# Nonparametric Empirical Likelihood Density Functionals Estimation and Applications

By

Ningning Wang

Bachelor of Science
Yantai Normal University
Yantai, CHINA
264300
July,2001

Master of Science
Huazhong University Science and Technology
Wuhan, CHINA
43500
July,2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2013

Nonparametric Empirical Likelihood Density Functionals Estimation and Applications

Dissertation Approved:

Ibrahim Ahmad

Thesis Adviser

Joshua Habiger

Lan Zhu

Lisa A. Mantini

**ABSTRACT**

Name: Ningning Wang

Date of Degree: DECEMBER,2013

Title of Study: NONPARAMETRIC EMPIRICAL DENSITY FUNCTIONAL ESTIMATION AND APPLICATIONS

Major Field: STATISTICS

Chapter 2 of this dissertation presents a nonparametric empirical likelihood estimation of kernel density functionals (**ELKDFE**), which are constructed based on a kernel density functional estimation (**KDFE**) and the concepts of empirical likelihood. The work focuses on estimating the integration of square density function and a known function which has a derivative of order p, for $p \geqslant 0$. In many applications there may be extra information available to use, hence the concept of empirical likelihood becomes useful in providing a systematic approach for capturing the extra information. So **ELKDFE** reduces the MSE, especially when the sample size is small to moderate, and the difference of MSE between those two estimates decreases as the sample size increases.

Secondly, in Chapters 3 and 4, two new kernel estimators are proposed, **GCA** and **LCA**, and their rationales, properties, empirical likelihood versions, data-driven bandwidth selection, and applications are given as well. The bandwidth of the new approach is much tighter, catching the density's humps and valleys is more accurate. These estimates can be used for fixed and sequential sampling. The empirical likelihood (**EL**) versions of the **GCA** and **LCA** are provided and shown to have smaller **AMISE** than that of the non-**EL** estimation, and the difference of MISE tends to shrink as the sample size increases.

The **GCA** and **LGA** estimates are applied to regression using a local polynomial setting. It is shown that the regression estimators based on **GCA** and **LGA** have smaller bias and variance than standard kernel regression estimators.

An investigation of the properties of cumulative distribution function estimation based on **GCA** and **LGA** shows that the new estimators have smaller MSE and better performance than standard kernel **CDF** estimation.

**ACKNOWLEDGMENTS**

I would like to express my sincere gratitude to my advisor, Dr. Ahmad, for providing able guidance and encouragement as I completed the work presented here. I would also like to thank Dr. Zhu, Dr. Habiger, and Dr. Mantini for serving as members of my dissertation committee.

I would also thank my department for giving me opportunities to teach, study, and do research. I am also grateful to other faculty and staff members in our department, without whom completion of my thesis would not have been possible.

Special thanks goes to my parents for their encouragement and support throughout my academic life, and to my brother, Penghua Wang, for his love, care, and encouragement, which allowed me to remain focused on my research.

I appreciate and thank all of my friends and colleagues for providing good suggestions and fruitful discussions.

Ningning Wang

[1]

[1] Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University

**TABLE OF CONTENTS**

## LIST OF FIGURES

# 1 Introduction and Literature Review

## 1.1 Introduction

Nonparametric density estimation has been widely used with an array of new tools for statistical analysis. The main advantage of this approach is that it allows the exploration of large amounts of data without making specific distributional assumptions. This approach is in contrast to parametric estimation, in which it is assumed that the density comes from a given family, and the parameters are estimated by various statistical methods. Nonparametric density estimation is currently found in many fields, such as economics, signal processing, and image processing and reconstruction. Early contributors to the theory of nonparametric estimation include Rosenblatt (1956) and Parzen (1962), and their methods are still the most commonly used approach up to today. Comprehensive descriptions of various approaches to nonparametric estimation have been provided by Silverman (1986), and Wand and Jones (1994) have depicted more recent developments. These researchers provide a though discussion of kernel estimation, including details about the assumptions of kernel weight, estimator properties such as bias and variance, and guidelines for choosing the smoothness parameter bandwidth $h$. Empirical likelihood based on kernel density estimation (**ELKDE**) was introduced by Chen (1997), who showed that **ELKDE** reduces **MSE** and variance.

Empirical likelihood was first introduced by Owen (1988, 1990) for constructing confidence regions or intervals. It has many useful properties: such as, automatic determination of the shape of confidence regions given the observed data set and a non-parametric version of Wilks's Theorem. For these reasons, empirical likelihood has found many applications, such as in smooth functions of means (DiCiccio et al., 1991), estimating equations (Qin and Lawless, 1994), non-parametric density and regression function estimation (Owen, 1988; Chen, 1996; Chen and Qin, 2000), quantile estimation (Chen and Hall, 1993), and empirical likelihood-based kernel estimation (Chen, 1997). Other useful sources of discussion about empirical likelihood include Owen (2001) and Chen and Keilegom (2009). In general, empirical likelihood combines the reliability of non-parametric methods with the effectiveness of the likelihood approach. The regions are invariant under transformations and often behave better than confidence regions based on asymptotic normality when the sample size is small, a characteristic we show prevailing in our research. Moreover, they are of natural shape and orientation since the regions are obtained by contouring a log likelihood ratio, and they often do not require the estimation of the variance, as the studentization is carried out internally via the optimization procedure. The empirical likelihood method is appealing not only in obtaining confidence regions, but in its unique attraction in parameter estimation and formulating goodness-of-fit tests. On the computational side, empirical likelihood involves maximizing non-parametric likelihood supported on the data subject to some constraints. Owen (1988) showed that empirical likelihood regions for mean (univariate and multivariate) are always convex, so there is a unique solution for $p_i$, where $p_i$ is the probability weight of the observed data.

The aim of this chapter is to review the most important aspects of kernel density estimation and empirical likelihood based on kernel methods. In the remainder of this chapter, an introduction of kernel density estimation is given in Section 1.2; Statistical results for the standard kernel density estimate is in Section 1.2.1; Bandwidth selection of kernel density estimation is shown in Section 1.2.2; The kernel smoothing applications: regression and cumulative distribution function **CDF** estimation are presented in Section 1.2.3; An empirical likelihood introduction and review are given in Section 1.3; Empirical likelihood for univariate mean in Section 1.3.1; Empirical likelihood-based kernel density estimation is given in Section1.3.2; Analysis of error criteria is given in Section 1.4; New kernel density estimators are proposed in Section 1.5.

## 1.2 Kernel Density Estimation

The kernel estimation method is an important method in non-parametric density and distribution functions fitting. Suppose $X_1, X_2, \cdots, X_n$ are a sample of independently and identically distributed random variable from some distribution with unknown density $f$. We are interested in estimating $f$. The kernel density estimate is

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

where $K$ is called the kernel, a bounded symmetric function satisfying $\int K(\mu)d\mu = 1$, $\int \mu K(\mu)d\mu = 0$, and $\int \mu^2 K(\mu)d\mu < \infty$, and $h$ is a positive number depending on n, usually called the bandwidth or window width and satisfies $h \to 0$ and $nh \to \infty$, as $n \to \infty$. Using the notation, $K_h(\mu) = h^{-1}K(\mu/h)$, the kernel density estimator (1.1) can be written as

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i). \tag{1.2}$$

For further information, refer to Wand and Jones (1994), Silverman (1986) and Alez (2012).

### 1.2.1 Statistical Result of Kernel Density Estimation

In this section, some theoretical properties of the standard kernel density estimator are derived. The assumptions and conditions are defined as in the previous section. So for a fixed $h$

$$\text{Bias}(\hat{f}(x)) = \frac{h^2}{2} f''(x)\mu_2(K) + o(h^2) \tag{1.3}$$

$$\text{Var}(\hat{f}(x)) = \frac{1}{nh} R(K)f(x) + o(\frac{1}{nh}), \tag{1.4}$$

where $R(K) = \int K^2(\mu)d\mu$. From these two equations, we have

$$\text{MSE}(\hat{f}(x)) = \frac{1}{nh}R(K)f(x) + \frac{h^4}{4} f''(x)^2\mu_2^2(K) + o(h^4 + \frac{1}{nh}). \tag{1.5}$$

The trade-off between bias and variance is controlled by MSE, when $h$ is decreasing , the Bias is decreasing but variance is increasing. So a small $h$ leads to a small Bias but large variance yields under smooth, and vice verse. As has already been pointed out, the smoothness of the estimate depends on the smoothing parameter $h$, and a closed-form expression can be obtained from minimizing the mean integrated square error (MISE)(1.12). We have

$$\text{MISE}(\hat{f}) = \frac{1}{nh}R(K) + \frac{h^4}{4} R(f''(x))\mu_2^2(K) + o(h^4 + \frac{1}{nh}). \tag{1.6}$$

Then the optimal bandwidth is achieved by minimizing AMISE (1.6)

$$h_{opt} = \left( \frac{R(K)}{n\mu_2^2(K)R(f'')} \right)^{1/5}.$$

Using this optimal bandwidth, we have

$$inf_{h>0}\text{MISE}(\hat{f}) = \frac{5}{4}[\mu_2^2(K)R^4(K)R(f'')]^{1/5}n^{-4/5}.$$

### 1.2.2 Bandwidth Selection

It is crucially important to select an appropriate bandwidth for the standard kernel density estimator. Since the early work on kernel methods emphasized asymptotic results, now determining an optimal $h$ has been the main research focus up to today. As **AMISE** contains the unknown function $R(f'')$, several "plug-in" procedures were proposed by estimating $R(f'')$ with $R(\hat{f}'')$ (see Scott and Terrell, 1987; Park and Marron, 1992). An automatic method for determining the optimal bandwidth is cross-validation (CV) which was first introduced by Rudemo (1982) and Bowman (1984). Scott and Terrell (1987) introduced biased cross-validation, which is considered a hybrid of cross-validation and plug-in, replacing an unknown value in AMISE with a cross-validation kernel estimator $\tilde{R}(f'')$. The recent kernel contrast method of Ahmad and Ran (2004) can be used for MISE minimization as well, but it is not really data adaptive. Moreover this method performs particularly well for regression, but not as well for density estimation. For more information about these methods, see the most exhaustive form comparison papers, by Jones et al. (1996) and Devroye et al. (1997) or the recent review paper by Heidenreich et al. (2013).

#### Cross-Validation Bandwidth Selection

Here we briefly introduce unbiased least square cross-validation, the idea of which is to consider the expansion of ISE in the following way

$$\text{ISE}(h) = \int \hat{f}(x)^2 dx - 2\int \hat{f}(x) f(x) dx + \int f^2(x) dx$$

Note that the last term is not dependent on $h$, so that we only need to consider the first two terms. The idea for choosing bandwidth is picking the one that minimizes

$$L(h) = \int \hat{f}(x)^2 dx - 2\int \hat{f}(x) f(x) dx$$

Consider the estimator

$$\mathrm{CV}(h) = \int \hat{f}(x)^2 dx - 2\frac{1}{n}\hat{f}_{-i}(X_i)$$

where

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

It is shown that $\mathrm{CV}(h)$ is the unbiased estimator of $\mathrm{MISE} - \int f^2(x)dx$. So the data-driven optimal bandwidth is

$$h_{CV} = arg\ min_h CV(h)$$

Biased cross-validation considers the asymptotic MISE, and its main idea is to replace the unknown quantity $R(f'')$ in equation (1.6) by cross-validation estimator

$$\begin{aligned}\tilde{R}(f'') =& R(\hat{f}'') - \frac{1}{nh^5}R(K'') \\ =& n^{-2}\sum_{i\neq j}(K'' * K'')(X_i - X_j).\end{aligned}$$

Then the biased cross-validation estimator (**BCV**) is given as

$$BCV(h) = \frac{R(K)}{nh} + \frac{h^4}{4}\mu_2^2(K)(f'').$$

So, the selected bandwidth is $h_{BCV} = argmin\ BCV(h)$.

### 1.2.3 Kernel Smoothing Applications: Regression and CDF Estimation

In this section, we describe nonparametric regression and **CDF** estimation based on standard kernel density estimation. There is a vast literature on flexible methods for estimating regression functions and **CDF**. The NW estimator proposed independently by Nadaraya (1964) and Watson (1964) is based on locally weighted averages. Another popular estimate is the integral kernel estimate proposed by Gasser and Miller (1979). An alternative method of smoothing, the locally weighted regression, appeared in the statistical literature by Stone (1977) and Cleveland (1979). This method is still widely used today. It estimates the regression function at a particular point by locally fitting $p$th degree polynomial to the data, via weighted least squares. The **CDF** estimation is obtained by integrating a kernel estimator of the density. There has recently been extensive work on the estimation by kernel method of probability densities and their derivatives; for a reference, see Wertz (1978) and Li and Racine (2007).

### 1.3 Empirical Likelihood

Empirical likelihood is a non-parametric method of inference based on a data-driven likelihood function. It allows the data analyst to use likelihood methods without assuming that the data come from a known family of distributions. The likelihood method is known to be efficient. For example, likelihood ratio tests have some good power properties. These tests can be modified to construct short confidence intervals or small confidence regions of the parameters. The empirical likelihood method combines reliability of the non-parametric methods and the flexibility and effectiveness of the likelihood approach. Now we will introduce the empirical likelihood.

**Definition** Let $X_1, X_2, \cdots, X_n$ be $i.i.d$ random variables with the distribution function $F$. The empirical cumulative distribution function (ECDF) of $X_1, X_2, \cdots, X_n$ is

$$F_n(x) = \frac{1}{n}\sum_{1}^{n} 1_{(X_i \leqslant x)},$$

where $1_{A(x)}$ represents the value 1 if the assertion $A(x)$ is true, and 0 otherwise.

**Definition** Assuming $X_1, X_2, \cdots, X_n$, are independent real random variable with common cumulative distribution function (CDF) $F$, the non-parametric likelihood of CDF of $F$ is

$$L(F) = \prod_{i=1}^{n}(F(X_i) - F(X_i-)),$$

where $F(x) = Pr(X \leqslant x)$ and $F(x-) = Pr(X < x)$, so $Pr(X = x) = F(x) - F(x-)$. Then for a CDF $F$, the ratios of the non-parametric likelihood for hypothesis tests and confidence intervals are defined in the following way,

$$R(F) = \frac{L(F)}{L(F_n)}.$$

Like parametric likelihood, suppose that we are interested in a parameter $\theta = T(F)$ for some functional T of the distribution. This $F$ is a member of a set $\mathbf{F}$ of distributions. Define the profile likelihood ratio function,

$$R(\theta) = sup\left\{R(F)|T(F) = \theta, F \in \mathbf{F}\right\}.$$

Empirical likelihood hypothesis tests reject $H_0 : T(F_0) = \theta_0$, when $R(\theta_0) < r_0$ for some threshold value $r_0$. Empirical likelihood confidence regions are of the form

$$\left\{\theta|R(\theta) \geqslant r_0\right\},$$

where threshold $r_0$ may be chosen using an empirical likelihood theorem (ELT) 1.3.1, a non-parametric analogue of Wilk's Theorem.

**Theorem 1.3.1 (ELT)** *Let $X_1, X_2, \cdots, X_n$ be independent random variables with common distribution $F_0$. Let $\mu_0 = E(X_i)$, and suppose that $0 < Var(X_i) < \infty$. Then $-2log(R(\mu_0))$ converges in distribution to $\chi^2_{(1)}$ as $n \to \infty$.*

First, the chi-squared limiting distribution is the same as the typically found for parametric likelihood models with one parameter, which is Wilk's Theorem. Second, it does not assume that $X_i's$ are bounded random variables. It is only required to have a bounded variance, which constrains how fast the sample maximum and minimum can grow as $n$ increases.

