV was first described in the context of density estimation by Rudemo (1982) and Bowman (1984). Given an estimate $\hat{f}(x;h)$ of a density $f(x)$, the integrated square error (ISE) of $\hat{f}(x;h)$ can be expressed as

$$\begin{aligned} ISE_{\hat{f}(x;h)} &= \int \left(\hat{f}(x;h) - f(x)\right)^2 dx \\ &= \int \hat{f}(x;h)^2 dx - 2\int \hat{f}(x;h) f(x) dx + \int f(x)^2 dx \end{aligned} \tag{2.4}$$

(Silverman, 1986; Wand & Jones, 1995). The $\int f(x)^2 dx$ term does not depend on h and can be safely ignored. The optimal bandwidth will minimize

$$Z(\hat{f}(x;h)) = \int \hat{f}(x;h)dx - 2\int \hat{f}(x;h)f(x)dx. \tag{2.5}$$

The general idea behind LSCV is similar to a leave-one out jackknife procedure. Using Equation 2.5 an estimate of $Z(\hat{f}(x;h))$ must be constructed from the data and then minimized with respect to h to find the optimal bandwidth. Rudemo (1982) noted that the second integral in Equation 2.5 can be written as $\mathbb{E}[\hat{f}(X)]$ where the expectation is with respect of the point of evaluation (as mentioned in Scott, 1992, p. 63). Note that $\mathbb{E}[g(x)] = \int g(x)f(x)dx$ by definition. To estimate $f(x)$, let $\hat{f}_{-i}(x)$ be the density estimate with all the data points except x_i giving

$$\hat{f}_{-i}(x_i) = (n-1)^{-1}h^{-1}\sum_{j\neq i}^{n} K\left(\frac{x - X_j}{h}\right). \tag{2.6}$$

From Equation 2.6 it can be seen that we are taking the average of $h^{-1}\sum_{i\neq j}^{n} K(x - X_j/h)$ from all the points except x_i. This procedure is then repeated for all the remaining data points to estimate the second integral in Equation 2.5. By combining Equation 2.6 with the first term of Equation 2.5 the LSCV function to be minimized is

$$LSCV_h = \int \hat{f}(x;h)^2 dx - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(x_i). \tag{2.7}$$

Minimizing Equation 2.7 with respect to h minimizes the MISE (see Wand & Jones, 1995, p. 63).

While LSCV has been extensively used in practice, research and simulation studies showed that LSCV tended to under smooth densities and had a high degree of sampling variability present. For example, Park and Marron (1990) found that LSCV performed poorly on almost all estimated densities compared with other bandwidth selectors due to the high amount of sampling variability present. These results were further validated by Cao et al. (1994) and Jones et al. (1996). Loader (1999), however, criticizes these studies censuring the behavior of LSCV, emphasizing that they

did not take into account the strengths and weaknesses of LSCV.

2.4 Biased Cross-Validation

Scott and Terrell (1987) created another cross-validation selection method to improve upon the short falls of LSCV known as biased cross-validation (BCV). This method is similar to LSCV and was created to lower the amount of sampling variability that was causing many of the problems mentioned previously. The difference between LSCV and BCV is that BCV is based on the formula for AMISE (see Equation 1.13), whereas, LSCV is based on ISE (see Equation 2.4). However, with BCV we want to replace the unknown estimator $R(f'')$, by an estimator

$$\widetilde{R(f'')} = R\left(\hat{f}''(x;h)\right) - (nh^5)^{-1}R(K''). \tag{2.8}$$

Scott and Terrell (1987) show that plugging in $R(f'')$ directly into Equation 2.1 produces a biased estimate by the amount $(nh)^{-1}R(K'')$. This explains why this term is being subtracted out in Equation 2.8. Plugging Equation 2.8 in Equation 1.13 gives the BCV formula

$$BCV_h = \frac{1}{4}h^4\mu_2(K)^2\widetilde{R(f'')} + (nh)^{-1}R(K) \tag{2.9}$$

(see Scott & Terrell, 1987 and Sheather, 2004 for details). According to Wand and Jones (1995), the attraction of BCV is that the asymptotic variance of the bandwidth is considerably lower than that of LSCV. This reduction in variance comes at a cost of BCV's tendency to oversmooth a density. Several simulation studies indicated that BCV typically performs better than LSCV (Cao et al., 1994; Jones et al., 1996; Park & Marron, 1990).

2.5 Plug-in Methods

Plug-in methods are a popular approach to bandwidth selection where the unknown quantity $R(f'')$ in Equation 2.1 is replaced with an estimate. This method dates back to Woodroofe (1970) and Nadaraya (1974) who laid the theoretical groundwork, yet the issue of selecting an accurate estimate for $R(f'')$ was not addressed until later. Park and Marron (1990) created a plug-in rule that performed superior to BCV and LSCV in asymptotic rate of convergence and a simulation study. Sheather and Jones (1991) further refined the plug-in rule created by Park and Marron (1990), creating the Sheather-Jones plug-in or Sheather-Jones direct plug-in (SJDP), which has performed well in simulation studies, asymptotic analyses, and on real data sets. To find the SJDP use Equation 1.13 (which is the equation for the AMISE) and estimate $R(f'')$ by $R(\hat{f}_g'')$, where g is a pilot bandwidth to be estimated. Next, solve Equation 1.13 for g as a pilot bandwidth estimate that must be estimated. The SJDP approach writes g as a function of h, for the estimate $R(\hat{f}'')$

$$g(h) = C(K)\left[\frac{R(f'')}{R(f''')}\right]^{1/7} h^{5/7}, \tag{2.10}$$

where $C(K)$ is a constant. It is necessary to estimate the unknown higher order functionals of f using kernel density estimates with the NROT method in place of g. The SJDP is given by estimating $g(h)$ and substituting it into Equation 1.13 which gives

$$h = \left[\frac{R(K)}{\mu_2(K)^2 R(\hat{f}''_{g(h)})}\right]^{1/5} n^{-1/5}. \tag{2.11}$$

The SJDP is the smoothing parameter that is the solution to Equation 2.11 (Sheather, 2004). See Sheather and Jones (1991) and Wand and Jones (1995, pp. 67-75) for additional details. The SJDP has performed excellent in simulation studies and on real data sets (Cao et al., 1994; Salgado-Ugarte & Perez-Hernandez, 2003; Sheather, 1992; Sheather & Jones, 1991; Jones et al., 1996).

2.6 The Current Study

The goal of KDE as an EDA approach is to estimate the true underlying density as accurately as possible. Different bandwidth selection algorithms lead to density estimates with varying levels of accuracy. The present study compares the five bandwidth selection methods reviewed above with respect to density recovery using the MISE, with sample size and true density shape as independent variables. Prior simulations involving more than three bandwidths with multiple densities have primarily focused on sample sizes of 100 or more (Cao et al., 1994; Devroye, 1997; Jones et al., 1996). Here, eight different sample sizes will be used, including three that are less than 100, because smaller samples are common in the social sciences. In addition, previous studies have considered a strong positively skewed density similar to a lognormal distribution; however, these studies have not considered a moderately positively skewed density (Cao et al., 1994; Jones et al., 1996; Marron & Wand, 1992). Here, moderately positively skewed densities are included among the true density shapes examined.