### 1.3.1 EL for Univariate Mean

To test whether $\mu = \mu_0$, we need to compute $R(\mu_0)$ and choose threshold value $r_0$ by Theorem 1.3.1. Then reject the value $\mu_0$ at the $\alpha$ level, when $-2logR(\mu_0) > \chi^{2,1-\alpha}_{(1)}$. Empirical likelihood determines the $p_i$ by maximizing the empirical likelihood ratio function $\prod_{i=1}^n np_i$ or $\sum_{i=1}^n log(np_i)$ subject to $\sum_{i=1}^n p_i(X_i - \mu_0) = 0$, $p_i > 0$, and $\sum_{i=1}^n p_i = 1$. The objective function $\sum_{i=1}^n log(np_i)$ is strictly concave on a convex set of weight vectors. So there exists a unique global maximum in the domain.

We may proceed using the Lagrange multiplier to find $p_i's$. Write

$$G = \sum_{i=1}^n log(np_i) - n\lambda \sum_{i=1}^n p_i(X_i - \mu_0) + \gamma(\sum_{i=1}^n p_i - 1)$$

Setting to zero the partial derivative of $G$ with respect to $p_i$ gives

$$\frac{\partial G}{\partial p_i} = \frac{1}{p_i} - n\lambda(X_i - \mu_0) + \gamma = 0.$$

Therefore,

$$p_i = \frac{1}{n}\frac{1}{1 + \lambda(X_i - \mu_0)}. \tag{1.7}$$

The value of $\lambda$ can be found by numerical search method, (for example, Newton's method or Brent's method), based on the equation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(X_i-\mu_0)}{1+\lambda(X_i-\mu_0)}=0.$$

### 1.3.2 Empirical Likelihood-Based Kernel Density Estimation(ELKDE)

In some statistical applications, additional information about $f$ is available: for exmaple, the mean or variance of a distribution is known. This additional information usually can be expressed as

$$E_X g_l(X)=0 \qquad (l=1,2,\cdots,q). \tag{1.8}$$

where $g_l(X)$ are some known real functions. **ELKDE** (Chen, 1997) uses empirical likelihood in conjunction with the kernel method to provide a systematic approach for capturing the extra information. Suppose the extra information can be formulated as equation (1.8), then **ELKDE** can be constructed by replacing $n^{-1}$ in equation (1.2) with the empirical likelihood $p_i$ under extra information (1.8). Specifically $p_i$ can be determined by maximizing a multinomial $\prod_1^n np_i$ subject to

$$\sum p_i=1 \qquad \text{and} \qquad \sum p_i g_l(X_i)=0 \qquad\qquad (l=1,2,\cdots,q).$$

Let $\lambda_1,\lambda_2,\cdots,\lambda_q$ be Lagrange multipliers corresponding to the $q$ constraints. Define $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_q)^T$ and $g(X_i)=\{g_1(X_i),g_2(X_i),\cdots,g_q(X_i)\}$. Then the weight $p_i$ are

$$p_i=n^{-1}\left\{1+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\right\}^{-1} \quad (i=1,2,\cdots,n), \tag{1.9}$$

where $\lambda$ is the solution of

$$\sum_{i=1}^{n} \frac{g_l(X_i)}{1+\boldsymbol{\lambda}^T \mathbf{g}(X_i)} = 0 \quad (l = 1, 2, \cdots, q).$$

**ELKDE** is obtained by replacing $n^{-1}$ in KDE (1.2) with the $p_i$ at equation (1.9) , so

$$\hat{f}_{el}(x) = \frac{1}{h} \sum_{i=1}^{n} p_i K_h(x - X_i) \tag{1.10}$$

It is shown that **ELKDE** has smaller variance and MSE than those of **KDE**. This is reasonable because **ELKDE** achieves a smaller variance by using unequal weights, which offers more flexibility than **KDE** using equal weight $n^{-1}$. In this Chapter, the **ELKDE** method is applied to estimate density functional, and it is shown that **ELKDE** has better performance than that of **KDE** in theoretical and simulation results .

## 1.4 Analysis of the Error Criteria

There are many criteria to evaluate $\hat{f}(t)$ as an estimator of $f(t)$ , such as the bias, square error, and distance error.

1. Bias

   Bias is the difference between an estimator's expectation and the true value of the parameter being estimated.

   $$\text{Bias}[f(x)] = \text{E}\{\hat{f}(x) - f(x)\}$$

2. Mean Squared Error (MSE) , Mean Integrated Square Error (MISE) and Integrated Squared Error (ISE)

   Mean squared error is the expected value of the square of the difference between the estimator and the true value of the parameter being estimated at a single point.

$$\text{MSE}[\hat{f}(x)] = \text{E}\left\{\hat{f}(x) - f(x)\right\}^2 \tag{1.11}$$

Mean integrated squared error is the expected value of the square of the difference between the estimator and the true value of the parameter being estimated at whole real line.

$$\text{MISE}[\hat{f}(x)] = \text{E}\int\left\{\hat{f}(x) - f(x)\right\}^2 dx \tag{1.12}$$

Integrated squared error globally measures the distance between the estimator and the true value of the parameter being estimated.

$$\text{ISE}[\hat{f}(x)] = \int\{\hat{f}(x) - f(x)\}^2 dx \tag{1.13}$$

3. Mean Distance Error (MDE) and Mean Integrated Distance Error(MIDE)
   The mean distance of using $\hat{f}(x)$ to estimate $f(x)$ is given by

$$\text{MDE}[\hat{f}(x)] = \text{E}|\hat{f}(x) - f(x)|.$$

   The MIDE is

$$\text{MIDE}[\hat{f}(x)] = \text{E}\int|\hat{f}(x) - f(x)|dx.$$

   The MSDE is

$$\text{MSDE}[\hat{f}(x)] = \text{E}\sup_x|\hat{f}(x) - f(x)|.$$

The Bias, ISE and MISE are discussed in this dissertation. For more details on the MDE, MIDE and MSDE see Devroye and Lugosi (1996, 2001), Ahmad (2002), and Ahmad and Ran (2004).

### 1.5 Proposed New Kernel Density Estimates

Standard kernel density estimation is still one of most active areas of research in nonparametric statistics. But there are drawbacks to this method, such as choice of smoothing parameter(s), and difficulty in catching humps and valleys. For example, if the small bandwidth $h$ is chosen, then the average kernel weight $K(\frac{x-X_i}{h})$ for some fixed $x$ is only based on relatively few observations, not for all observations. So the estimate pays too much attention to the local data and does not allow for variation across the sample. But if the bandwidth is too large, then the estimate is too smooth and cannot catch details such as humps and valleys.

In view of the flaws of standard kernel smoothing, two new kernel density estimators **GCA** and **LCA** and their empirical likelihood versions are proposed in Chapter 3. Suppose the $X_1, \cdots, X_n$ are independently and identically distributed from the unknown distribution $f$, and these bandwidth of these two estimators is $ih$ instead of $h$. So the bandwidth has two parts: one is the smoothing parameter $h$, and the other is the scale coefficient i. When choosing smaller $h$, for the fixed $x$, the value of $K(\frac{x-X_i}{h})$ in standard kernel estimate is almost zero when $X_i$ is far away from $x$. In this situation, standard kernel density estimation is more "wiggy". But in the methods proposed in Chapter 3, the ratio of $\frac{x-X_i}{h}$ divided by coefficient $i$, the value of $K(\frac{x-X_i}{ih})$ is not dependent on the distance between $x$ and observation $X_i$, so the average of $K(\frac{x-X_i}{ih})$ at each point $x$ is dependent on the entire sample data instead of just the local data (the data close to the $x$) in the standard **KDE**. These methods of choosing difference bandwidth do very well on balance between the local data and the whole sample data. Simulation study show that the new estimators can catch humps and valleys better that the standard **KDE**. The empirical likelihood version of **GCA** and **LCA** show that when the sample size is small to moderate, these methods are significantly better at catching humps and valleys than those of **GCA** and **LCA**. And when the sample size increases, the advantages shrink.

The applications of the proposed methods in the regression and **CDF** estimation of are developed in Chapter 4.

# 2 Nonparametric Empirical Likelihood Estimation of Density Functional

## 2.1 Introduction of Density Functional Estimation

Immediately following the introduction of the kernel density estimation by Fix and Hodges (1951) and the study of its functional properties by Rosenblatt (1956), Parzen (1962),Watson and Leadbetter (1963) and Nadaraya (1964), many authors saw the potential of using a kernel density methodology to study inferential problems. The methodology was subsequently used in estimating regression (Nadaraya, 1964; Watson, 1964), testing goodness of fit (Bickel and Rosenblatt, 1973), testing independence (Rosenblatt, 1975; Ahmad and Li, 1997a), testing symmetry (Ahmad and Li, 1997b), and testing positive aging (Ahmad, 2000). Many books have been written on the subject. For univariate density estimation, more recent work has been conducted by Wand and Jones (1994), Bowman and Azzalini (1997), Simonoff (1996), Alez (2012) and Pons (2011), and in the multivariate case by Scott (1992) and Klemelä (2009). For econometric application, see Pagan and Ullah (1999), and Li and Racine (2007). Finally, for regression applications, see Hardle (1990).

Of particular interest to researchers is the subject of estimating density functionals of the type $\int \gamma(x)f(x)^2dx = I(\gamma; f)$, where $\gamma(x)$ is some known continuous function that has the $p^{th}$ derivative, for $p \geqslant 0$ . For $\gamma(x) = 1$ or $x$ , Ahmad and Amezziane (2011) studied the basic kernel estimates properties of $I(1; f)$ and $I(x; f)$. These special cases are the location $(I(x; f))$ and scale $(I(1; f))$ parameters. Applications of estimates of $I(\gamma; f)$ are found in several areas. Among them many authors used variations of $I(\gamma; f)$ and the estimates in evaluating the power of the nonparametric tests (Aubuchon and Hettmansperger, 1984) or obtaining estimates of the smoothing parameter (Sheather and Jones, 1991; Jones et al., 1991; Birge and Massart, 1995).

## 2.2 Methodology

In this work, $I(\gamma, f)$ is estimated by the kernel density functional estimation (**KDFE**) as follows:

$$\hat{I}(\gamma; f) = \frac{2}{n(n-1)h} \sum_{i<j} \left( \frac{\gamma(X_i) + \gamma(X_j)}{2} \right) K\left( \frac{X_i - X_j}{h} \right) \tag{2.1}$$

Moreover, in many applications there exists extra information which can be represented by

$$\mathrm{E}(g_l(x)) = 0, l = 1, \cdots, L, \tag{2.2}$$

where $g_l(x)$ are some known real-valued functions. Using the concept of empirical likelihood (see Owen, 2001), in conjunction with the kernel method, provides a systematic approach for capturing the extra data information. The estimator (2.1) assigns an equal probability weight $1/(n(n+1))$ to each data pair. However, if the extra data information is available as (2.2), then empirical likelihood based on kernel estimation is constructed by replacing $1/(n(n+1))$ in (2.1) with empirical likelihood weights $p_i p_j$, where $p_i's$ are the solution of the multinomial likelihood $\prod_{i=1}^n p_i$ subject to:

$$\sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i g_l(X_i) = 0, l = 1, \cdots, L.$$

Let $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_L)'$ be the Lagrange multiplier and $\mathbf{g}(X_i) = (g_1(X_i), \cdots, g_L(X_i))'$. Then

$$p_i = \frac{1}{n}\{1 + \boldsymbol{\lambda}'\mathbf{g}(x_i)\}^{-1}, i = 1, \cdots, n, \tag{2.3}$$

where $\boldsymbol{\lambda}$ is the solution to

$$\sum_{i=1}^{n} \frac{g_l(x_i)}{1 + \boldsymbol{\lambda}'\mathbf{g}_l(x_i)} = 0, l = 1, \cdots, L.$$

Hence, the empirical likelihood based on kernel density functional estimation (**ELKDFE**) of $I(\gamma, f)$ is

$$\hat{I}_{el}(\gamma; f) = \frac{1}{h} \sum_{i \neq j} p_i p_j \left( \frac{\gamma(X_i) + \gamma(X_j)}{2} \right) K \left( \frac{X_i - X_j}{h} \right), \tag{2.4}$$

where $p_i$ is given in (2.3).

## 2.3 Statistical Result

In order to study the mean squared error (MSE) and expectation of **ELKDFE** in comparison to those of the **KDFE**, we need the following customary conditions on $K, h$ and $f$:

1. The density function $f$ has $p^{th}$ continuous derivative, where $p$ is an integer and $p > 1$.

2. The kernel $K(\cdot)$ is a symmetric probability density with mean $\mu_k = 0$ and variance $\mu_2(K) = \sigma_k^2 < \infty$.

3. The sequence of constant $\{h_n\}$, $h_n \equiv h$ is such that $h \to 0$ and $nh \to \infty$ as $n \to \infty$

In this section, the expectation and MSE of **ELKDFE** and **KDFE** are investigated, and it is shown that the Bias and the MSE of **ELKDFE** are both smaller than those of **KDFE**. The following is the main result.

**Theorem 2.3.1**

$$\mathrm{E}(\hat{I}_{el}) = \mathrm{E}(\hat{I}) - \frac{1}{n}\int \mathbf{g}^T(y)\Sigma^{-1}\mathbf{g}(y)\gamma(y)f^2(y)dy + o(n^{-1}), \tag{2.5}$$

*and*

$$\mathrm{MSE}(\hat{I}_{el}) = \mathrm{MSE}(\hat{I}) - \frac{2}{n}\int \gamma(y)f^2(y)dy \int \mathbf{g}^T(y)\Sigma^{-1}\mathbf{g}(y)\gamma(y)f^2(y)dy + o(n^{-1}), \tag{2.6}$$

*where* $\mathbf{g}(\cdot)$ *is the vector of extra information(eg, mean, variance·) and* $\Sigma = \mathbf{cov}(g_i, g_j)$. *In addition,*

$$\mathrm{E}(\hat{I}) = \int \gamma(y)f^2(y)dy + \mu_2(K)h^2C_1 + o(h^2) \tag{2.7}$$

*and*

$$\mathrm{MSE}(\hat{I}) = \frac{1}{n^2h}\int \gamma^2(y)f^2(y)dy \int K^2(\mu)d\mu + \mu_2^2(K)h^4C_1^2 + \frac{4}{n}C_2, \tag{2.8}$$

*with*

$$C_1 = \frac{1}{2}\int \gamma(y)f''(y)f(y)dy + \frac{1}{4}\int \gamma''(y)f^2(y)dy + \frac{1}{2}\int \gamma'(y)f'(y)f(y)dy \tag{2.9}$$

*and*

$$C_2 = \int \gamma^2(y)f^3(y)dy - \left[\int \gamma(y)f^2(y)dy\right]^2, \tag{2.10}$$

*both assumed finite.*

This theorem shows the difference of the expectation and MSE between **ELKDFE** and **KDFE**. Also, the difference of MSE between these two estimators is not dependent on the bandwidth $h$, so the optimal bandwidth for both methods can be obtained from equation (2.8). Thus it is given by

$$h_{opt} = n^{-2/5} \left\{ \frac{\int \gamma^2(y) f^2(y) dy \int K^2(\mu) d\mu}{4C_1 \mu_2^2(K)} \right\}^{1/5}. \tag{2.11}$$

## 2.4 Applications

Define the location-scale family distributions as

$$f(x;\mu,\sigma) = \frac{1}{\sigma} f_0 \left( \frac{x-\mu}{\sigma} \right), \tag{2.12}$$

where $f_0$ is bounded, and almost everywhere continuous probability density function (**pdf**). Consider the following functional,

$$\begin{aligned} I(x; f(\cdot;\mu,\sigma)) = \int x f^2(x;\mu,\sigma) dx &= \int \left( \frac{y\sigma + \mu}{\sigma} \right) f_0^2(y) dy \\ &= I(x; f_0) + \frac{\mu}{\sigma} I(1; f_0). \end{aligned}$$

Which leads to

$$\mu = \sigma \frac{I(x; f(\cdot;\mu,\sigma))}{I(1; f_0)} - \sigma \frac{I(x; f_0)}{I(1; f_0)}, \tag{2.13}$$

and

$$\sigma = \frac{I(1; f(\cdot;\mu,\sigma))}{I(1; f_0)}. \tag{2.14}$$

Thus estimating $\mu$ and $\sigma$ is reduced to estimating $I(\gamma; f)$, where $\gamma(x) = x$ for $\mu$ and $\gamma(x) = 1$ for $\sigma$, provided the $I(\gamma; f_0)$ known. Suppose we wish to test $H_{01}$ :

$\sigma = \sigma_0$ or $H_{02} : \mu = \mu_0$, then we have $I(1; f(x, \mu, \sigma)) = I(1; f_0)\sigma$ under $H_{01}$ or $\mu = c_1 I(x; f(x; \mu, \sigma)) + c_2$ under $H_{02}$, where $c_1 = \sigma / I(1, f_0)$ and $c_2 = -\frac{\sigma I(x;f_0)}{I(1;f_0)}$. Hence testing $H_{01}$ or $H_{02}$ is equivalent to testing $H_{01}^* : I(1; f(x; \mu, \sigma) = I(1; f_0)$ or $H_{02}^* : I(x; f(x; \mu, \sigma) = I(x; f_0) + \frac{\mu_0}{\sigma} I(1; f_0)$ respectively. To test two or more samples $H_{01} : \sigma_1 = \cdots = \sigma_k$ or $H_{02} : \mu_1 = \cdots = \mu_k$, we only need to test $H_{01} : I_1(1; f_1(x; \mu, \sigma)) = \cdots = I_k(1; f_k(x; \mu, \sigma))$ or $H_{02} : I_1(x; f_1(x; \mu, \sigma)) = \cdots = I_k(x; f_1(x; \mu, \sigma))$. Also notice that $I(\gamma; f_0)$ is not required in those cases. In this work, by using extra information $\mathbf{g}$, it shows that both the Bias and MSE of the **ELKDFE** are distinctly smaller than those of the **KDFE**.

### 2.4.1 Location Parameter

By equation (2.13), estimating $\mu$ is reduced to $I(x; f)$. If the extra information $\mathbf{g}$ is the location function $\mathbf{g}(y) = \mathbf{g_0}(y - \mu)$, then Theorem 2.3.1 can be expressed as follows:

$$\text{Bias}(\hat{I}_{el}) = \text{Bias}(\hat{I}) - \frac{\mu}{n} \int \mathbf{g}_0^T(y) \Sigma^{-1} \mathbf{g}_0(y) f_0^2(y; \mu) dy + o(h^2), \tag{2.15}$$

$$\text{MSE}(\hat{I}_{el}) = \text{MSE}(\hat{I}) - \frac{4}{n} \mu^2 \int f_0^2(y; \mu) dy \int \mathbf{g_0}^T(y) \Sigma^{-1} \mathbf{g_0}(y) dy + o(n^{-1}), \tag{2.16}$$

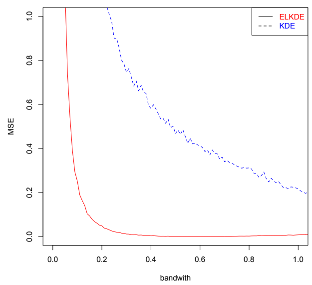
where

$$\text{Bias}(\hat{I}) = \frac{\mu h^2}{2} \mu_2(K) \int f_0''(y; \mu) f_0(y; \mu) dy + o(h^2) \tag{2.17}$$

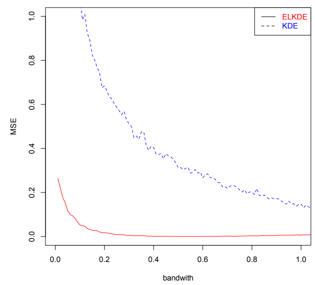
Equation (2.16) shows that the empirical likelihood based on the kernel method has reduced the MSE, and this reduction decreases when sample size $n$ increases.

#### Simulation

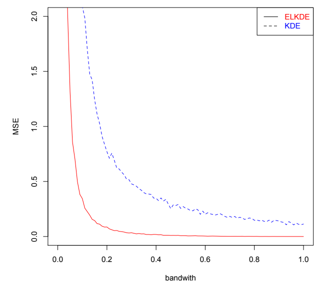
Generate the data from $N(2, 1)$, $Laplace(2, 1)$ and $Cauchy(2, 1)$ for the location parameter study with the sample sizes 50 and 100, with 1000 replications. Figure 1 shows that the MSE of the **ELKDFE** is smaller than that of the **KDFE**, and the difference in MSE decreases as the sample size increases. **ELKDFE** performs better for small and moderate sample sizes, and this advantage shrinks when the sample size becomes large. The MSE of **ELKDFE** is close to zero when $h$ is increasing. When choosing the proper bandwidth $h$, MSE of **ELKDFE** is close to zero. Figure 2 shows that the **ELKDFE** is closer to the true value $\mu = 2$ than that of **KDFE**. From these three cases, the **ELKDFE** not only reduces the MSE, but also provides bias correction with a proper bandwidth, which is shown in Theorem 2.3.1.



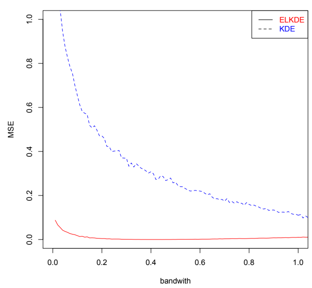

**(a)** N(2,1) n=15



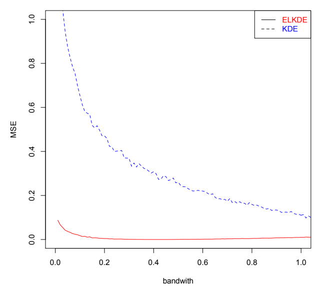

**(b)** Laplace(2,1) n=15



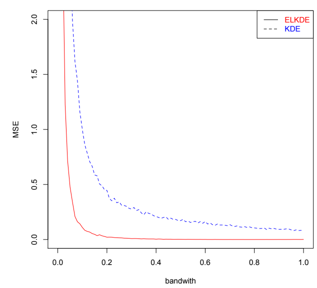

**(c)** Cauchy(2,1) n=15



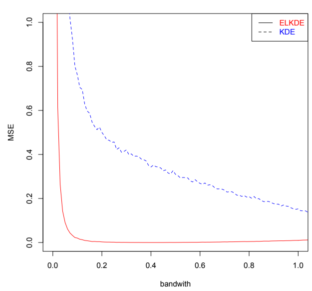

**(d)** N(2,1) n=25



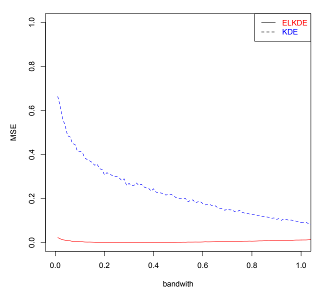

**(e)** Laplace(2,1) n=25



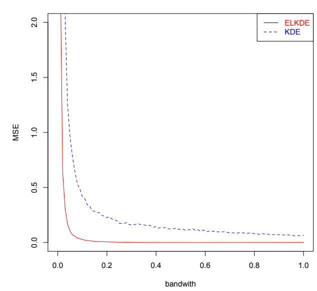

**(f)** Cauchy(2,1) n=25



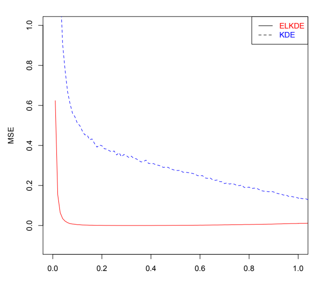

**(g)** N(2,1) n=50



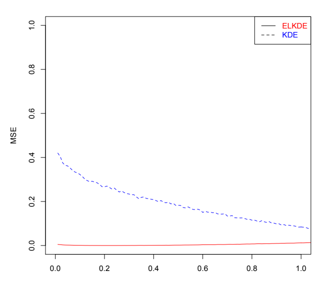

**(h)** Laplace(2,1) n=50



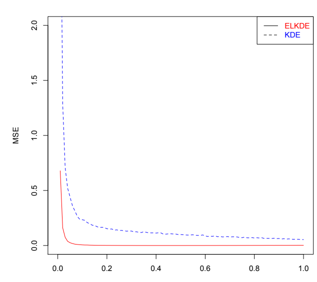

**(i)** Cauchy(2,1) n=50




**(j)** N(2,1) n=100

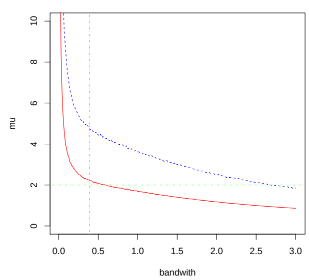



**(k)** Laplace(2,1) n=100

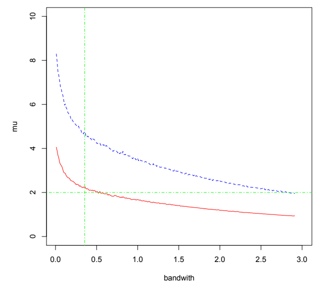



**(l)** Cauchy(2,1) n=100

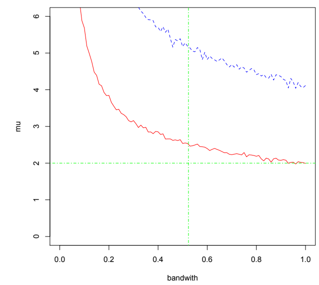
**Figure 1:** MSE of $\hat{\mu}$ from different distributions and sample size



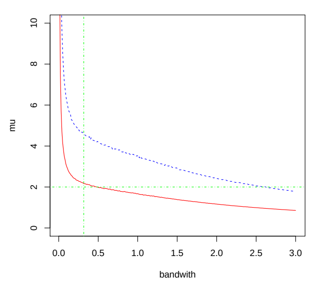

**(a)** N(2,1) n=15



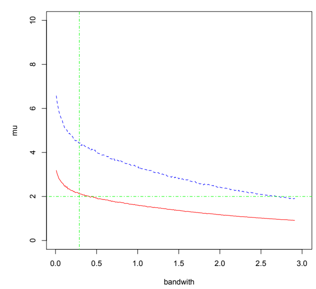

**(b)** Laplace(2,1) n=15



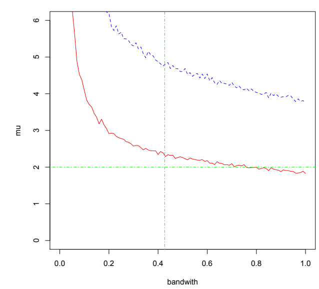

**(c)** Cauchy(2,1) n=15



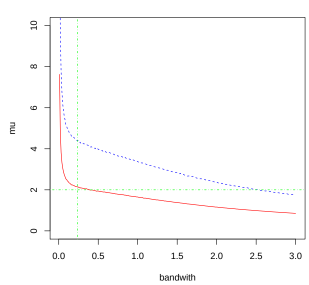

**(d)** N(2,1) n=25



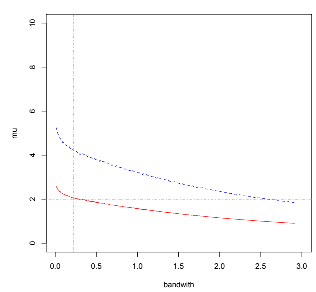

**(e)** Laplace(2,1) n=25



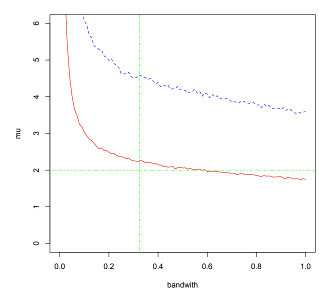

**(f)** Cauchy(2,1) n=25



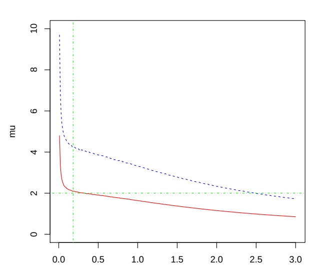

**(g)** N(2,1) n=50



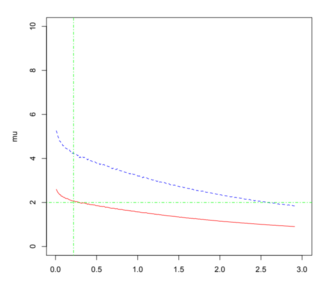

**(h)** Laplace(2,1) n=50



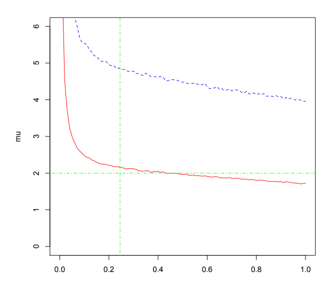

**(i)** Cauchy(2,1) n=50




**(j)** N(2,1) n=100




**(k)** Laplace(2,1) n=100




**(l)** Cauchy(2,1) n=100

**Figure 2:** Estimated $\hat{\mu}$ from different distributions and sample sizes. Red line is the **ELKDFE**; blue line is the **KDFE**; horizontal green line is $\mu = 2$; vertical green line is optimal bandwidth based on equation (2.11)

### 2.4.2 Scale Parameter

From equation (2.14), estimating scale parameter $\sigma$ is equivalent to $I(1; f)$. If extra information $\mathbf{g}(x)$ is given, then Theorem 2.3.1 can be expressed as follows:

$$\text{Bias}(\hat{I}_{el}) = \text{Bias}(\hat{I}) - \frac{1}{n}\int \mathbf{g}^T(y)\Sigma^{-1}\mathbf{g}(y)f^2(y;\sigma)dy + o(h^2) \tag{2.18}$$

$$\text{MSE}(\hat{I}_{el}) = \text{MSE}(\hat{I}) - \frac{4}{n}\int f^2(y;\sigma)\int \mathbf{g}^T(y)\Sigma^{-1}\mathbf{g}(y)f^2(y;\sigma)dy + o(n^{-1}) \tag{2.19}$$

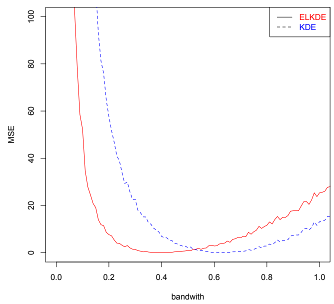
where

$$\text{Bias}(\hat{I}) = \frac{h^2}{2}\mu_2(K)\int f''(y;\mu)f(y;\mu)dy + o(h^2), \tag{2.20}$$

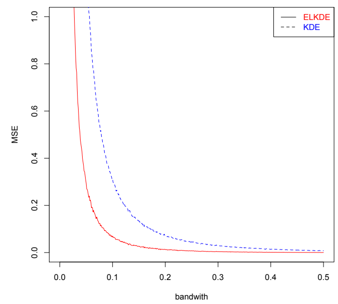
Equations (2.18) and (2.19) show that the **ELKDFE** not only reduces MSE but also reduces Bias. The difference decreases as the sample size $n$ increases.

**Simulation**

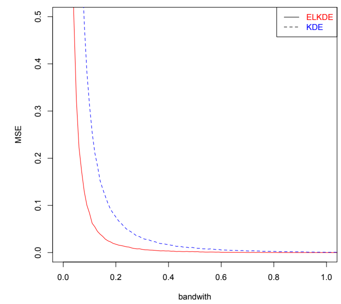
Generate the data from $N(0,1)$, $Laplace(0,1)$ and $Cauchy(0,1)$ for the scale parameter study with the sample sizes 15, 25, 50, and 100, with 1000 replications . Figure 3 shows that the **ELKDFE** has a smaller MSE, and that difference decreases as the sample size increases. The **ELKDFE** works for small and moderate sample sizes, and this advantage shrinks when the sample sizes become large. The estimated difference decreases when the sample size increases, which is shown in Theorem 2.3.1.