Chapter 3

Methods

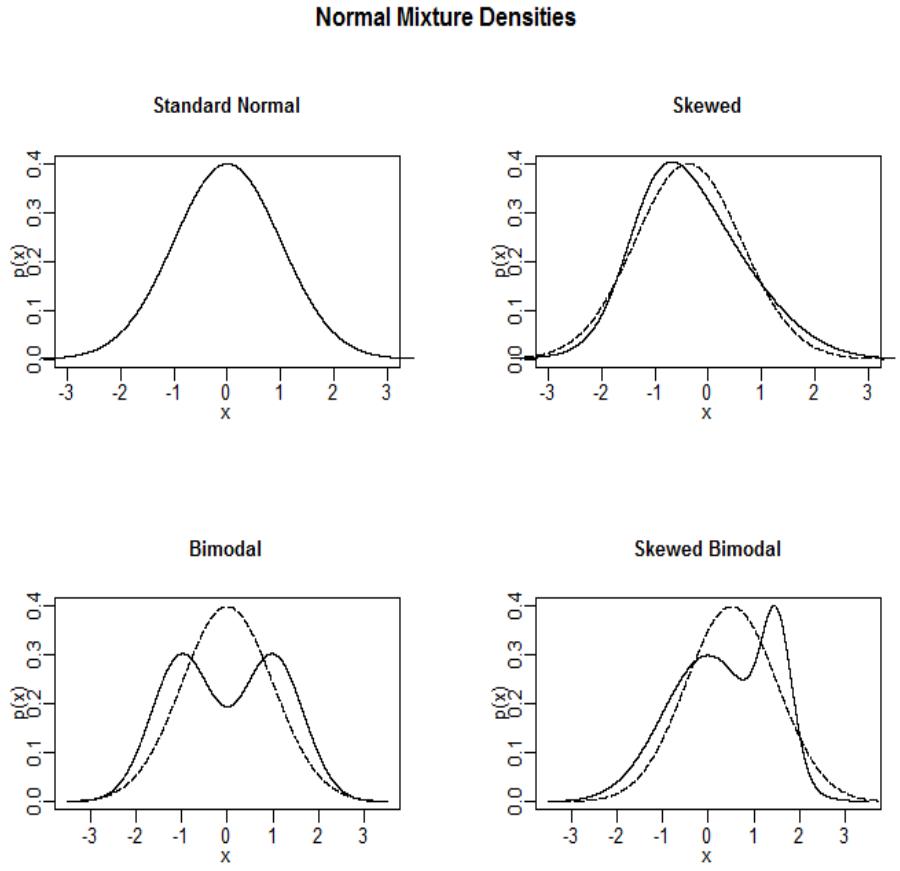
3.1 Simulation Method

An R program (version 2.15) generated the data, executed, and processed the output. The R program utilized the "ks" package to randomly generate data from four Normal mixture densities; a standard Normal, bimodal, positively skewed, and skewed bimodal (R Core Team, 2012; Duong, 2012). R code for the Standard Normal density is available in the appendix. The equations for the Normal mixtures are given in Table 3.1 and graphical illustrations are given in Figure 3.1. These Normal mixtures were chosen to give a general representation of the types of densities most often encountered in the social and behavioral sciences. The positively skewed density is representative of psychopathologies such as anxiety and depression (Carleton et al., 2005; Van Dam & Earleywine, 2011). Both the bimodal and asymmetric bimodal densities can occur in stereotype research as well as intergroup relations (Fiske et al., 2002; Van Boven & Thompson, 2003).

Table 3.1: Equations for the Normal Mixture Densities

Density	Equation
MW.nm1-Normal	$N(0,1)$
MW.nm2-Skewed	$\frac{3}{5}N\left(0,\left(\frac{9}{8}\right)^2\right)+\frac{1}{5}N\left(-\frac{1}{2},\left(\frac{2}{3}\right)^2\right)+\frac{1}{5}N\left(-\frac{13}{12},\left(\frac{5}{9}\right)^2\right)$
MW.nm6-Bimodal	$\frac{1}{2}N\left(-1,\left(\frac{2}{3}\right)^2\right)+\frac{1}{2}N\left(1,\left(\frac{2}{3}\right)^2\right)$
MW.nm8-Skewed Bimodal	$\frac{3}{4}N(0,1)+\frac{1}{4}N\left(\frac{3}{2},\left(\frac{1}{3}\right)^2\right)$

Figure 3.1: Normal Mixture Densities



Note. The dashed line indicates a standard Normal density

The density function within the "stats" package of R was used to calculate the respective bandwidths from each Normal mixture distribution using the five different bandwidth selection methods (NROT, SROT, LSCV, BCV, and SJDP). The Normal kernel was used throughout all simulations and the number of grid points was set equal to the sample size being evaluated. For each density and bandwidth selection method eight sample sizes were evaluated: 25, 50, 75, 100, 250, 500, 1000, and 2000. Sample sizes were chosen to both cover a range typically encountered in psychology, and show how the bandwidth selectors behave with a large sample.

To provide a picture of how the different bandwidth selection methods behave, 32 density graphs were constructed (four Normal mixtures and eight sample sizes). For each of the graphs, the base 10 log was taken for both the optimal bandwidth and each of the 1000 bandwidths selected for

Table 3.2: Exact MISE for the Optimal Bandwidths

Sample Size	Standard Normal	Skewed	Bimodal	Skewed Bimodal
25	0.01373	0.01490	0.01824	0.02219
50	0.00869	0.00948	0.01187	0.01508
75	0.00660	0.00722	0.00907	0.01172
100	0.00541	0.00593	0.00745	0.00972
250	0.00283	0.00311	0.00389	0.00518
500	0.00172	0.00189	0.00234	0.00314
1000	0.00103	0.00113	0.00140	0.00188
2000	0.00061	0.00068	0.00083	0.00112

a given method. Each of the 1000 log draws was subtracted from the log of the optimal bandwidth and stored in a vector. This was repeated for all bandwidth selection methods in each of the 32 graphs. Then a kernel density estimate was plotted using the SJDP bandwidth method for each vector. The resulting zero point along the x-axis of each graph represents the optimal bandwidth. The more concentrated the density estimate is around this point, the better the performance.

3.2 Exact MISE Calculations

To compare the performance of each method, the optimal bandwidth that minimizes the MISE must be calculated for each condition. When using the Normal kernel, a Normal mixture density's MISE has a closed form expression and can be calculated (Marron & Wand, 1992). The calculations to solve for the optimal bandwidth were conducted with the "Hmise.mixt" function within the "ks" R package (Duong, 2012). The "Hmise.mixt" function calculates the optimal bandwidth by specifying a Normal mixture density and a sample size by numerically minimizing the closed form expressions for the MISE given in Equation 1.10. These values were used to calculate the bias and MSE for each normal mixture density. In addition, the "ks" package calculated the exact MISE for each condition using the "mise.mixt" function for each sample size for each of the four normal mixture densities. The MISEs are presented in Table 3.2.

Chapter 4

Results

4.1 Overview

Results are presented in Tables 4.1-4.4 for each true density. The optimal bandwidth is given in bold to the right of each sample size. For each of the five bandwidth selection methods, the mean bandwidth (M), standard deviation (STD), bias, and MSE were calculated for each condition. The density graphs using each bandwidth selection algorithm for a given Normal mixture and sample size are depicted in Figures 4.1-4.4.

To summarize the results and provide practical conclusions, a ranking system was devised based on Cao et al. (1994). Although the ranking strategy was arbitrary, it was designed to provide practitioners with a general idea of relative performance of the bandwidth selection methods given the conditions of the present study. Each of the five bandwidth selection methods was given one total score reflecting how close it was on average to the optimal bandwidth (bias), how variable the bandwidth was over replications (STD), and the interaction between the bias and STD over replications (MSE). The ideal bandwidth selection method should have a low bias, low standard deviation, and low MSE. For each bandwidth selection method, the best bandwidth for a given sample size and true density was given five points for the lowest outcome (best performance), four points for second lowest, and so on to one point for the highest outcome (worst performance). A

higher number of points correspond to better performance.