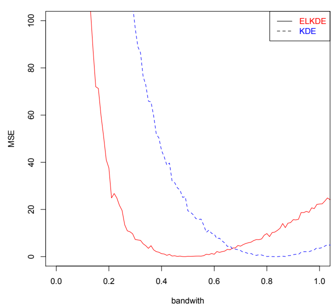

**(a)** Normal n=15



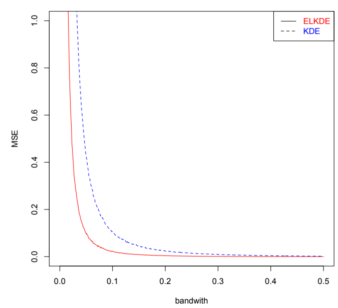

**(b)** Laplace n=15



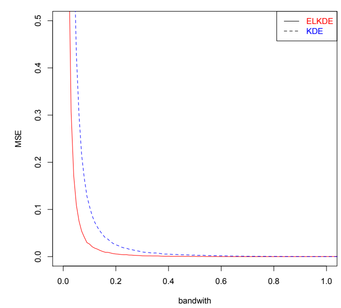

**(c)** Cauchy n=15



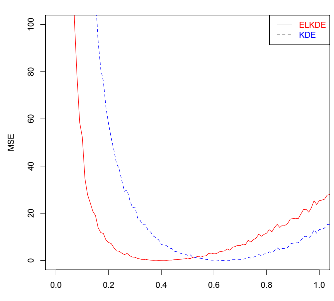

**(d)** Normal n=25



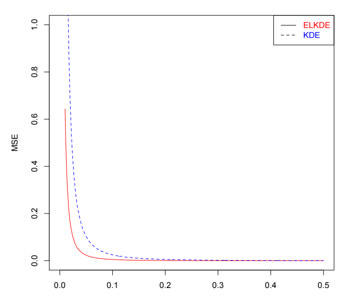

**(e)** Laplace n=25



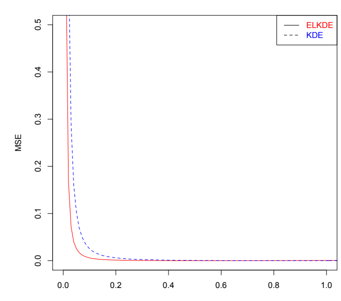

**(f)** Cauchy n=25



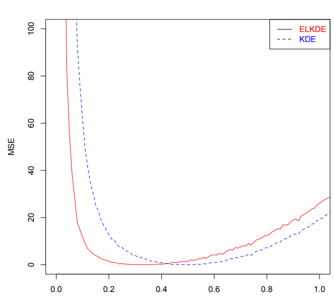

**(g)** Normal n=50



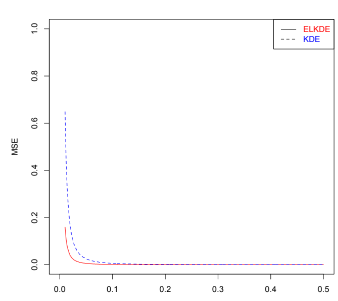

**(h)** Laplace n=50



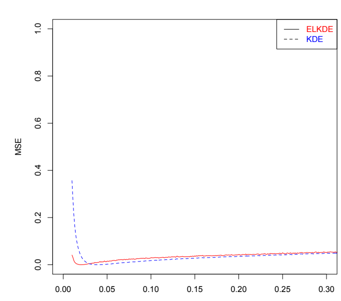

**(i)** Cauchy n=50




**(j)** Normal n=100

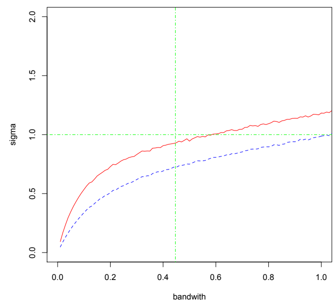



**(k)** Laplace n=100

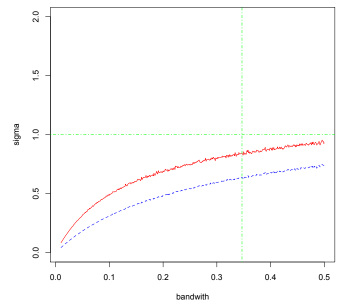



**(l)** Cauchy n=100

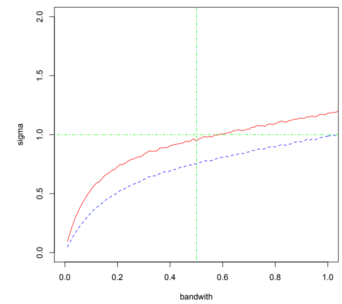
**Figure 3:** MSE of $\hat{\sigma}$ from different distributions and sample size



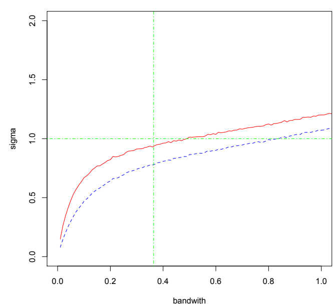

**(a)** Normal n=15



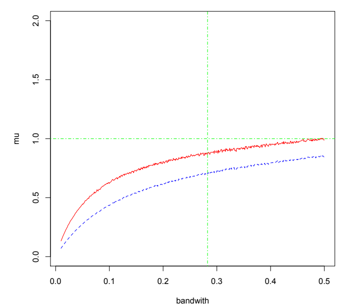

**(b)** Laplace n=15



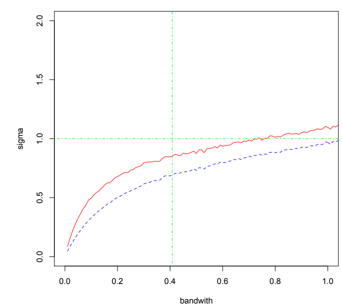

**(c)** Cauchy n=15



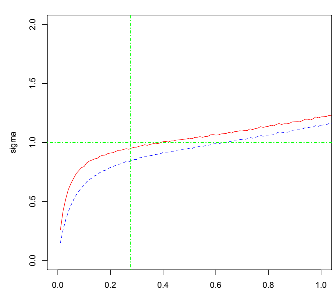

**(d)** Normal n=25



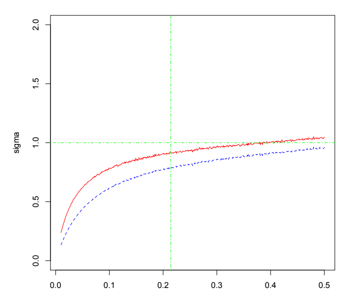

**(e)** Laplace n=25



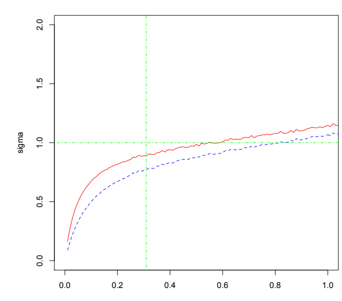

**(f)** Cauchy n=25



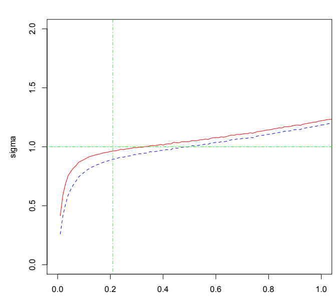

**(g)** Normal n=50



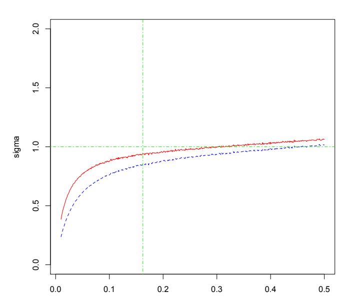

**(h)** Laplace n=50



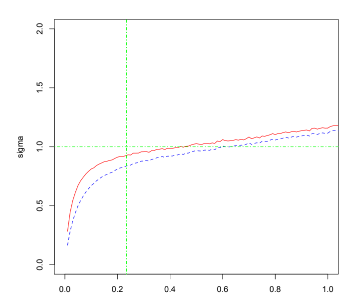

**(i)** Cauchy n=50




**(j)** Normal n=100




**(k)** Laplace n=100




**(l)** Cauchy n=100

**Figure 4:** Estimated $\hat{\sigma}$ from different distributions and sample size. Red line is the **ELKDFE**; blue line is the **KDFE**; horizontal green line is $\sigma = 1$; vertical green line is optimal bandwidth based on equation (2.11).

## 2.5 Appendix: proof the Theorem 2.3.1

**Proof** Assume function $\gamma(x)$ has $p^{th}$ derivative, $\gamma^{(p)}(x) \neq 0$ and $\gamma^{(p+i)}(x) = 0$ for $i = 1, \cdots, k$. First show the equation (2.7) and (2.8).

$$
\begin{aligned}
\mathrm{E}(\hat{I}) =& \frac{2}{n(n-1)h} \sum_{i<j} \mathrm{E}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2}\right) K\left(\frac{X_i - X_j}{h}\right) \\
=& \frac{2}{n(n-1)h} \sum_{i<j} \int \left(\frac{\gamma(x)+\gamma(y)}{2}\right) K\left(\frac{x-y}{h}\right) f(x)f(y)dxdy \\
=& \int \frac{1}{h}\left(\frac{\gamma(x)+\gamma(y)}{2}\right) K(\mu) f(x)f(y)dxdy \\
=& \int \left(\frac{\gamma(y+\mu h)+\gamma(y)}{2}\right) K(\mu) f(y+\mu h)f(y)d\mu dy \\
=& \int \left(\frac{\gamma(y)+\mu h\gamma'(y)+\frac{(\mu h)^2}{2}\gamma''(y)+\gamma(y)}{2}\right) K(\mu) f(y+\mu h)f(y)d\mu dy \\
=& \int \left(\gamma(y)+\frac{\mu h}{2}\gamma'(y)+(\frac{\mu h}{2})^2\gamma''(y)\right) K(\mu) \\
& \{f(y)+\mu h f'(y)+\frac{(\mu h)^2}{2}f''(y)\} f(y)d\mu dy \\
=& \int \gamma(y)f^2(y)dy + \frac{\mu_2(K)h^2}{2}\int \gamma(y)f''(y)f(y)dy \\
& + \frac{\mu_2(K)h^2}{4}\int \gamma''(y)f^2(y)dy + \frac{\mu_2(K)h^2}{2}\int \gamma(y)f'(y)f(y)dy + o(h^2).
\end{aligned}
\tag{2.21}
$$

Now we will work on the $\mathrm{Var}(\hat{I})$. It is not difficult to show that

$$\mathrm{Var}(\hat{I}) = \mathrm{Var}\frac{1}{n(n-1)h}\sum_{i\neq j}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2}\right)K\left(\frac{X_i-X_j}{h}\right)$$
$$= \frac{1}{n^2(n-1)^2}\mathrm{Var}\{\sum_{i\neq j}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\}$$
$$= \frac{2}{n(n-1)}\mathrm{Var}\{\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right)K\left(\frac{X_1-X_2}{h}\right)\}$$
$$+ \frac{4(n-2)}{n(n-1)}\mathrm{Cov}\{\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right)K\left(\frac{X_1-X_2}{h}\right),$$
$$\left(\frac{\gamma(X_2)+\gamma(X_3)}{2h}\right)K\left(\frac{X_2-X_3}{h}\right)\}$$
$$= I_1 + I_2.$$

$$\mathrm{E}\{\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right)K\left(\frac{X_1-X_2}{h}\right)\}^2$$
$$= \iint\left(\frac{\gamma(x)+\gamma(y)}{2h}\right)^2 K\left(\frac{x-y}{h}\right)^2 f(x)f(y)dxdy$$
$$= \iint\left(\frac{\gamma(y+\mu h)+\gamma(y)}{2h}\right)^2 hK(\mu)^2 f(y+\mu h)f(y)d\mu dy$$
$$= \frac{1}{h}\iint\left(\gamma(y)+\frac{\mu h}{2}\gamma'(y)+(\frac{\mu h}{2})^2\gamma''(y)\right)^2 K(\mu)^2$$
$$\{f(y)+\mu h f'(y)+\frac{(\mu h)^2}{2}f''(y)\}f(y)d\mu dy$$
$$= \frac{1}{h}\int\gamma^2(y)f^2(y)dy\int K^2(\mu)d\mu$$
$$+ h\int K^2(\mu)\mu^2 d\mu C_1 + o(h),$$

where $C_1 = \frac{1}{4}\int\gamma''(y)f^2(y)dy + \frac{1}{2}\gamma^2(y)f''(y)f(y)dy + \int\gamma(y)\gamma'(y)f'(y)f(y)dy.$

Then

$$
\begin{aligned}
I_1 =&\frac{2}{n(n-1)}\mathrm{E}\{\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right)K\left(\frac{X_1-X_2}{h}\right)\} \\
-&\frac{2}{n(n-1)}[\mathrm{E}\{\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right)K\left(\frac{X_1-X_2}{h}\right)\}]^2 \\
=&\frac{2}{n(n-1)h}\int\gamma^2(y)f^2(y)dy\int K^2(\mu)d\mu+\frac{2C_1}{n(n-1)}\int K^2(\mu)\mu^2d\mu \\
-&\frac{2}{n(n-1)}\{\int\gamma(y)f^2(y)dy+\frac{\mu_2(K)h^2}{2}\int\gamma(y)f''(y)f(y)dy+ \\
&\frac{\mu_2(K)h^2}{4}\int\gamma''(y)f^2(y)dy+\frac{\mu_2(K)h^2}{2}\int\gamma(y)f'(y)f(y)dy\}^2+o(h^2) \\
=&\frac{2}{n^2h}\int\gamma^2(y)f^2(y)dy\int K^2(\mu)d\mu+o(n^{-2}),
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{E}\{\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right)K\left(\frac{X_1-X_2}{h}\right)\left(\frac{\gamma(X_2)+\gamma(X_3)}{2h}\right)K\left(\frac{X_2-X_3}{h}\right)\}\\
=&\iiint\left(\frac{\gamma(x)+\gamma(y)}{2h}\right)\left(\frac{\gamma(y)+\gamma(z)}{2h}\right)K\left(\frac{x-y}{h}\right)\\
&K\left(\frac{y-z}{h}\right)f(x)f(y)f(z)dxdydz\\
=&\iiint\left(\frac{\gamma(y+\mu h)+\gamma(y)}{2}\right)\left(\frac{\gamma(y)+\gamma(y-\nu h)}{2}\right)\\
&K(\mu)K(\nu)f(y+\mu h)f(y)f(y-\nu h)d\mu dyd\nu\\
=&\iiint\left(\frac{\gamma(y+\mu h)+\gamma(y)}{2}\right)\left(\frac{\gamma(y)+\gamma(y-\nu h)}{2}\right)\\
&K(\mu)K(\nu)f(y+\mu h)f(y)f(y-\nu h)d\mu dyd\nu\\
=&\iiint\left(\gamma(y)+\frac{\mu h}{2}\gamma'(y)+(\frac{\mu h}{2})^2\gamma''(y)\right)\left(\gamma(y)-\frac{\nu h}{2}\gamma'(y)+(\frac{\nu h}{2})^2\gamma''(y)\right)\\
&K(\mu)K(\nu)\{f(y)+\mu hf'(y)+\frac{(\mu h)^2}{2}f''(y)\}\\
&f(y)\{f(y)-\nu hf'(y)+\frac{(\nu h)^2}{2}f''(y)\}d\mu dyd\nu\\
=&\int\gamma^2(y)f^3(y)dy+\frac{\mu_2(K)h^2}{2}\int\gamma''(y)\gamma(y)f^2(y)dy\\
&+\mu_2(K)h^2\int\gamma^2(y)f''(y)f^2(y)dy+o(h^2).
\end{aligned}
$$

So that we have

$$
\begin{aligned}
&\frac{n(n-1)}{4(n-2)} I_2 \\
=&\mathrm{E}\{\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right) K\left(\frac{X_1-X_2}{h}\right)\left(\frac{\gamma(X_2)+\gamma(X_3)}{2h}\right) K\left(\frac{X_2-X_3}{h}\right)\} \\
&-\mathrm{E}\left(\frac{\gamma(X_1)+\gamma(X_2)}{2h}\right) K\left(\frac{X_1-X_2}{h}\right) \mathrm{E}\left(\frac{\gamma(X_2)+\gamma(X_3)}{2h}\right) K\left(\frac{X_2-X_3}{h}\right) \\
=&\int \gamma^2(y) f^3(y) dy + \frac{\mu_2(K)h^2}{2}\int \gamma''(y)\gamma(y) f^2(y) dy \\
&+\mu_2(K)h^2 \int \gamma^2(y) f''(y) f^2(y) dy + o(h^2) \\
&-\{\int \gamma(y) f^2(y) dy + \frac{\mu_2(K)h^2}{2}\int \gamma(y) f''(y) f(y) dy + \\
&\frac{\mu_2(K)h^2}{4}\int \gamma''(y) f^2(y) dy + \frac{\mu_2(K)h^2}{2}\int \gamma(y) f'(y) f(y) dy\}^2 \\
=&\int \gamma^2(y) f^3(y) dy - [\int \gamma(y) f^2(y) dy]^2 + O(h^2).
\end{aligned}
$$

Then we get that

$$
\mathrm{Var}(\hat{I}) = \frac{1}{n^2 h}\int \gamma^2(y) f^2(y) dy \int K^2(\mu) d\mu + \frac{4C_2}{n} + o(n^{-1}), \tag{2.22}
$$

where $C_2 = \int \gamma^2(y) f^3(y) dy - [\int \gamma(y) f^2(y) dy]^2$.