Table 4.1: Results for the Standard Normal Density by Sample Size and Method

Method	**M**	**STD**	**Bias**	**MSE***	**Method**	**M**	**STD**	**Bias**	**MSE***
Norm									
n = 25	**0.609**				**n = 50**	**0.5199**			
SJDP	0.496	0.123	-0.114	2.816	SJDP	0.453	0.083	-0.067	1.130
NROT	0.502	0.099	-0.107	2.130	NROT	0.457	0.060	-0.063	0.751
SROT	0.426	0.084	-0.183	4.057	SROT	0.388	0.051	-0.132	1.992
LSCV	0.517	0.155	-0.093	3.266	LSCV	0.454	0.126	-0.066	2.036
BCV	0.594	0.086	-0.016	0.767	BCV	0.521	0.054	0.001	0.288
n = 75	**0.475**				**n = 100**	**0.446**			
SJDP	0.421	0.064	-0.054	0.698	SJDP	0.399	0.057	-0.047	0.545
NROT	0.422	0.046	-0.052	0.481	NROT	0.403	0.037	-0.043	0.323
SROT	0.359	0.039	-0.116	1.497	SROT	0.342	0.032	-0.104	1.175
LSCV	0.422	0.105	-0.053	1.377	LSCV	0.392	0.101	-0.053	1.309
BCV	0.475	0.040	0.000	0.160	BCV	0.449	0.034	0.004	0.117
n = 250	**0.365**				**n = 500**	**0.315**			
SJDP	0.341	0.032	-0.024	0.158	SJDP	0.297	0.022	-0.018	0.082
NROT	0.344	0.019	-0.021	0.080	NROT	0.301	0.012	-0.014	0.034
SROT	0.292	0.016	-0.073	0.557	SROT	0.256	0.010	-0.059	0.363
LSCV	0.329	0.076	-0.036	0.711	LSCV	0.285	0.064	-0.030	0.504
BCV	0.373	0.019	0.008	0.045	BCV	0.321	0.016	0.006	0.028
n = 1000	**0.272**				**n = 2000**	**0.236**			
SJDP	0.261	0.015	-0.011	0.035	SJDP	0.227	0.010	-0.009	0.018
NROT	0.264	0.007	-0.008	0.013	NROT	0.230	0.004	-0.005	0.005
SROT	0.224	0.006	-0.048	0.237	SROT	0.196	0.004	-0.040	0.163
LSCV	0.250	0.053	-0.022	0.331	LSCV	0.217	0.045	-0.018	0.238
BCV	0.277	0.012	0.005	0.017	BCV	0.238	0.010	0.002	0.010

Note. * indicates that the value of MSE has been multiplied by 100.

Table 4.2: Results for the Skewed Density by Sample Size and Method

Method	**M**	**STD**	**Bias**	**MSE***	**Method**	**M**	**STD**	**Bias**	**MSE***
Norm									
n = 25	**0.595**				**n = 50**	**0.501**			
SJDP	0.507	0.121	-0.088	2.234	SJDP	0.453	0.080	-0.048	0.877
NROT	0.521	0.103	-0.074	1.614	NROT	0.474	0.066	-0.027	0.511
SROT	0.442	0.088	-0.152	3.093	SROT	0.403	0.056	-0.099	1.291
LSCV	0.534	0.162	-0.061	3.003	LSCV	0.458	0.136	-0.044	2.030
BCV	0.619	0.092	0.025	0.900	BCV	0.542	0.061	0.041	0.539
n = 75	**0.455**				**n = 100**	**0.425**			
SJDP	0.421	0.062	-0.034	0.497	SJDP	0.400	0.053	-0.026	0.344
NROT	0.441	0.047	-0.015	0.239	NROT	0.420	0.041	-0.006	0.171
SROT	0.374	0.040	-0.081	0.815	SROT	0.356	0.035	-0.069	0.598
LSCV	0.420	0.117	-0.035	1.495	LSCV	0.401	0.104	-0.024	1.138
BCV	0.493	0.044	0.037	0.330	BCV	0.463	0.040	0.038	0.301
n=250	**0.345**				**n=500**	**0.296**			
SJDP	0.334	0.031	-0.011	0.106	SJDP	0.290	0.021	-0.006	0.045
NROT	0.358	0.022	0.013	0.063	NROT	0.313	0.014	0.017	0.049
SROT	0.304	0.019	-0.041	0.206	SROT	0.266	0.012	-0.030	0.106
LSCV	0.329	0.081	-0.016	0.681	LSCV	0.285	0.063	-0.011	0.414
BCV	0.376	0.028	0.031	0.176	BCV	0.318	0.022	0.022	0.094
n=1000	**0.254**				**n=2000**	**0.220**			
SJDP	0.251	0.014	-0.004	0.022	SJDP	0.218	0.010	-0.002	0.009
NROT	0.274	0.009	0.020	0.046	NROT	0.240	0.005	0.020	0.044
SROT	0.233	0.007	-0.022	0.052	SROT	0.204	0.004	-0.016	0.027
LSCV	0.245	0.053	-0.009	0.292	LSCV	0.212	0.043	-0.007	0.193
BCV	0.268	0.017	0.013	0.045	BCV	0.229	0.013	0.009	0.024

Note. * indicates that the value of MSE has been multiplied by 100.

Table 4.3: Results for the Bimodal Density by Sample Size and Method

Method	**M**	**STD**	**Bias**	**MSE***	**Method**	**M**	**STD**	**Bias**	**MSE***
Norm									
n=25	**0.603**				**n=50**	**0.472**			
SJDP	0.584	0.112	-0.019	1.281	SJDP	0.504	0.082	0.032	0.771
NROT	0.657	0.080	0.054	0.930	NROT	0.580	0.045	0.108	1.371
SROT	0.558	0.068	-0.045	0.657	SROT	0.493	0.038	0.021	0.186
LSCV	0.569	0.186	-0.034	3.575	LSCV	0.482	0.148	0.010	2.196
BCV	0.716	0.076	0.113	1.858	BCV	0.623	0.049	0.151	2.526
n=75	**0.418**				**n=100**	**0.385**			
SJDP	0.449	0.068	0.032	0.557	SJDP	0.421	0.059	0.035	0.476
NROT	0.536	0.033	0.119	1.513	NROT	0.507	0.026	0.122	1.555
SROT	0.455	0.028	0.038	0.219	SROT	0.431	0.022	0.045	0.255
LSCV	0.429	0.127	0.012	1.628	LSCV	0.398	0.117	0.013	1.387
BCV	0.572	0.042	0.154	2.561	BCV	0.536	0.047	0.151	2.499
n=250	**0.304**				**n=500**	**0.258**			
SJDP	0.332	0.035	0.027	0.197	SJDP	0.275	0.022	0.017	0.077
NROT	0.422	0.014	0.117	1.389	NROT	0.368	0.008	0.109	1.203
SROT	0.358	0.012	0.053	0.299	SROT	0.312	0.007	0.054	0.295
LSCV	0.308	0.082	0.003	0.675	LSCV	0.255	0.056	-0.003	0.317
BCV	0.391	0.054	0.087	1.050	BCV	0.299	0.036	0.041	0.292
n=1000	**0.221**				**n=2000**	**0.190**			
SJDP	0.232	0.014	0.011	0.032	SJDP	0.197	0.008	0.007	0.012
NROT	0.320	0.005	0.099	0.981	NROT	0.278	0.003	0.089	0.789
SROT	0.272	0.004	0.051	0.259	SROT	0.236	0.003	0.047	0.219
LSCV	0.219	0.043	-0.002	0.186	LSCV	0.186	0.033	-0.004	0.107
BCV	0.242	0.020	0.021	0.083	BCV	0.201	0.012	0.012	0.028

Note. * indicates that the value of MSE has been multiplied by 100.

Table 4.4: Results for the Skewed Bimodal Density by Sample Size and Method

Method	M	STD	Bias	MSE*	Method	M	STD	Bias	MSE*
Norm									
n=25	**0.555**				**n=50**	**0.408**			
SJDP	0.527	0.107	-0.027	1.210	SJDP	0.465	0.077	0.057	0.911
NROT	0.586	0.084	0.031	0.809	NROT	0.529	0.047	0.121	1.680
SROT	0.498	0.072	-0.057	0.842	SROT	0.450	0.039	0.041	0.325
LSCV	0.516	0.180	-0.038	3.387	LSCV	0.427	0.147	0.018	2.203
BCV	0.649	0.079	0.094	1.512	BCV	0.572	0.049	0.163	2.906
n=75	**0.351**				**n=100**	**0.318**			
SJDP	0.414	0.061	0.063	0.773	SJDP	0.386	0.050	0.068	0.713
NROT	0.486	0.034	0.135	1.951	NROT	0.461	0.029	0.143	2.130
SROT	0.413	0.029	0.062	0.470	SROT	0.391	0.025	0.073	0.600
LSCV	0.371	0.128	0.020	1.681	LSCV	0.340	0.109	0.023	1.250
BCV	0.522	0.038	0.172	3.089	BCV	0.494	0.035	0.176	3.211
n=250	**0.242**				**n=500**	**0.201**			
SJDP	0.295	0.032	0.053	0.385	SJDP	0.241	0.021	0.040	0.201
NROT	0.385	0.014	0.144	2.080	NROT	0.335	0.009	0.134	1.795
SROT	0.327	0.012	0.085	0.744	SROT	0.284	0.008	0.083	0.697
LSCV	0.248	0.070	0.006	0.492	LSCV	0.205	0.048	0.003	0.228
BCV	0.386	0.052	0.144	2.345	BCV	0.274	0.050	0.072	0.778
n=1000	**0.170**				**n=2000**	**0.144**			
SJDP	0.197	0.014	0.028	0.094	SJDP	0.163	0.008	0.018	0.039
NROT	0.292	0.006	0.122	1.489	NROT	0.254	0.003	0.109	1.197
SROT	0.248	0.005	0.078	0.609	SROT	0.215	0.003	0.071	0.506
LSCV	0.171	0.034	0.001	0.117	LSCV	0.142	0.025	-0.002	0.064
BCV	0.199	0.023	0.030	0.141	BCV	0.159	0.012	0.015	0.036

Note. * indicates that the value of MSE has been multiplied by 100.

4.2 Aggregated Tabular Results

The aggregated results are as follows (scores are in parentheses): SJDP (314), SROT (311), NROT (306), BCV (287), and LSCV (222). As can be seen, the overall winner is the SJDP closely followed by SROT and NROT. However, these results do not tell the whole story because they are aggregated. To provide further detail, the rankings are broken down by sample size and presented in Table 4.5. One result that is consistent between both the aggregated results and the results in Table 4.5 is that LSCV is last in every category. Considering Tables 4.1-4.4, LSCV typically has the lowest bias but it always has the highest variance. Table 4.5 better describes how the optimal

method depends on sample size. Looking at sample sizes of 25 to 100, SROT is the best method closely followed by NROT, and then SJDP. Once the sample size increases to 250, SJDP supplants SROT and dominates for sample sizes of 250 to 2000. A final interesting result is the unique behavior of the BCV at higher sample sizes. When sample sizes are 25 to 100, BCV closely tails SJDP, and at a sample size of 250 is several points behind both SROT and NROT. At a sample size of 500, BCV ties with NROT, and at 1000 and 2000 BCV, surpasses both SROT and NROT.

Table 4.5: Rankings of Bandwidth Selection Methods Aggregated by Sample Size

Sample Size							
25		**50**		**75**		**100**	
Method	Total	Method	Total	Method	Total	Method	Total
NROT	42	SROT	44	SROT	42	SROT	42
BCV	41	NROT	43	NROT	41	NROT	40
SROT	38	SJDP	33	SJDP	35	SJDP	36
SJDP	34	BCV	33	BCV	34	BCV	34
LSCV	25	LSCV	27	LSCV	28	LSCV	28
Sample Size							
250		**500**		**1000**		**2000**	
Method	Total	Method	Total	Method	Total	Method	Total
SJDP	43	SJDP	45	SJDP	45	SJDP	43
NROT	39	SROT	36	BCV	38	BCV	41
SROT	39	NROT	35	NROT	34	SROT	36
BCV	31	BCV	35	SROT	34	NROT	32
LSCV	28	LSCV	29	LSCV	29	LSCV	28

4.3 Graphical Results

Results are graphically shown in Figures 4.1-4.4. The vertical black line at zero on the x-axis denotes the optimal bandwidth, with plots left and right of this line indicating oversmoothing and undersmoothing respectively. For all true density shapes, LSCV varied tremendously in every condition shown by the large tail of its distribution. SJDP had greater variability in sample sizes 25 to 100 compared with samples of 250 to 2000. For samples of 250 to 2000, SJDP had the best balance of small bias, variance, and MSE for all true density shapes. SROT slightly undersmoothed for the Normal and positively skewed curves (Figures 4.1 and 4.2) with sample sizes of 100 or

smaller and the undersmoothing became more pronounced for samples of 250 to 2000. SROT oversmoothed the bimodal and skewed bimodal curves (Figures 4.3 and 4.4) for sample sizes of 250 or larger. NROT performed quite well for the Normal distribution (Figure 4.1) with most of its distribution centered around the optimal bandwidth and similarly for sample sizes of less than 100 for the positively skewed distribution (Figure 4.2), with slight oversmoothing occurring for sample sizes 250 and greater. NROT oversmoothed the bimodal and skewed bimodal curves (Figures 4.3 and 4.4) for all sample sizes with the degree of oversmoothing increasing with sample size. BCV oversmoothed for all density shapes in sample sizes of 100 or less and for all density shapes BCV started approaching the optimal bandwidth with the degree of convergence increasing with sample size.