Now,

$$
\begin{aligned}
\mathrm{MSE}(\hat{I}) =&(\mathrm{Bias}\hat{I})^2 + \mathrm{Var}(\hat{I}) \\
=&\mu_2^2(K) h^4 (C_1)^2 + \frac{1}{n^2 h}\int \gamma^2(y) f^2(y) dy \int K^2(\mu) d\mu + \frac{4C_2}{n}.
\end{aligned}
$$

This gives us the optimal bandwidth $h$

$$
h_{opt} = n^{-2/5}\left(\frac{\int \gamma^2(y) f^2(y) dy \int K^2(\mu) d\mu}{4C_1 \mu_2^2(K)}\right)^{1/5} \tag{2.23}
$$

Next we show the equations (2.5) and (2.6).

First,

$$\mathbf{E}(\hat{I}_{el}) = \mathrm{E}\sum_{i\neq j} p_i p_j \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right). \tag{2.24}$$

Plug in $p_i$ from equation (2.3), then

$$\begin{aligned}
&\mathrm{E}(\hat{I}_{el}) \\
&= \mathrm{E}\sum_{1\leq i\neq j\leq n} \frac{1}{n}\frac{1}{1+\boldsymbol{\lambda}^T \mathrm{g}(X_i)}\frac{1}{n}\frac{1}{1+\boldsymbol{\lambda}^T \mathrm{g}(X_j)} \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right) \\
&= \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \mathrm{E}\frac{1}{1+\boldsymbol{\lambda}^T \mathbf{g}(X_i)}\frac{1}{1+\boldsymbol{\lambda}^T \mathbf{g}(X_j)} \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right).
\end{aligned}$$

By using that $\boldsymbol{\lambda} = O_p(n^{-1/2})$ and Taylor series expansion, then

$$
\begin{aligned}
&\mathrm{E}\frac{1}{1+\boldsymbol{\lambda}^T\mathbf{g}(X_i)}\frac{1}{1+\boldsymbol{\lambda}^T\mathbf{g}(X_j)}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
=&\mathrm{E}\left\{1-\boldsymbol{\lambda}^T\mathbf{g}(X_i)+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\boldsymbol{\lambda}+o_p(n^{-1})\right\}\\
&\left\{1-\boldsymbol{\lambda}^T\mathbf{g}(X_j)+\boldsymbol{\lambda}^T\mathbf{g}(X_j)\mathbf{g}^T(X_j)\boldsymbol{\lambda}+o_p(n^{-1})\right\}\\
&\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
=&\mathrm{E}\left\{1-\boldsymbol{\lambda}^T\mathbf{g}(X_i)+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\boldsymbol{\lambda}-\boldsymbol{\lambda}^T\mathbf{g}(X_j)\right.\\
&\left.+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\boldsymbol{\lambda}^T\mathbf{g}(X_j)+\boldsymbol{\lambda}^T\mathbf{g}(X_j)\mathbf{g}^T(X_j)\boldsymbol{\lambda}\right\}\\
&\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
=&\mathrm{E}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)-E\left\{\boldsymbol{\lambda}^T(\mathbf{g}(X_i)+\mathbf{g}(X_j))\right\}\\
&\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
&+\mathrm{E}\left\{\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\boldsymbol{\lambda}+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\boldsymbol{\lambda}^T\mathbf{g}(X_j)+\boldsymbol{\lambda}^T\mathbf{g}(X_j)\mathbf{g}^T(X_j)\boldsymbol{\lambda}\right\}\\
&\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right).
\end{aligned}
$$

So,

$$
\begin{aligned}
\mathrm{E}(\hat{I}_{el})=&\frac{1}{n^2}\sum_{i\neq j}\mathrm{E}\{\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
&-E\left\{\boldsymbol{\lambda}^T(\mathbf{g}(X_i)+\mathbf{g}(X_j))\right\}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
&+\mathrm{E}\left\{\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\boldsymbol{\lambda}+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\boldsymbol{\lambda}^T\mathbf{g}(X_j)+\boldsymbol{\lambda}^T\mathbf{g}(X_j)\mathbf{g}^T(X_j)\boldsymbol{\lambda}\right\}\\
&\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\}\\
=&\mathrm{E}(\hat{I})-E_1+E_{21}+E_{22}+E_{23}+o(n^{-1}),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathrm{E}(\hat{I}) =& \frac{1}{n^2}\sum_{i\neq j} \mathrm{E}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right)\\
E_1 =& \frac{1}{n^2}\sum_{i\neq j} E\left\{\boldsymbol{\lambda}^T(\mathbf{g}(X_i)+\mathbf{g}(X_j))\right\}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right)\\
E_{21} =& \frac{1}{n^2}\sum_{i\neq j} \mathrm{E}\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right)\\
E_{22} =& \frac{1}{n^2}\sum_{i\neq j} \mathrm{E}\boldsymbol{\lambda}^T\mathbf{g}(X_i)\boldsymbol{\lambda}^T\mathbf{g}(X_j)\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right)\\
E_{23} =& \frac{1}{n^2}\sum_{i\neq j} \mathrm{E}\boldsymbol{\lambda}^T\mathbf{g}(X_j)\mathbf{g}^T(X_j)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right).
\end{aligned}
$$

Now, we need to find $E_1$,$E_{21}$,$E_{22}$ and $E_{23}$, respectively. Using Taylor expansion for $\boldsymbol{\lambda}$, we have following equation:

$$
\boldsymbol{\lambda} = \Sigma^{-1}\frac{1}{n}\sum \mathbf{g}(X_i) + O_p(n^{-1}), \tag{2.25}
$$

where $\Sigma_{lm} = Cov(g_l(X), g_m(X))$, and $\mu_2(K) = \int \mu^2 K(\mu)d\mu$. Plug in $\boldsymbol{\lambda}$ as in (2.25). Hence

$$
E_1 = \frac{1}{n^3}\sum_{i\neq j,k} \mathrm{E}\{\Sigma^{-1}\sum \mathbf{g}(X_k)\}^T\{\mathbf{g}(X_i)+\mathbf{g}(X_j)\}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right).
$$

There are three cases to consider,$k = i \neq j$,$k = j \neq i$ and $k \neq i \neq j$, then

$$
\begin{aligned}
E_1 =& \frac{n-1}{n^2}\mathrm{E}\mathbf{g}^T(X_i)\Sigma^{-1}\left\{\mathbf{g}(X_i)+\mathbf{g}(X_j)\right\}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right)\\
&+ \frac{n-1}{n^2}\mathrm{E}\mathbf{g}^T(X_j)\Sigma^{-1}\left\{\mathbf{g}(X_i)+\mathbf{g}(X_j)\right\}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right)\\
&+ \frac{(n-1)(n-2)}{n^2}\mathrm{E}\mathbf{g}^T(X_k)\Sigma^{-1}\left\{\mathbf{g}(X_i)+\mathbf{g}(X_j)\right\}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right),
\end{aligned}
$$

since $X_1, \cdots, X_n$ are independent, and $\mathrm{E}g_l(X) = 0$, the third term is equal to zero, then

$$
\begin{aligned}
E_1 =& \frac{1}{n} \iint \mathbf{g}(x)^T \Sigma^{-1} \left\{ \mathbf{g}(x) + \mathbf{g}(y) \right\} \left( \frac{\gamma(x) + \gamma(y)}{2h} \right) \\
& K\left( \frac{x-y}{h} \right) f(x) f(y) dx dy \\
+& \frac{1}{n} \iint \mathbf{g}^T(y) \Sigma^{-1} \left\{ \mathbf{g}(x) + \mathbf{g}(y) \right\} \left( \frac{\gamma(x) + \gamma(y)}{2h} \right) \\
& K\left( \frac{x-y}{h} \right) f(x) f(y) dx dy + o(n^{-1}) \\
=& \frac{1}{n} \iint \mathbf{g}^T(y + \mu h) \Sigma^{-1} \left\{ \mathbf{g}(y + h\mu) + \mathbf{g}(y) \right\} \\
& \left( \gamma(y) + \frac{\mu h}{2} \gamma'(y) + (\frac{\mu h}{2})^2 \gamma''(y) \right) K(\mu) f(y + \mu h) f(y) d\mu dy \\
+& \frac{1}{n} \iint \mathbf{g}^T(y) \Sigma^{-1} \left\{ \mathbf{g}(y + \mu h) + \mathbf{g}(y) \right\} \\
& \left( \gamma(y) + \frac{\mu h}{2} \gamma'(y) + (\frac{\mu h}{2})^2 \gamma''(y) \right) K(\mu) f(y + \mu h) f(y) d\mu dy \\
=& \frac{1}{n} \iint \left\{ \mathbf{g}(y) + \mu h \mathbf{g}'(y) + \frac{\mu^2 h^2}{2} \mathbf{g}''(y) \right\}^T \Sigma^{-1} \\
& \left\{ \mathbf{g}(y) + \mu h \mathbf{g}'(y) + \frac{\mu^2 h^2}{2} \mathbf{g}''(y) + \mathbf{g}(y) \right\} \left( \gamma(y) + \frac{\mu h}{2} \gamma'(y) + (\frac{\mu h}{2})^2 \gamma''(y) \right) \\
& K(\mu) \left\{ f(y) + \mu h f'(y) + \frac{h^2 \mu^2}{2} f''(y) \right\} f(y) d\mu dy \\
&+ \frac{1}{n} \iint \mathbf{g}^T(y) \Sigma^{-1} \left\{ \mathbf{g}(y) + h\mu \mathbf{g}'(y) + \frac{\mu^2 h^2}{2} \mathbf{g}''(y) + \mathbf{g}(y) \right\} \\
& \left( \gamma(y) + \frac{\mu h}{2} \gamma'(y) + (\frac{\mu h}{2})^2 \gamma''(y) \right) K(\mu) \\
& \left\{ f(y) - \mu h f'(y) + \frac{h^2 \mu^2}{2} f''(y) \right\} f(y) d\mu dy \\
=& \frac{4}{n} \int \mathbf{g}^T(y) \Sigma^{-1} \mathbf{g}(y) \gamma(y) f^2(y) dy + o(n^{-1}).
\end{aligned}
$$

Next,

$$
\begin{aligned}
E_{21} =& \frac{1}{n^2}\sum_{i\neq j} \mathrm{E}\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
=& \frac{1}{n^4}\sum_{i\neq j,k,l} \mathrm{E}(\Sigma^{-1}\mathbf{g}(X_k))^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\Sigma^{-1}\mathbf{g}(X_l)\\
& \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right).
\end{aligned}
$$

There are three cases, $k=l\neq i\neq j$,$k=l=i\neq j$ and $k=l=j\neq i$, also we know $\mathrm{E}\mathrm{g}(X_i)=0$, so the rest of cases are equal to zero. The second and third case are order of $n^{-2}$, so only the first case is considered.

$$
\begin{aligned}
E_{21} = &= \frac{1}{n^4}\mathrm{E}\sum_{l=k\neq i\neq j}^{n} \mathbf{g}(X_k)^T\Sigma^{-1}\mathbf{g}(X_i)\mathbf{g}^T(X_i)\Sigma^{-1}\mathbf{g}(X_k)\\
& \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)+o(n^{-1})\\
&= \frac{(n-1)(n-2)}{n^3}\mathrm{E}\mathbf{g}(X_k)^T\Sigma^{-1}\mathbf{g}(X_i)\mathbf{g}^T(X_i)\Sigma^{-1}\mathbf{g}(X_k)\\
& \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\\
&= \frac{1}{n}\mathrm{E}\mathbf{g}^T(X_i)\Sigma^{-1}\mathbf{g}(X_k)\mathbf{g}(X_k)^T\Sigma^{-1}\mathbf{g}(X_i)\\
& \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)+o(n^{-1})\\
&= \frac{1}{n}\mathrm{E}\mathbf{g}^T(X_i)\Sigma^{-1}\mathbf{g}(X_i)\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)+o(n^{-1})\\
&= \frac{1}{n}\iint \mathbf{g}^T(x)\Sigma^{-1}\mathbf{g}(x)\left(\frac{\gamma(x)+\gamma(y)}{2h}\right)K\left(\frac{x-y}{h}\right)f(x)f(y)dxdy+o(n^{-1})\\
&= \frac{1}{n}\iint \mathbf{g}^T(x)\Sigma^{-1}\mathbf{g}(x)\left(\gamma(x)-\frac{\mu h}{2}\gamma'(x)+(\frac{\mu h)}{2})^2\gamma''(x)\right)\\
& K(\mu)f(x)f(x-\mu h)dxd\mu+o(n^{-1})\\
&= \frac{1}{n}\int \mathbf{g}^T(x)\Sigma^{-1}\mathbf{g}(x)\gamma(x)f^2(x)dx+o(n^{-1}).
\end{aligned}
$$

Similarly, $E_{23} = E_{22} = \frac{1}{n}\int \mathbf{g}^T(x)\Sigma^{-1}\mathbf{g}(x)\gamma(x)f^2(x)dx + o(n^{-1})$. From above calculations, then $\mathrm{E}(\hat{I}_{el})$

$$\begin{aligned}
\mathrm{E}(\hat{I}_{el}) =& \mathrm{E}(\hat{I}) - E_1 + E_{21} + E_{22} + E_{23} + o(n^{-1}) \\
=& \mathrm{E}(\hat{I}) - \frac{1}{n}\int \mathbf{g}^T(y)\Sigma^{-1}\mathbf{g}(y)\gamma(y)f^2(y)dy + o(n^{-1}),
\end{aligned}$$

where $\mathrm{E}(\hat{I})$ is equation (2.21).