Figure 4.1: Graphical Densities of Standard Normal

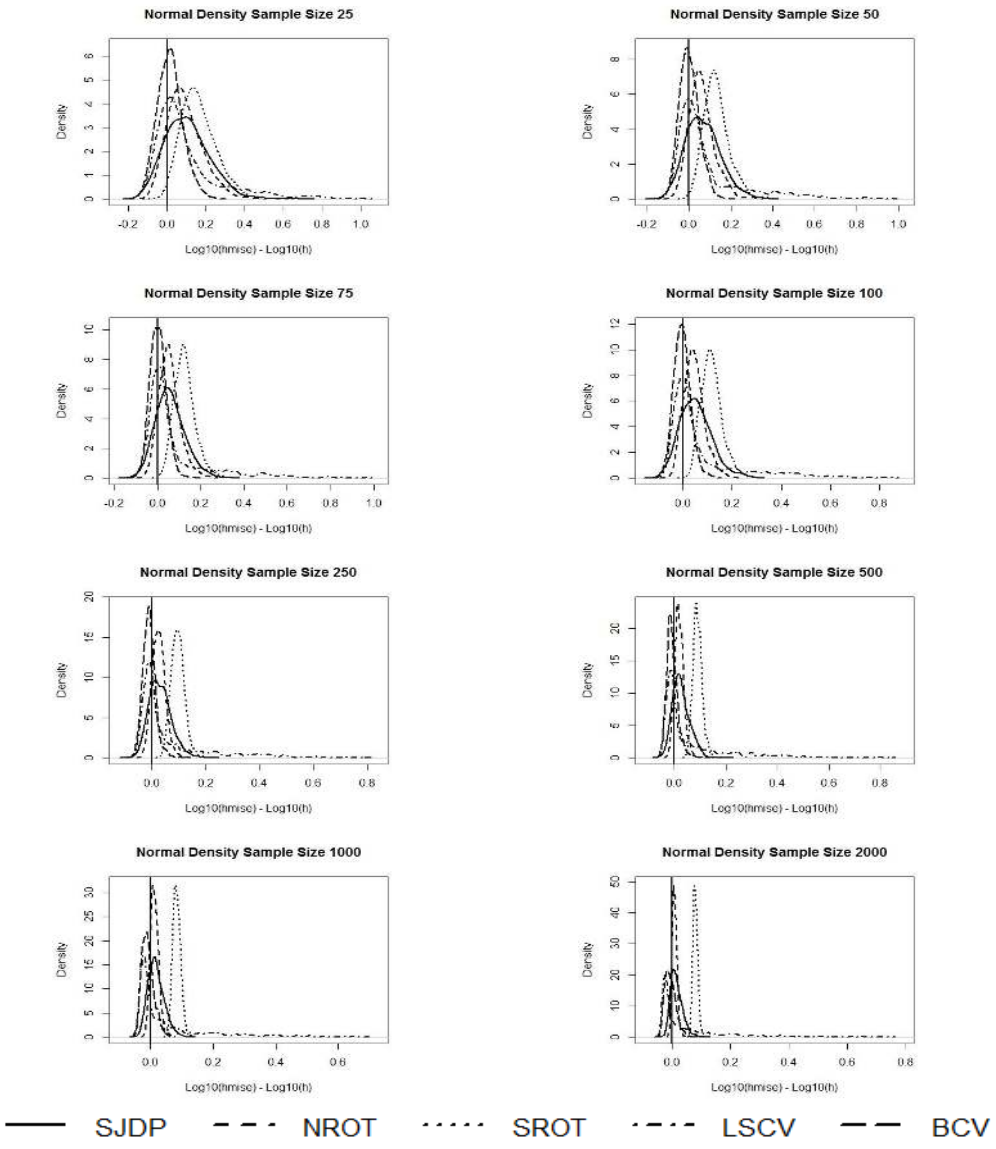


Figure 4.2: Graphical Densities of Skew Normal