Now, calculate $\mathrm{E}(\hat{I}_{el}^2)$.

$$\begin{aligned}
\mathrm{E}(\hat{I}_{el}^2) =& \mathrm{E}\sum_{i\neq j} p_i p_j \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right) \\
& \sum_{k\neq l} p_k p_l \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right) K\left(\frac{X_k - X_l}{h}\right) \\
=& \frac{1}{n^4}\mathrm{E}\sum_{i\neq j}\sum_{k\neq l} \frac{1}{[1+\lambda^T\mathbf{g}(X_i)][1+\lambda^T\mathbf{g}(X_j)][1+\lambda^T\mathbf{g}(X_k)][1+\lambda^T\mathbf{g}(X_l)]} \\
& \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) K\left(\frac{X_i - X_j}{h}\right)\left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right) K\left(\frac{X_k - X_l}{h}\right).
\end{aligned}$$

Using Taylor expression, then

$$
\begin{aligned}
\mathrm{E}(\hat{I}_{el}^2) =& \frac{1}{n^4}\mathrm{E}\sum_{i,j}\sum_{k,l}[1-\boldsymbol{\lambda}^T\mathbf{g}(X_i)+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}^T(X_i)\boldsymbol{\lambda}+o(n^{-1})] \\
& [1-\boldsymbol{\lambda}^T\mathbf{g}(X_j)+\boldsymbol{\lambda}^T\mathbf{g}(X_j)\mathbf{g}^T(X_j)\boldsymbol{\lambda}+o(n^{-1})] \\
& [1-\boldsymbol{\lambda}^T\mathbf{g}(X_k)+\boldsymbol{\lambda}^T\mathbf{g}(X_k)\mathbf{g}^T(X_k)\boldsymbol{\lambda}+o(n^{-1})] \\
& [1-\boldsymbol{\lambda}^T\mathbf{g}(X_l)+\boldsymbol{\lambda}^T\mathbf{g}(X_l)\mathbf{g}^T(X_l)\boldsymbol{\lambda}+o(n^{-1})] \\
& \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
=& \frac{1}{n^4}\mathrm{E}\sum_{i,j}\sum_{k,l}[1-\boldsymbol{\lambda}^T(\mathbf{g}(X_i)+\mathbf{g}(X_j)+\mathbf{g}(X_k)+\mathbf{g}(X_l)) \\
&+\boldsymbol{\lambda}^T(\mathbf{g}(X_i)\mathbf{g}^T(X_j)+\mathbf{g}(X_i)\mathbf{g}^T(X_k)+\mathbf{g}(X_i)\mathbf{g}^T(X_l)+\mathbf{g}(X_j)\mathbf{g}^T(X_k) \\
&+\mathbf{g}(X_j)\mathbf{g}^T(X_l)+\mathbf{g}(X_k)\mathbf{g}^T(X_l))\boldsymbol{\lambda}] \\
& \left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right)\left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right).
\end{aligned}
$$

Simplifying can get

$$
\begin{aligned}
\mathrm{E}(\hat{I}_{1el}^2) = & \mathrm{E}(\hat{I}_1^2) - \mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_i)+\mathbf{g}(X_j))\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right) \\
& K\left(\frac{X_i-X_j}{h}\right)\left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
& -\mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_k)+\mathbf{g}(X_l))\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right) \\
& \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
& +\mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_i)\mathbf{g}^T(X_j)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right) \\
& \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
& +\mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_i)\mathbf{g}^T(X_k)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right) \\
& \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
& +\mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_i)\mathbf{g}^T(X_l)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right) \\
& \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
& +\mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_j)\mathbf{g}^T(X_k)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right) \\
& \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
& +\mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_j)\mathbf{g}^T(X_l)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right) \\
& \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
& +\mathrm{E}\boldsymbol{\lambda}^T(\mathbf{g}(X_k)\mathbf{g}^T(X_l)\boldsymbol{\lambda}\left(\frac{\gamma(X_i)+\gamma(X_j)}{2h}\right)K\left(\frac{X_i-X_j}{h}\right) \\
& \left(\frac{\gamma(X_k)+\gamma(X_l)}{2h}\right)K\left(\frac{X_k-X_l}{h}\right) \\
= & \mathrm{E}(\hat{I}^2) - 2E_1\mathrm{E}(\hat{I}) + 2\mathrm{E}(\hat{I})E_{22} + E_1E_1 + o(n^{-1}).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\mathrm{Var}(\hat{I}_{el}) =& \mathrm{E}(\hat{I}_{el}^2) - (\mathrm{E}\hat{I}_{el})^2 \\
=& \mathrm{E}(\hat{I}^2) - 2E_1\mathrm{E}(\hat{I}) + 2\mathrm{E}(\hat{I})E_{22} + E_1E_1 \\
& -[\mathrm{E}(\hat{I}) - E_1 + E_{21} + E_{22} + E_{23}]^2 \\
=& \mathrm{E}(\hat{I}^2) - 2E_1\mathrm{E}(\hat{I}) + 2\mathrm{E}(\hat{I})E_{22} + E_1E_1 \\
& -\{[\mathrm{E}(\hat{I})]^2 + (E_1)^2 + (E_{21} + E_{22} + E_{23})^2 - 2\mathrm{E}(\hat{I})E_1 \\
& \quad - 2E_1(E_{21} + E_{22} + E_{23}) + 2\mathrm{E}(\hat{I})(E_{21} + E_{22} + E_{23})\} \\
=& \mathrm{Var}(\hat{I}) - 2\mathrm{E}(\hat{I})(E_{21} + E_{23}) - (E_{21} + E_{22} + E_{23})^2 \\
& \quad + 2E_1(E_{21} + E_{22} + E_{23}) + o(n^{-1}).
\end{aligned}
$$

Finally, $\mathrm{MSE}(\hat{I}_{el})$ can be written as

$$
\begin{aligned}
& \mathrm{MSE}(\hat{I}_{el}) = \mathrm{Var}(\hat{I}_{el}) + \mathrm{Bias}(\hat{I}_{el})^2 \\
=& \mathrm{Var}(\hat{I}) - 2\mathrm{E}(\hat{I})(E_{21} + E_{23}) - (E_{21} + E_{22} + E_{23})^2 \\
& + 2E_1(E_{21} + E_{22} + E_{23}) + (\mathrm{Bias}(\hat{I}) - E_1 + E_{21} + E_{22} + E_{23})^2 + o(n^{-1}) \\
=& \mathrm{MSE}(\hat{I}) + E_1E_1 - 2\mathrm{Bias}(\hat{I})E_1 + 2\mathrm{Bias}(\hat{I})(E_{21} + E_{22} + E_{23}) \\
& - 2\mathrm{E}(\hat{I})(E_{21} + E_{23}) + o(n^{-1}) \\
=& \mathrm{MSE}(\hat{I}) + E_1E_1 - 2\mathrm{Bias}(\hat{I})E_1 + 2\mathrm{Bias}(\hat{I})(E_{21} + E_{22} + E_{23}) \\
& - 2(\mathrm{Bias}(\hat{I}) + I)(E_{21} + E_{23}) + o(n^{-1}) \\
=& \mathrm{MSE}(\hat{I}) + E_1E_1 - 2\mathrm{Bias}(\hat{I})E_1 + 2\mathrm{Bias}(\hat{I})E_{22} - 2(I)(E_{21} + E_{23}) + o(n^{-1}) \\
=& \mathrm{MSE}(\hat{I}) - \frac{2}{n}\int \gamma(y)f^2(y)dy \int \mathbf{g}^T(y)\Sigma^{-1}\mathbf{g}(y)\gamma(y)f^2(y)dy + o(n^{-1}).
\end{aligned}
$$

# 3 New Kernel Density Estimations and their Empirical Likelihood Version

## 3.1 Introduction of Kennel Density Estimation

The kernel method is a popular tool for the non-parametric estimation of the probability density function $f$. Suppose the independent and identical distribution sample $X_1, \cdots, X_n$ from a continuous distribution, a kernel density estimator for $f$ at an arbitrary point $x$, is

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

where $K$ is a kernel function and $h$ is a smoothing parameter that controls the smoothness of the fit. The choice of the shape of the kernel function is not a particularly important one. However, the choice of the value of the bandwidth is very important to trade off between the bias and the variance. When the bandwidth $h$ is increasing, the bias is increasing but the variance is decreasing. The estimate of point $x$ is the average of $\frac{1}{h}K(\frac{x-X_i}{h})$, and the kernel $K$ is bounded. Also, in the symmetric density function for small $h$, the estimate only pays attention to the local data (the observations close to the point x) because the value of $K(\frac{x-X_i}{h})$ is almost equal to zero when $X_i$ is far away from $x$. But for the large $h$, the value of $K(\frac{x-X_i}{h})$ is extremely close for most observations, so in this case, it smooths away some details, such as humps and valleys. Based on these reasons, the kernel density estimation has some drawbacks, such as difficulty in catching humps and valleys and finding the bandwidth. In this chapter, two new kernel density estimators are proposed in Section 3.1.1, and their empirical likelihood versions are also provided in Section 3.4.

### 3.1.1 New Kernel Density Estimations

Bandwidth plays an important role in the kernel density estimation. If bandwidth $h$ is small, the estimate pays too much attention to the particular data set and does not allow for variation across the sample. If bandwidth $h$ is large, the estimate is too smooth, in that it smooths away some details. To solve this problem, the bandwidth $h$ of new estimations has two factors: one is the smoothing parameter $h$ which controls smoothness, and the other is a scale coefficient which balances the local data and data from the whole sample. Under the same assumptions on the standard **KDE**, since the $X_i$ is the random sample, for the sake of simplicity, choose the scale coefficient is the index of $X_i$. So the bandwidth for $X_1$ is $h$, $X_2$ is $2h$, and so on, $X_n$ is $nh$. For fixed $x$, the value of $K(\frac{x-X_i}{ih})$ is not only dependent on the distance between $x$ and $X_i$, it is also dependent on the scale coefficient $i$. As a result, these estimators are smooth enough and are also able to catch the humps and valleys. By minimizing the $AMISE$, we reach the optimal $h_{new.opt} = O(n^{-6/5})$, which is smaller than standard kernel optimal bandwidth $h = O(n^{-1/5})$. The new estimators therefore have smaller bandwidth $ih$; only the $n^{th}$ of the data $X_n$ has the same order of the standard kernel optimal bandwidth. Theoretical and simulation results show that the new kernel density estimators have better performance than that of standard **KDE**.

### 3.1.2 First New kernel Density Estimation: Local Coefficient Adjustment(LCA)

The first kernel estimator **LCA** is defined as follows:

$$\hat{f}_{LCA}(x) = \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{i}K\left(\frac{x-X_i}{ih}\right). \tag{3.1}$$

In this estimate, each kernel has adjusted coefficient $1/i$ and the bandwidth coefficient is $i$. Also $\int \hat{f}_{LCA}(x)dx = 1$ when $\int K(\mu)d\mu = 1$. So $\hat{f}_{LCA}(x)$ is the density function. This estimate transforms $ih* = h$ in the standard **KDE**. So this transformation makes the value of $K(\frac{x-X_i}{ih})$ dependent not only on the distance of $x$ and $X_i$. In this way, by choosing a small $h$, this estimate is able to catch more details and is also sufficiently smooth, since the estimate pays attention to the whole sample data, not only some local data.

### 3.1.3 Second New Kernel Density Estimation: Global Coefficient Adjustment(GCA)

With the idea of **LCA**, choosing the same bandwidth, but this estimator has the global coefficient adjustment $\frac{2}{n+1}$ instead of $\frac{1}{i}$ for each $K(\frac{x-X_i}{ih})$ in **LCA**. Then **GCA** is defined as follows:

$$\hat{f}_{GCA}(x) = \frac{2}{n(n+1)h} \sum_{i=1}^{n} K\left(\frac{x-X_i}{ih}\right), \tag{3.2}$$

and $\int \hat{f}_{GCA}(x)dx = 1$ when $\int K(\mu)d\mu = 1$ , so **GCA** is density function. Since the coefficient of this estimator is fixed, adding one more observation yields the following result,

$$\begin{aligned}
\hat{f}_{GCA.n+1}(x) =& \frac{2}{(n+1)(n+2)h} \left\{ \sum_{i=1}^{n} K\left(\frac{x-X_i}{ih}\right) + K\left(\frac{x-X_{n+1}}{(n+1)h}\right) \right\} \\
=& \frac{n}{n+2}\hat{f}_{GCA.n}(x) + \frac{2}{(n+1)(n+2)h} K\left(\frac{x-X_{n+1}}{(n+1)h}\right).
\end{aligned}$$

Hence, $\hat{f}_{GCA}(x)$ is a recursive estimate that can be used in sampling schemes.

## 3.2 New Kernel Density Estimation Properties

In this section, four different measurement errors (see Section 1.4), are discussed for estimators **GCA** and **LCA**: bias, variance, MSE, and MISE.

### 3.2.1 Bias and Variance

Theorem 3.2.1 shows the bias and variance based on the **GCA** and **LCA**.

**Theorem 3.2.1**

$$\text{Bias}[\hat{f}_{GCA}(x)] = \frac{1}{4}n^2h^2\mu_2(K)f''(x) + o(n^2h^2), \tag{3.3}$$

$$\text{Bias}[\hat{f}_{LCA}(x)] = \frac{1}{6}n^2h^2\mu_2(K)f''(x) + o(n^2h^2), \tag{3.4}$$

$$\text{Var}[\hat{f}_{GCA}(x)] = \frac{2}{n^2h}f(x)R(K) + o(n^{-1}, h), \tag{3.5}$$

$$\text{Var}[\hat{f}_{LCA}(x)] = \frac{(\sum_{i=1}^{n}\frac{1}{i})}{n^2h}f(x)R(K) + o(n^{-1}, h), \tag{3.6}$$

*where* $\mu_2(K) = \int \mu^2 K(\mu)d\mu < \infty$, *and* $R(K) = \int K^2(\mu)d\mu$.

By minimizing AMISE, we can solve the new method optimal bandwidth $h_{new.opt} = O(n^{-6/5})$ for both estimators, while the standard **KDE** has the optimal bandwidth $h_{opt} = O(n^{-1/5})$.

### 3.2.2 MSE and MISE

Combine the Bias and the Var in Theorem 3.2.1, then MSE is given by the following theorem.

**Theorem 3.2.2**

$$\text{MSE}\hat{f}_{GCA}(x) = \frac{2}{n^2h}R(K)f(x) + \frac{1}{16}n^4h^4\mu_2^2(K)(f''(x))^2 + o(n^{-1}), \tag{3.7}$$

$$\text{MSE}\hat{f}_{LCA}(x) = \frac{\sum_{i=1}^{n}(\frac{1}{i})}{n^2h}R(K)f(x) + \frac{1}{36}n^4h^4\mu_2^2(K)(f''(x))^2 + o(n^{-1}). \tag{3.8}$$

*This leads to the* AMISE *expression*

$$\text{AMISE}_{GCA} = \frac{2}{n^2 h} R(K) + \frac{1}{16} n^4 h^4 \mu_2^2(K) R(f''), \tag{3.9}$$

*and*

$$\text{AMISE}_{LCA} = \frac{\sum_{i=1}^{n}(\frac{1}{i})}{n^2 h} R(K) + \frac{1}{36} n^4 h^4 \mu_2^2(K) R(f''). \tag{3.10}$$

*Then* AMISE-*optimal bandwidth is*

$$h_{AMISE.GCA} = n^{-6/5} \left[ \frac{8R(K)}{\mu_2^2(K) R(f'')} \right]^{1/5}, \tag{3.11}$$

*and*

$$h_{AMISE.LCA} = n^{-6/5} \left[ \frac{9R(K) \sum_{i=1}^{n}(\frac{1}{i})}{\mu_2^2(K) R(f'')} \right]^{1/5}. \tag{3.12}$$

Theorem 3.2.2 shows that optimal bandwidth $h = o(n^{-6/5})$ for the **GCA** and $h = o(n^{-6/5}(\sum_{i=1}^{n} \frac{1}{i})^{1/5})$, when $n$ is large, $\sum_{i=1}^{n}(\frac{1}{i}) \sim \log n$, so optimal bandwidth of **GCA** is larger than that of **GCA**. But the optimal bandwidths of **GCA** and **LCA** are both smaller than those of standard **KDE**.

### 3.2.3 Estimation of Density functionals

An important component of bandwidth selectors is the estimation of integrated squared density derivatives. The general integrated squared density derivative functional is

$$R(f^{(p)}) = \int f^{(p)}(x)^2 dx.$$

Under sufficient smoothness assumption on $f$, using integration by parts can get,

$$R(f^{(s)}) = (-1)^s \int f^{(2s)}(x) f(x) dx.$$

Therefore, it is sufficient to study the functional estimation of this form

$$\psi = \int f^{(r)}(x)f(x)dx,$$

for r even. Note that the sign of $\psi_{2s}$ is the same as that of $(-1)^s$ and $\psi_r = 0$ if r is odd. Also note that

$$\psi_r = \mathrm{E}\{f^{(r)}(x)\}.$$

This motivates the density estimator

$$\hat{\psi}_r(h) = \int f^{(r)}(x)dF_n(x) = n^{-1}\sum_{i=1}^{n}\hat{f}^{(r)}(X_i;h). \tag{3.13}$$

So for the **GCA** and **LCA**, we have

$$\begin{aligned}
&\hat{\psi}_r(h)_{GCA}\\
&= \frac{2}{n^2(n+1)h}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{jh}\right)^r K^{(r)}\left(\frac{X_i - X_j}{jh}\right)\\
&= \frac{2K^{(r)}(0)}{n(n+1)h^{r+1}}\sum_{i=1}^{n}(\frac{1}{j})^r + \frac{2}{n^2(n+1)h}\sum_{i\neq j}^{n}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i - X_j}{jh}\right),
\end{aligned} \tag{3.14}$$

$$\begin{aligned}
&\hat{\psi}_r(h)_{LCA}\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{jh}\right)^{r+1} K^{(r)}\left(\frac{X_i - X_j}{jh}\right)\\
&= \frac{K^{(r)}(0)}{n^2h^{r+1}}\sum_{i=1}^{n}(\frac{1}{j})^{r+1} + \frac{1}{n^2h}\sum_{i\neq j}^{n}(\frac{1}{jh})^{r+1} K^{(r)}\left(\frac{X_i - X_j}{jh}\right),
\end{aligned} \tag{3.15}$$

where $h$ is the bandwidth and $K$ is kernel density.

**Theorem 3.2.3**

$$\begin{aligned}
\mathrm{E}\hat{\psi}_r(h)_{GCA} &= \frac{2K^{(r)}(0)}{n(n+1)h^{r+1}}\sum_{i=1}^{n}(\frac{1}{j})^r + \frac{n-1}{n}\psi_r + \frac{n^2h^2}{4}\mu_2(K)\psi_{r+2} + o(n^2h^2) \\
\mathrm{E}\hat{\psi}_r(h)_{LCA} &= \frac{K^{(r)}(0)}{n^2h^{r+1}}\sum_{i=1}^{n}(\frac{1}{j})^{r+1} + \frac{n-1}{n}\psi_r + \frac{n^2h^2}{6}\mu_2(K)\psi_{r+2} + o(n^2h^2) \\
\mathrm{Var}\hat{\psi}_r(h)_{GCA} &= \frac{3}{n}\int f(y)[f^{(r)}(y)]^2dy - \frac{8}{n}[\int f(y)f^{(r)}(y)dy]^2 \\
\mathrm{Var}\hat{\psi}_r(h)_{LCA} &= \frac{2}{nh^{2r+1}}\psi_0 R(K^{(r)}) + \frac{4}{n}\{\int f(y)[f^{(r)}(y)]^2dy] - \psi_r\},
\end{aligned}$$

*where* $\psi_{r+s} = \int f^{(r)}(x)f^{(s)}(x)dx$

From this theorem, we see that $\hat{\psi}_r(h)$ is an asymptotically unbiased estimator. Also the variance of $\hat{\psi}_r(h)$ does not depend on $h$.

## 3.3 Bandwidth Selection

The choice of the bandwidth $h$ is more important for the behavior of $\hat{f}$ than the choice of kernel $K$. A small value of $h$ makes the estimate look "wiggly" and shows spurious features, whereas too large a value of $h$ will lead to an estimate that is too smooth, in the sense that it is too biased and may not reveal structural features, such as bimodality. Figure 5 shows a mixture normal distribution $0.5N(-1, 4/9) + 0.5N(1, 4/9)$ by using **GCA** estimator for different values of $h$ based on a sample of 100 observations for 1000 replications. And Figure 6 uses the **LCA** estimator. These two figures both show that for $h = 0.001$, the estimators are not smooth, but they still catch humps and valleys very well. Even if we choose a smaller bandwidth for each replication data set by choosing a random bandwidth $ih$ instead of a fixed $h$, the estimates still allow for variation across samples. Unlike standard **KDE**, if we choose a smaller bandwidth, the estimate focuses on particular data and is overly noisy for most of the whole sampled data. In these two figures, a compromise is reached with $h = 0.005$ represented by the red dotted lines. These estimates are not overly noisy and recover the essential structure of the true density. When $h = 0.015$, these estimates are overly smooth, since the bimodality structure has been smoothed away.

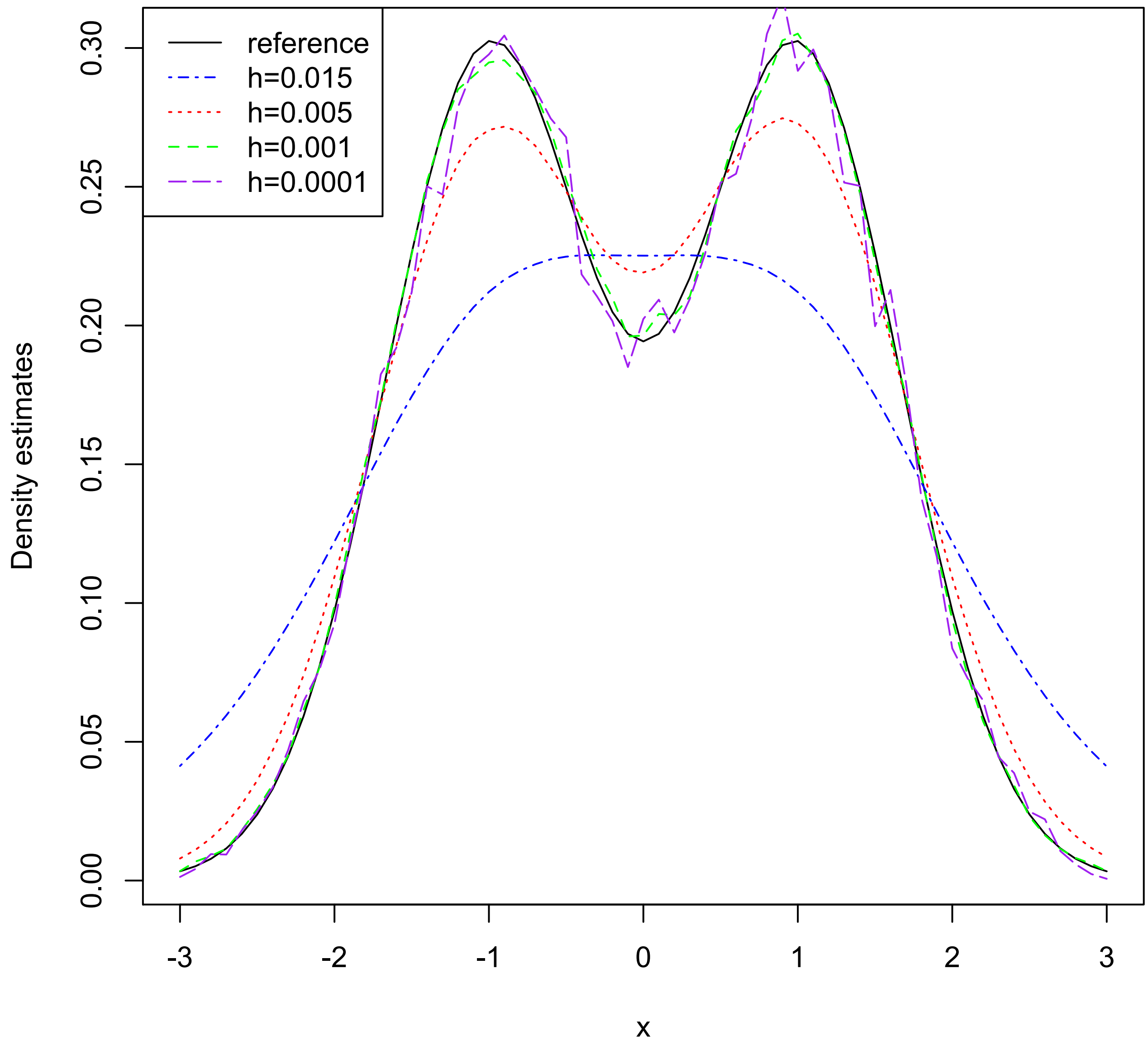



**Figure 5:** Density estimate by **GCA** from bimodal distribution $0.5N(-1, 4/9) + 0.5N(1, 4/9)$ for different bandwidths

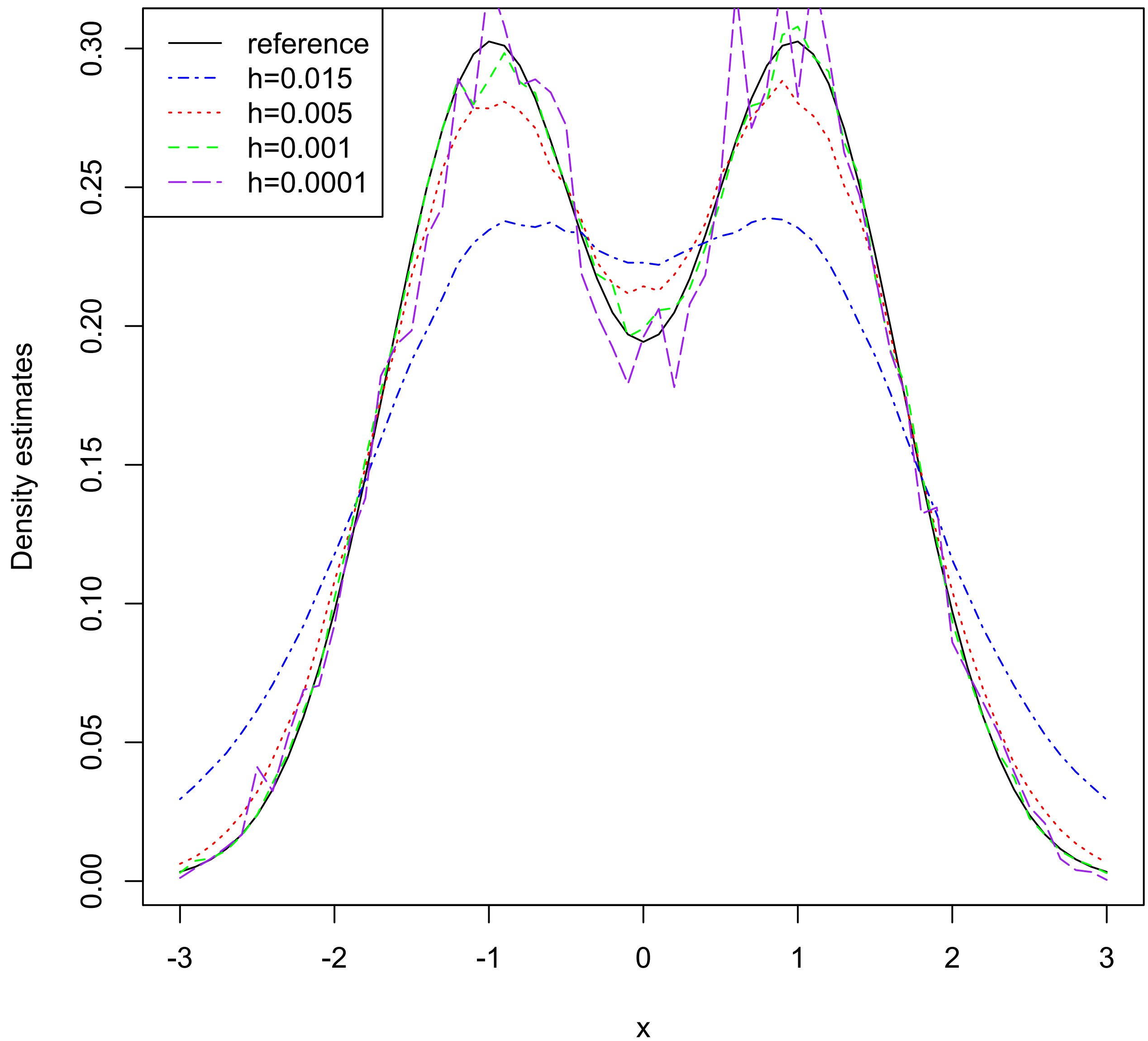



**Figure 6:** Density estimate by **LCA** from bimodal distribution $0.5N(-1, 4/9) + 0.5N(1, 4/9)$ for different bandwidth

In this section, the performance measures are mean integrated squared error **MISE** (1.12) and integrated squared error **ISE** (1.13). Based on these error criteria, we introduce a reliable data-driven estimator of the optimal bandwidth, the cross-validation via plug-in method which tries to minimize the **MISE** to find $h_{opt}$.

### 3.3.1 Unbiased Cross Validation Method

The idea of unbiased cross-validation was introduced by Rudemo (1982) and Bowman (1984). In this section, we employ this unbiased least squared cross-validation of bandwidth selection in the new kernel estimate **GCA** and **LCA**. We will begin our description of selection of bandwidth selectors. Ideally, for each sample, we would like to construct a density estimate to minimize the **ISE** (1.13). Least squares cross-validation attempts to address **ISE** rather than **MISE**. Its motivation comes from expanding the MISE of $\hat{f}(.;h)$ to obtain

$$\text{MISE}\hat{f}(x;h) = \text{E}\int(\hat{f}(x))^2 - 2\text{E}\int \hat{f}(x)f(x)dx + \int f^2(x)dx. \tag{3.16}$$

Since the last term does not depend on $h$, minimizing the MISE$\hat{f}(x,h)$ is equivalent to minimizing

$$\text{MISE}\hat{f}(x,h) - \int f^2(x)dx = \text{E}\int(\hat{f}(x))^2 - 2\text{E}\int \hat{f}(x)f(x)dx. \tag{3.17}$$

Then consider the cross-validation estimator

$$\begin{aligned}\text{LSCV}(h) &\equiv \int \hat{f}(x)^2dx - 2\int \hat{f}_{-i}(x)dF_n(x)\\ &= \int \hat{f}(x)^2dx - 2\sum_{i=1}^{n}\hat{f}_{-i}(X_i),\end{aligned} \tag{3.18}$$

where $F_n(x)$ is empirical cumulative density function (**ECDF**) based on the sample with $X_i$ deleted.

$$\hat{f}_{-i}(x)_{GCA} = \frac{1}{n(n-1)h}\{\sum_{j=1}^{i-1} K\left(\frac{x-X_j}{jh}\right) + \sum_{j=i+1}^{n} K\left(\frac{x-X_j}{(j-1)h}\right)\}$$

and

$$\hat{f}_{-i}(x)_{LCA} = \frac{1}{(n-1)h}\sum_{j=1}^{n}\frac{1}{j}K\left(\frac{x-X_j}{jh}\right)$$

Now we check the expectations of LSCV$(h)_{GCA}$ and LSCV$(h)_{LCA}$.

$$\begin{aligned}
&\frac{1}{n}\mathrm{E}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)_{GCA}\\
=&\sum_{i=1}^{n}\sum_{j=1}^{i-1}\frac{1}{n^2(n-1)h}K\left(\frac{y-X_j}{jh}\right)f(y)dy + \sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{1}{n(n-1)h}K\left(\frac{y-X_j}{jh}\right)f(y)dy\\
=&\mathrm{E}\int \hat{f}(y)_{GCA}f(y)dy.
\end{aligned}$$

Same approach on LSCV$(h)_{LCA}$,

$$\frac{1}{n}\mathrm{E}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)_{LCA} = \mathrm{E}\int \hat{f}(y)_{LCA}f(y)dy.$$

So,

$$\begin{aligned}
\mathrm{E}\{\mathrm{LSCV}(h)_{GCA}\} =&\mathrm{MISE}(h)_{GCA} - R(f),\\
\mathrm{E}\{\mathrm{LSCV}(h)_{LCA}\} =&\mathrm{MISE}(h)_{LCA} - R(f).
\end{aligned}$$

Hence, for the fixed bandwidth, LSCV$(h)_{GCA}$ and LSCV$(h)_{LCA}$ are unbiased estimators.