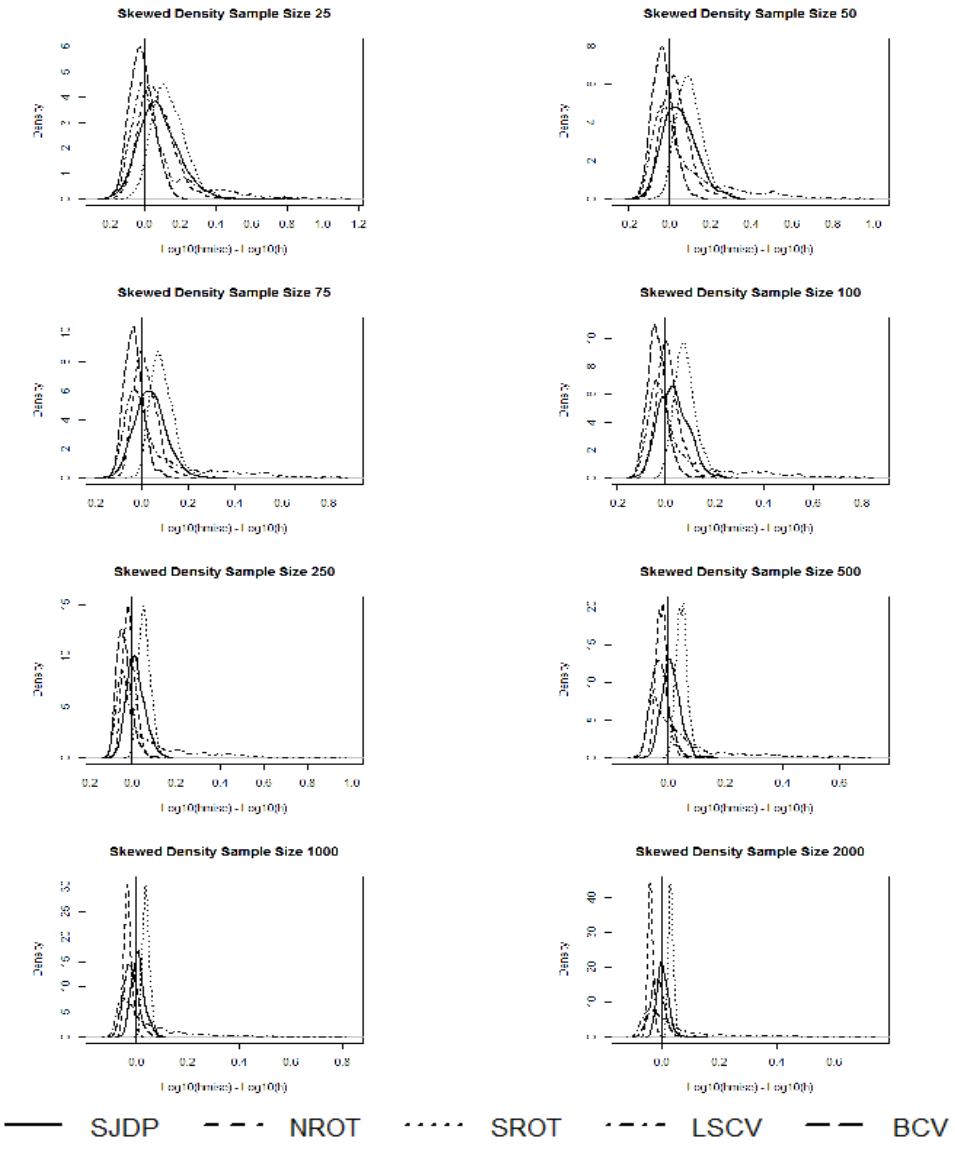


Figure 4.3: Graphical Densities of Bimodal

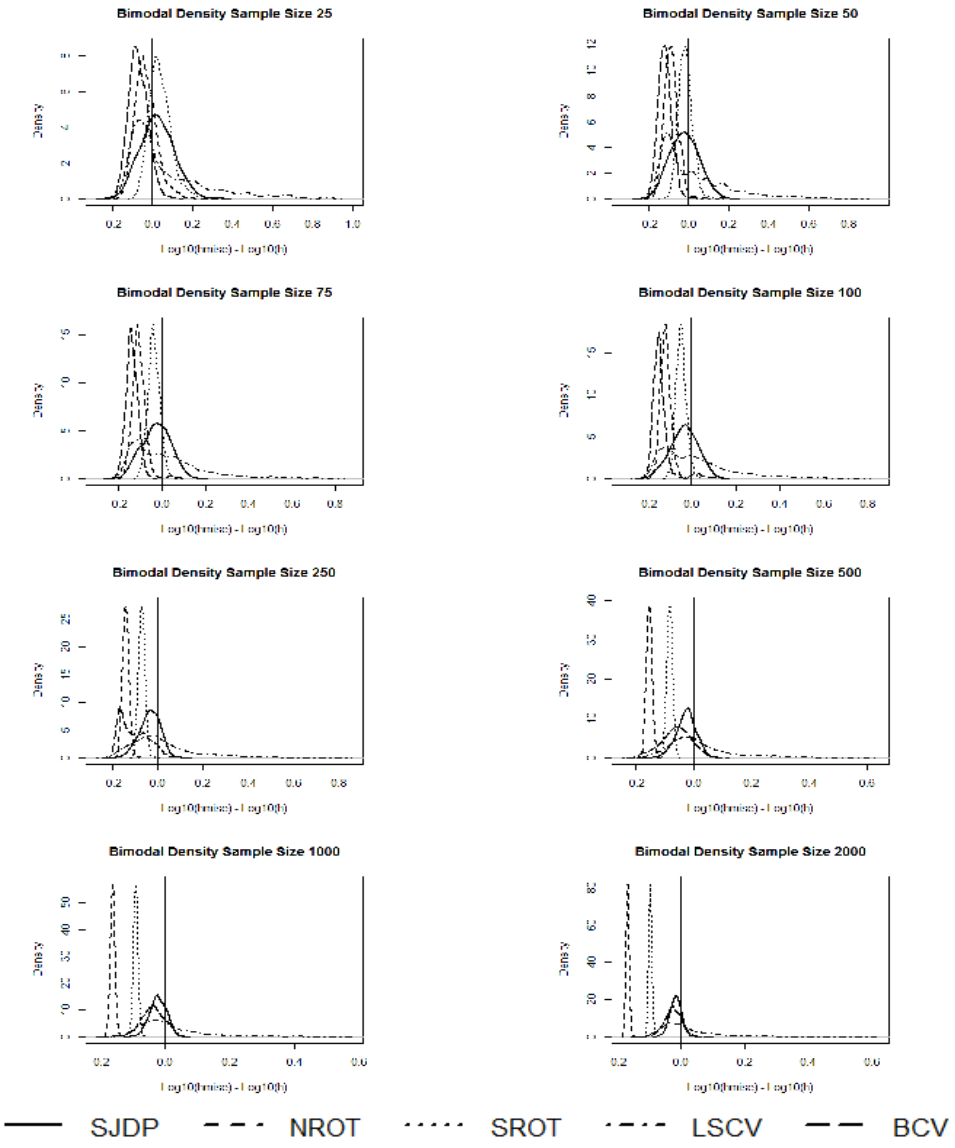
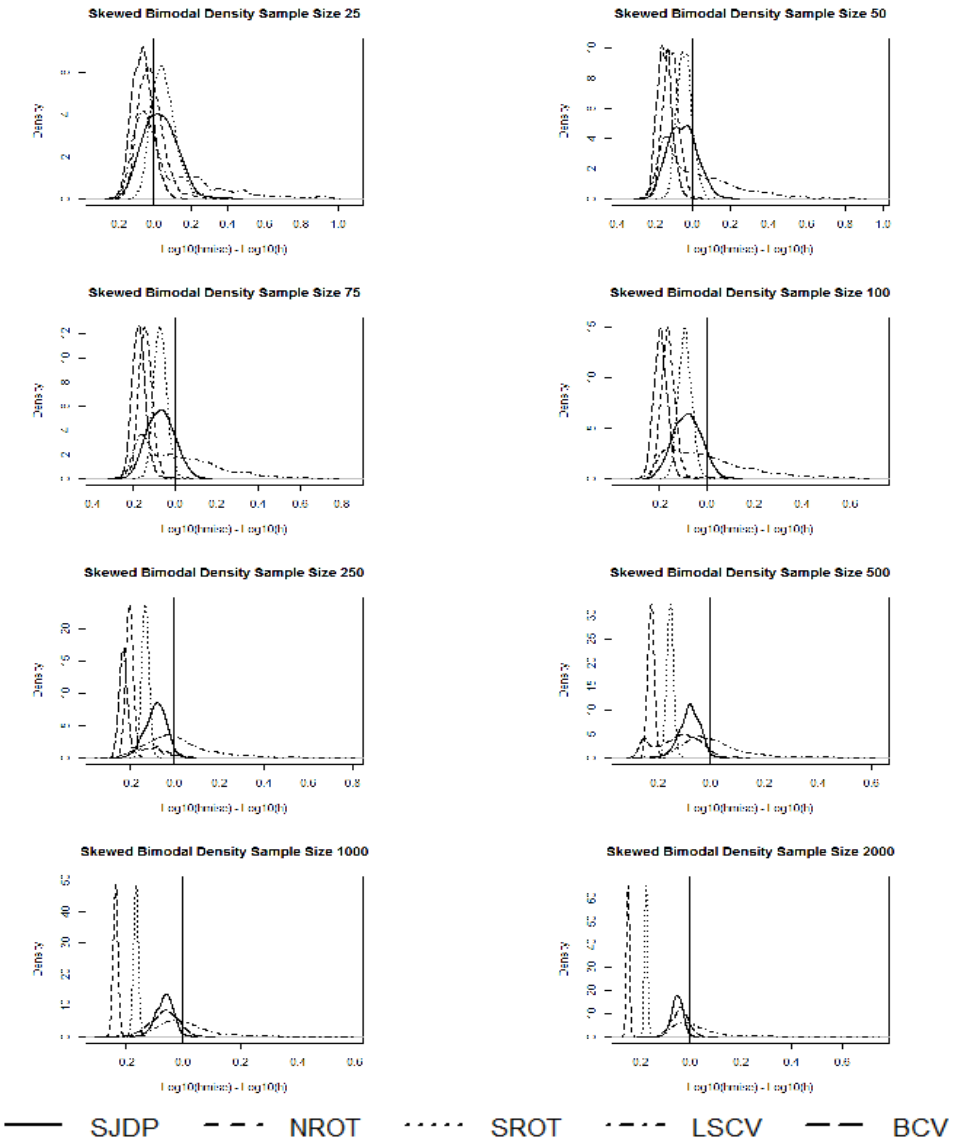


Figure 4.4: Graphical Densities of Skewed Bimodal



Chapter 5

Empirical Example

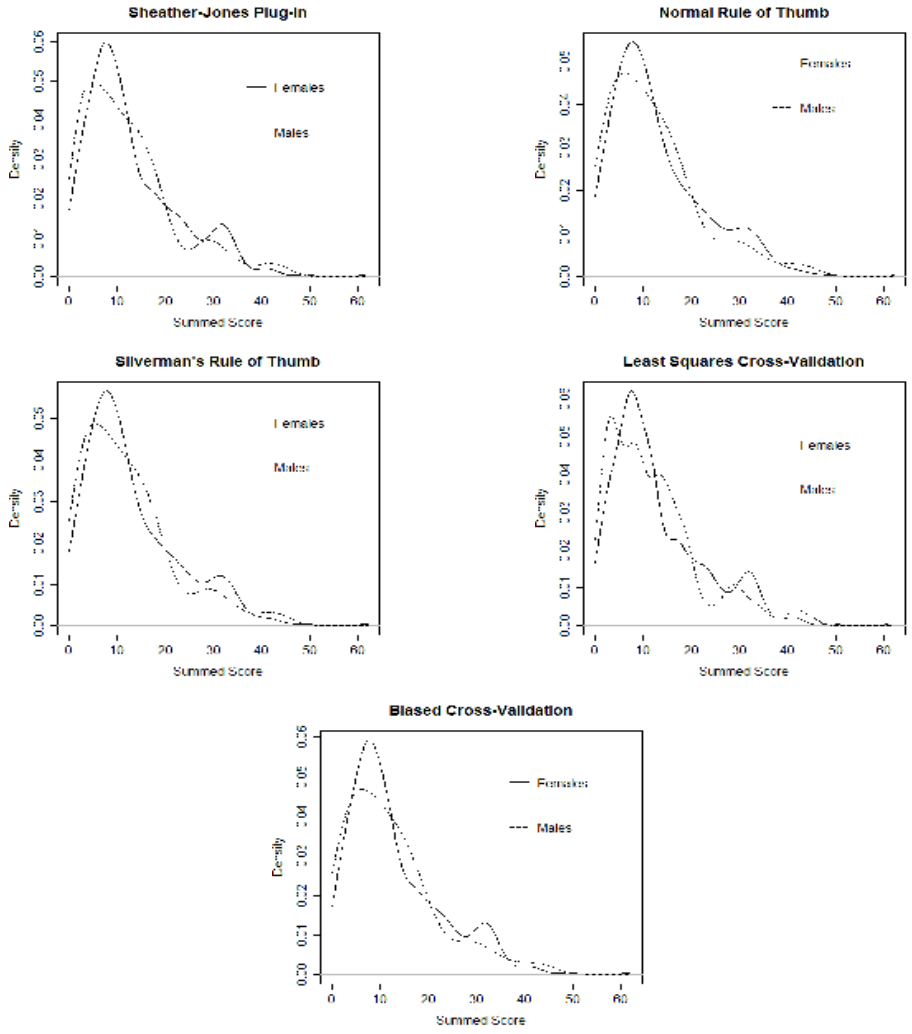
An empirical data set was analyzed to illustrate KDE and the bandwidth selection algorithms. The variable is the summed score distribution on the Obsessive-Compulsive Inventory-Revised (OCI-R; Foa et al., 1998) in a sample of 544 undergraduate students from data described by Tolin et al. (2003). The mean age of participants was 19.01 (SD = 1.87), 74.45% were female and 25.55% were male, 25% were members of ethnic groups other than Caucasian, and 75% were Caucasian. Summed scores are frequently used by psychologists and often presumed to have certain statistical properties. It is a useful exploratory tactic to examine the distribution with KDE.

The summed scores on the entire OCI-R were calculated in R (version 2.15) and the densities of the summed scores were plotted using the density function in the stats package. Each of the five bandwidth selection methods (SJDP, NROT, SROT, LSCV, and BCV) were used to estimate the density of summed scores. The bandwidths chosen by each method are: SJDP (1.77), NROT (2.47), SROT (2.10), LSCV (1.22), and BCV (1.95). The graphs are shown in Figure 5.1.

Consistent with simulation results, LSCV is probably undersmoothing, which creates the spurious modes seen in Figure 5.1. NROT provides the most smoothing, tends to smooth away the mode at around 31, and completely smoothed away the mode around 42. In this example BCV, SROT, and SJDP provide very similar estimates; however, SJDP gives the most vivid picture of the second mode around 31 without spurious artifacts of the data. Given that the sample size is

544, coupled with the results of the simulation study, we prefer the SJDP density for this example. The distribution is clearly skewed and multimodal which is useful information for researchers and practitioners using the OCI-R scores.

Figure 5.1: Density Plots of Discrimination Parameters for the FNE



Chapter 6

Discussion

This study compared the performance of five bandwidth selection methods with respect to density recovery using the MISE, for varying sample sizes and true density shapes. Simulations showed that overall the SJDP bandwidth selector performed best on the three outcome criteria (bias, STD, and MSE). This result is consistent with previous research (Cao et al., 1994; Jones et al., 1996; Mugdadi & Jeter, 2010). Also consistent with previous research was the result that LSCV performed poorly and is not recommended for general use (Cao et al., 1994; Scott & Terrell, 1987; Jones et al., 1996).

The performance of the bandwidth selection methods varied according to sample size. First, for sample sizes at or below 100, SROT was best. Mugdadi and Jeter (2010) compared SJDP, BCV, and LSCV for sample sizes of 10 to 100 using three of the same true densities in the present study. They concluded that SJDP performed better overall than BCV and LSCV, but they did not compare SROT and NROT at those sample sizes. The finding of SROT being superior at lower sample sizes is interesting because it shows that for these four densities SROT would be the recommended method, closely followed by the NROT, which has not been found before. At some point between a sample size of 100 and 250, SJDP usurps SROT and becomes the best performer. It can be seen that SJDP outperforms the other methods by a sizeable amount, which explains the method being superior overall. This finding is consistent with past research (Cao et al., 1994; Jones et al., 1996).

Thus, SJDP is recommended at higher sample sizes.

The simulation results should be qualified by several limitations. First, results are specific to the four true densities. While these densities were chosen to be representative of those frequently observed in psychological research, the underlying density of a random sample is always unknown and the present results may not generalize to all other density shapes that may be observed in practice. Second, although the choice of the kernel is trivial, results are specific to the Normal kernel and it is unknown how these results would differ if other kernels were used.

Future research in this area could look at how adaptive KDE algorithms perform on the four true densities used here and whether they outperform the standard methods. Adaptive KDE algorithms alter the bandwidth size across the area of the density depending on the concentration of the points surrounding the point of interest. Bandwidth size is smaller when there are a large number of points concentrated around the point of interest and lower with a low concentration of points. Wilcox (2004) used an adaptive KDE algorithm to illustrate the utility of the method, but no simulation study was conducted. It would also be interesting to compare the performance of the five bandwidth selection methods with additional true density shapes in the future.

References

Akiskal, H. & Benazzi, F. (2006). The DSM-IV and ICD-10 categories of recurrent [major] depressive and bipolar II disorders: Evidence that they lie on a dimensional spectrum. *Journal of Affective Disorders*, 92(1), 45–54.

Behrens, J. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2), 131–160.

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

Bowman, A. & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations.* Oxford University Press.

Cao, R., Cuevas, A., & Gonzalez Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics & Data Analysis*, 17(2), 153–176.

Carleton, R., Asmundson, G., & Taylor, S. (2005). Fear of physical harm: Factor structure and psychometric properties of the injury/illness sensitivity index. *Journal of Psychopathology and Behavioral Assessment*, 27(4), 235–241.

Colley, S. (2011). *Vector Calculus.* Boston, MA: Pearson Education.

Devroye, L. (1997). Universal smoothing factor selection in density estimation: Theory and practice. *Test*, 6(2), 223–320.

Duong, T. (2012). *ks: Kernel smoothing*. R package version 1.8.10.

Fiske, S., Cuddy, A., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902.

Foa, E., Kozak, M., Salkovskis, P., Coles, M., & Amir, N. (1998). The validation of a new obsessive–compulsive disorder scale: The Obsessive–Compulsive Inventory. *Psychological Assessment*, 10(3), 206–214.

Jeter, J. (2005). Comparison between the exact and the approximate bandwidths in the kernel density estimation. Master's thesis, Southern Illinois University Carbondale.

Jones, C., Marron, J., & Sheather, S. (1996). Progress in data-based bandwidth selection for kernel density estimation. *Computational Statistics*, 11, 337–381.

Kline, R. (2008). *Becoming a behavioral science researcher: A guide to producing research that matters*. New York: Guilford Press.

Lange, K. (2010). *Numerical Analysis for Statisticians*. New York, NY: Springer.

Loader, C. (1999). Bandwidth selection: Classical or plug-in? *The Annals of Statistics*, 27(2), 415–438.

Marmolejo-Ramos, F. & Matsunaga, M. (2009). Getting the most from your curves: Exploring and reporting data using informative graphical techniques. *Tutorials in Quantitative Methods for Psychology*, 5, 40–50.

Marron, J. & Wand, M. (1992). Exact mean integrated squared error. *The Annals of Statistics*, 20, 712–736.

Mugdadi, A. & Jeter, J. (2010). A simulation study for the bandwidth selection in the kernel density estimation based on the exact and the asymptotic MISE. *Pakistan Journal of Statististics*, 26, 239–265.

Nadaraya, E. (1974). On the integral mean square error of some nonparametric estimates for the density function. *Theory of Probability & Its Applications*, 19(1), 133–141.

Osberg, L. & Smeeding, T. (2006). Fair inequality? Attitudes toward pay differentials: the United States in comparative perspective. *American Sociological Review*, 71(3), 450–473.

Park, B. & Marron, J. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409), 66–72.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9, 65–78.

Salgado-Ugarte, I. & Perez-Hernandez, M. (2003). Exploring the use of variable bandwidth kernel density estimators. *Stata Journal*, 3(2), 133–147.

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

Scott, D. & Terrell, G. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400), 1131–1146.

Sheather, S. (1992). The performance of six popular bandwidth selection methods on some real data sets. *Computational Statistics*, 7, 225–281.

Sheather, S. (2004). Density estimation. *Statistical Science*, 19, 588–597.

Sheather, S. & Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, 683–690.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman & Hall/CRC.

Tolin, D., Woods, C., & Abramowitz, J. (2003). Relationship between obsessive beliefs and obsessive–compulsive symptoms. *Cognitive Therapy and Research*, 27(6), 657–669.

Van Boven, L. & Thompson, L. (2003). A look into the mind of the negotiator: Mental models in negotiation. *Group Processes & Intergroup Relations*, 6(4), 387–404.

Van Dam, N. & Earleywine, M. (2011). Validation of the Center for Epidemiologic Studies Depression Scale–Revised : Pragmatic depression assessment in the general population. *Psychiatry Research*, 186(1), 128–132.

Wand, M. & Jones, M. (1995). *Kernel Smoothing*. London: Chapman & Hall/CRC.

Weisstein, E. W. (2013). Convolution – from Wolfram MathWorld. Retrieved from: `http://mathworld.wolfram.com/Convolution.html`.

Wilcox, R. (2001). *Nonparametric estimation*. Hoboken, NJ: John Wiley & Sons, Ltd.

Wilcox, R. (2004). Kernel density estimators: An approach to understanding how groups differ. *Understanding Statistics*, 3(4), 333–348.

Wilcox, R. (2006). Graphical methods for assessing effect size: Some alternatives to Cohen's d. *The Journal of Experimental Education*, 74(4), 351–367.

Wolter, K. (2007). *Introduction to Variance Estimation*. Springer Series in Statistics. New York, NY: Springer.

Woodroofe, M. (1970). On choosing a delta-sequence. *The Annals of Mathematical Statistics*, 41, 1665–1671.

Appendix A

R Code for Standard Normal Simulation

```
##############################################################################
    ##################Kernel Density Simulation Code SS 25#####################################
    ###Clear the workspace rm(list=ls())
    ###Set Working Directory setwd('C:/Users/Jared Harpole/Desktop/R_Working.Directory/Normal Density/KS_NormalDensity/GraphsGray')
    library(nor1mix)
    library(ks)
    sink("KDESimulation_25Norm_ks_Gray.txt", append = TRUE) ###################################
    reps <- 1000 #Specifies the number of replications
    Bandwidth <- matrix(NA, reps, 6) #Specifies the bandwidth matrix for analyzing bandwidth results across replications
    BW.sd <- list() # Specifies the Result CI for Density [j]
    BW.mn <- list() # Specifies the Bandwidth CI for Density [j]
    colnames(Bandwidth) <- c("SJ-ste","SJ-dpi", "nrd","nrd0", "ucv", "bcv") #Specifies the column names of Bandwidth
    #Generate data for positively skewed distribution
    mus <- c(0, 0)
    sigmas <- c(1, 1)
    props <- c(.5, .5)
    N <- 25
    #Specifies the number of observations sampled from the mixture distribution #######
```

```
set.seed(10001) ##############################################################################################
for (i in 1:reps) {
#Randomly Draw N observations from Marron & Wand Mixture Distribution specified above.
MW1_10 <- rnorm.mixt(n=N, mus, sigmas, props)
# Calculate the density values using each bandwidth method (aka, SJ, Normal, LSCV, BCV)
MW1_10_SJ <-density(MW1_10, bw="SJ-ste", n=N)
MW1_10_SJdp <- density(MW1_10, bw="SJ-dpi", n=N)
MW1_10_nrd <-density(MW1_10, bw="nrd", n=N)
MW1_10_nrd0 <- density(MW1_10, bw="nrd0", n=N)
MW1_10_ucv <-suppressWarnings(density(MW1_10, bw="ucv", n=N))
MW1_10_bcv <-suppressWarnings(density(MW1_10, bw="bcv", n=N))
# Calculates the bandwidths used in each repetition
SJbw <- MW1_10_SJ$bw
SJdpbw <- MW1_10_SJdp$bw
NRDbw <- MW1_10_nrd$bw
NRD0bw <- MW1_10_nrd0$bw
UCVbw <- MW1_10_ucv$bw
BCVbw <- MW1_10_bcv$bw
#Creates a vector with the bandwidths used in each repetition
Bandwidth[i,] <- c(SJbw,SJdpbw, NRDbw,NRD0bw, UCVbw, BCVbw)
}
#Output of Bandwidth
temp2 <- as.data.frame(Bandwidth)
BW.mn <- sapply(temp2, mean)
BW.sd <- sapply(temp2, sd)
print("Mean Bandwidth By Method")
print(BW.mn)
print("Standard Deviation By Method")
print(BW.sd) sink()
###Graph the bandwidths
samp <- N
hmise.opt <- hmise.mixt(mus=mus, sigmas=sigmas, props=props, samp=samp, deriv.order=0)
hmise.opt
```

```
###Create Bandwidth.g matrix for graphing the density
BW2 <- log10(hmise.opt)-log10(Bandwidth[,2])
BW3 <- log10(hmise.opt)-log10(Bandwidth[,3])
BW4 <- log10(hmise.opt)-log10(Bandwidth[,4])
BW5 <- log10(hmise.opt)-log10(Bandwidth[,5])
BW6 <- log10(hmise.opt)-log10(Bandwidth[,6])
###Function for plotting multiple densities
plot.5.dens <- function(s) {
knot.x <- NULL
knot.y <- NULL
for (i in 1:length(s)) {
knot.x <- c(knot.x, density(s[[i]])$x)
knot.y <- c(knot.y, density(s[[i]])$y)
}
xr <- range(knot.x)
yr <- range(knot.y)
plot(density(s[[1]], bw = "SJ-dpi"), xlim = xr, ylim = yr, ylab = "Density", xlab ="Log10(hmise) - Log10(h)",
main = "Normal Density Sample Size 25")
for (i in 1:length(s)) { lines(density(s[[i]], bw = "SJ-dpi"), xlim = xr, ylim = yr, col = i*(2*i)^((1+(-1)^i)/2),
lty = i, lwd = 2) } }
###Plotting the densities
plot.5.dens(list(BW2, BW3, BW4, BW5,BW6))
abline(a = NULL, b = NULL, h = NULL, v = 0, lwd=2)
lines(density(BW2, bw = "SJ-dpi"), col = 1, lty = 1, lwd = 2)
lines(density(BW4, bw = "SJ-dpi"), col = 1, lty = 3, lwd = 2)
lines(density(BW6, bw = "SJ-dpi"), col = 1, lty = 5, lwd = 2)
### Copy plot to file
dev.copy(png, 'Normal_SS_25.png') dev.off()
```