Now we will introduce least squares unbiased cross-validation for the **GCA**.

$$
\begin{aligned}
&\widehat{\mathrm{LSCV}}_{GCA}(h) \\
&= \int \hat{f}(x)^2 dx - 2\int \hat{f}_{-i}(x) dF_n \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\int \frac{2}{n(n+1)h} K\left(\frac{x-X_i}{ih}\right)\frac{2}{n(n+1)h} \\
&- \frac{4}{n^2(n-1)h}\sum_{i=1}^{n}\{\sum_{j=1}^{i-1} K\left(\frac{x-X_j}{jh}\right) + \sum_{j=i+1}^{n} K\left(\frac{x-X_j}{(j-1)h}\right)\} \\
&= \frac{2R(K)}{n(n+1)h} + \frac{4}{n^2(n+1)(n-1)}\sum\sum_{i\neq j}\int \frac{1}{h^2} K\left(\frac{x-X_i}{ih}\right) K\left(\frac{x-X_j}{jh}\right) dx \\
&- \frac{4}{n^2(n-1)h}\sum_{i=1}^{n}\{\sum_{j=1}^{i-1} K\left(\frac{x-X_j}{jh}\right) + \sum_{j=i+1}^{n} K\left(\frac{x-X_j}{(j-1)h}\right)\} \qquad (3.19) \\
&= \frac{2R(K)}{n^2 h} + \frac{4}{n^4}\sum\sum_{i\neq j}\int \frac{1}{h^2} K\left(\frac{x-X_i}{ih}\right) K\left(\frac{x-X_j}{jh}\right) dx \\
&- \frac{4}{n^3 h}\sum_{i=1}^{n}\{\sum_{j=1}^{i-1} K\left(\frac{x-X_j}{jh}\right) + \sum_{j=i+1}^{n} K\left(\frac{x-X_j}{(j-1)h}\right)\}. \qquad (3.20)
\end{aligned}
$$

The last equation replaces $n \pm 1$ to n.

$$
\begin{aligned}
&\widehat{\text{LSCV}}_{LCA}(h) \\
&= \int \hat{f}(x)^2 dx - 2\int \hat{f}_{-i}(x) dF_n \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\int \frac{1}{n^2h^2ij} K\left(\frac{x-X_i}{ih}\right) K\left(\frac{x-X_j}{jh}\right) dx \\
&- \frac{2}{n(n-1)h}\sum\sum_{i\neq j}\frac{1}{j}K\left(\frac{X_i-X_j}{jh}\right) \\
&= \frac{\sum_{i=1}^{n}\frac{1}{i}R(K)}{nh} + \frac{1}{n^2h^2}\sum\sum_{i\neq j}\int\frac{1}{ij}K\left(\frac{x-X_i}{ih}\right)K\left(\frac{x-X_j}{jh}\right)dx \\
&- \frac{2}{n(n-1)h}\sum\sum_{i\neq j}\frac{1}{j}K\left(\frac{X_i-X_j}{jh}\right) \qquad (3.21) \\
&= \frac{\sum_{i=1}^{n}\frac{1}{i}R(K)}{nh} + \frac{1}{n^2h^2}\sum\sum_{i\neq j}\int\frac{1}{ij}K\left(\frac{x-X_i}{ih}\right)K\left(\frac{x-X_j}{jh}\right)dx \\
&- \frac{2}{n^2h}\sum\sum_{i\neq j}\frac{1}{j}K\left(\frac{X_i-X_j}{jh}\right). \qquad (3.22)
\end{aligned}
$$

### 3.3.2 Bias Cross Validation Method

The idea of biased least squares cross-validation methods for the classic **KDE** goes back to Scott and Terrell (1987). In this section, we employ this biased cross-validation method of bandwidth selection in the new kernel estimate **GCA** and **LCA**. The motivation comes from asymptotic expansion for AMISE as given in (3.9), and (3.10) contains only one unknown quantity ( $R(\hat{f}^{(p)}_{GCA})$ and $R(\hat{f}^{(p)}_{LCA})$) , where $\hat{f}_{GCA}$ and $\hat{f}_{LCA}$ are new kernel estimators which are defined in section 3.1.1 and $(p)$ is the $p$ derivatives. The $\mathbf{BCV}_{GCA}$ and $\mathbf{BCV}_{LCA}$ objective functions are obtained by replacing the unknown $R(f'')$ in the (3.9) and (3.10) by the estimators

$$
\begin{aligned}
\widetilde{R(f''_{GCA})} =& \frac{4}{n^4h^6}\sum_{i\neq j}\int (\frac{1}{ji})^2 K''\left(\frac{x-X_i}{ih}\right)K''\left(\frac{x-X_j}{jh}\right)dx, \\
\widetilde{R(f''_{LCA})} =& \frac{1}{n^2h^6}\sum_{i\neq j}\int (\frac{1}{ji})^3 K''\left(\frac{x-X_i}{ih}\right)K''\left(\frac{x-X_j}{jh}\right)dx.
\end{aligned}
$$

These two selectors, $\widetilde{R(f''_{GCA})}$ and $\widetilde{R(f''_{LCA})}$, used the data set of the $i \neq j$ case. These selectors use cross-validation techniques.

$$\mathbf{BCV}(h)_{GCA} = \frac{2}{n^2 h} R(K) + \frac{n^4}{16} h^4 \mu_2^2(K) \widetilde{R(f'')}_{GCA}, \tag{3.23}$$

$$\mathbf{BCV}(h)_{LCA} = \frac{\sum_{i=1}^{n} \frac{1}{i}}{n^2 h} R(K) + \frac{1}{9} n^4 h^4 \mu_2^2(K) \widetilde{R(f'')}_{LCA}. \tag{3.24}$$

## 3.4 Empirical Likelihood Based on GCA and LCA Estimation

In some statistical applications, additional information about $f$ is available: the mean or variance of a distribution may be known, such as when estimating equations. This additional information usually can be expressed as (1.10).

### 3.4.1 Empirical Likelihood Based on GCA(ELGCA)

**ELGCA** uses empirical likelihood in conjunction with the new kernel method (**GCA**) to provide a systematic approach for capturing the extra information. Suppose the extra information can be formulated as equation (1.8), then, **ELGCA** can be constructed by replacing $n^{-1}$ in equation (3.2) with the empirical likelihood $p_i$ under extra information (1.8). Then $p_i$ can be determined by maximizing a multinational $\prod_1^n np_i$ subject to

$$\sum p_i = 1, \qquad \sum ip_i = \frac{n+1}{2} \quad \text{and} \qquad \sum p_i g_l(X_i) = 0 \qquad\qquad (l = 1, 2, \cdots, q).$$

The second constraint makes the equation (3.26) to be density function. Let $\lambda_1, \lambda_2, \cdots, \lambda_q$ be Lagrange multipliers corresponding to the $q$ constraints. Define $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_q)^T$ and $\mathbf{g}(X_i) = \{g_1(X_i), g_2(X_i), \cdots, g_q(X_i)\}$. Then the weight $p_i$ are

$$p_i = n^{-1} \left\{1 + \boldsymbol{\lambda}^T \mathbf{g}(X_i)\right\}^{-1} \quad (i = 1, 2, \cdots, n), \tag{3.25}$$

where $\boldsymbol{\lambda}$ is the solution of

$$\sum_{i=1}^{n} \frac{g_l(X_i)}{1+\boldsymbol{\lambda}^T \mathbf{g}(X_i)} = 0 \quad (l = 1, 2, \cdots, q).$$

**ELGCA** is obtained by replacing $n^{-1}$ with the $p_i$ (3.25) in (3.2), so

$$\hat{f}_{el.GCA}(x) = \frac{2}{(n+1)h} \sum_{i=1}^{n} p_i K\left(\frac{x - X_i}{ih}\right). \tag{3.26}$$

It is easy to check that $\hat{f}_{el.GCA}(x)$ is a density function.

### 3.4.2 Empirical Likelihood Based on LCA Estimation (ELLCA)

This section, similar to Section 3.4.1, uses the empirical likelihood technique to apply the **LGA** estimation. Suppose the extra information about $f$ is available and can be expressed as the 1.8. Then $p_i$ can be determined by maximizing a multinational $\prod_{i=1}^{n} np_i$

$$\sum p_i = 1, \quad \text{and} \qquad \sum p_i g_l(X_i) = 0 \qquad\qquad (l = 1, 2, \cdots, q).$$

**ELLCA** is obtained by replacing $n^{-1}$ with the $p_i$ at equation 1.9 in the **LCA** (3.1). So **ELLCA** can be expressed

$$\hat{f}_{el.LGA}(x) = \frac{1}{h} \sum_{i=1}^{n} \frac{p_i}{i} K\left(\frac{x - X_i}{ih}\right). \tag{3.27}$$

It easy to check $\hat{f}_{el.LGA}(x)$ is a density function.

### 3.4.3 Bias and Variance of ELGCA and ELLCA

In this section, the bias and variance of the new empirical likelihood-based kernel density estimators are investigated, and the performance of all estimators is compared. We assume the function $g_l$ and kernel $K$ satisfied the following conditions:

1. for $l = 1, \cdots, q, g_l$ are smooth functions with enough derivatives;

2. $Eg_l^{(k)}(X) < \infty$ for nonnegative integer $k \leqslant 4$;

3. $K$ is symmetric about zero and is the probability density.

**Theorem 3.4.1**

$$\mathrm{E}(\hat{f}_{el.GCA}) = \mathrm{E}(\hat{f}_{GCA}) + o(n^{-1}), \tag{3.28}$$

$$\mathrm{E}(\hat{f}_{el.LCA}(x)) = \mathrm{E}(\hat{f}_{LCA}) + o(n^{-1}), \tag{3.29}$$

$$\mathrm{Var}(\hat{f}_{el.GCA}(x)) = \mathrm{Var}(\hat{f}_{GCA}(x)) - \frac{1}{n}\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)^2 + o(n^{-1}), \tag{3.30}$$

$$\mathrm{Var}(\hat{f}_{el.LCA}(x)) = \mathrm{Var}(\hat{f}_{LCA}(x)) - \frac{1}{n}\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)^2 + o(n^{-1}), \tag{3.31}$$

where $\mathrm{E}\hat{f}_{GCA}$,$\mathrm{E}\hat{f}_{LCA}$,$\mathrm{Var}\hat{f}_{GCA}$ and $\mathrm{Var}\hat{f}_{LCA}$ are defined in Theorem 3.2.1.

Theorem 3.4.1 shows that the difference between $\mathrm{E}(\hat{f}_{el.GCA})$ and $\mathrm{E}(\hat{f}_{GCA})$ is $o(n^{-1})$, so is between $\mathrm{E}(\hat{f}_{el.GCA})$ and $\mathrm{E}(\hat{f}_{GCA})$. Also it is obvious that the coefficient of $n^{-1}$ is always negative in the equation (3.30) and (3.31), there is an $O(n^{-1})$ reduction in the variance of $\hat{f}_{el.GCA}(x)$ and $\hat{f}_{GCA}(x)$, so is variance of $\hat{f}_{el.LCA}(x)$ and $\hat{f}_{LCA}(x)$. Using the empirical likelihood technique can reduce the variance with $O(n^{-1})$, and this reduction decreases as the sample size increases. Simulations show that when n is greater than 25, $\hat{f}_{el.GCA}(x)$ and $\hat{f}_{GCA}(x)$ are almost the same for standard normal distribution, and so are $\hat{f}_{el.LCA}(x)$ and $\hat{f}_{LCA}(x)$.
Immediately from the Theorem 3.4.1, the MISE for both estimators has the following results,

$$\mathrm{MISE}\hat{f}_{el.GCA} = \mathrm{MISE}\hat{f}_{GCA} - \frac{1}{n}\int \mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)^2dx + o(n^{-1}) \tag{3.32}$$

$$\mathrm{MISE}\hat{f}_{el.LCA} = \mathrm{MISE}\hat{f}_{LCA} - \frac{1}{n}\int \mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)^2dx + o(n^{-1}) \tag{3.33}$$

There is a reduction in the mean integrated square error due to using the extra information by the empirical likelihood, since the coefficients of $n^{-1}$ in the equation (3.32) and (3.33) are always positive.

### 3.5 Simulation Study

In this section, the performance of six estimators will be compared: **GCA**, **ELGCA**, **LCA**, **ELLCA**, standard **KDE**, and **ELKDE**. Data will be generated from standard normal distribution, mixture normal $0.5N(-1,4/9)+0.5N(1,4/9)$ and mixture normal $0.75N(0,1)+0.25N(1.5,4/9)$, with 1000 replications and sample size n=15, 25, 50, 100 and 500. All figures show that **GCA** and **LCA** are significantly better than standard **KDE**, especially at catching humps and valleys and on the tails (extreme points). When the sample size is small, the empirical likelihood version is significantly better than the non-empirical likelihood version. The difference between the estimators shrinks as the sample size increases.

For the normal distribution, when sample size n=15 (see figure 7), **GCA** and **ELGCA** are almost identical, and both are close to true density function. **ELLCA** is slightly better than **LCA**, and all four estimators are significantly better than **ELKDE** and **KDE**. Especially on the mode and two tails, the two new estimators and their empirical likelihood versions are better than **KDE** and **ELKDE**. When the sample size n=25 (see figure 8), the empirical likelihood version estimators are almost the same as their standard estimators. But **GCA** is better than **LCA**, and both estimators are better than **KDE**. And when sample size increases, the difference between these estimators decreases. When sample size n=500, these six estimators are almost the same.

For the mixture normal $0.75N(0,1)+0.25N(1.5,4/9)$, when sample size n=15 (see figure 9), **ELGCA** is best at catching the right hump, followed by **ELLCA**, **GCA**, **LCA**, **ELKDE**, and finally, **KDE**. For catching the valley, **GCA** is best, followed by **LCA**, **ELGCA**, **ELLCA** and **KDE**. For catching the left hump, **ELGCA** is almost the same as **GCA**, and both are better than **ELLCA** and **LCA**. All four of these estimators are significantly better than **ELKDE** and **KDE**. Figure 10 shows that the empirical likelihood version estimates are significant better than standard estimates for sample size n=25, n=50. When n=100, empirical likelihood version estimations are slightly better than standard estimators. When sample size n=500, empirical likelihood version estimations are almost the same as standard estimators. But for all cases, **LCA** is better than **GCA**, and both of these estimators are better than **KDE**. Also, when sample sizes increase, the difference between these estimators decreases. But the new estimators are better than classic **KDE**.

For the mixture normal distribution $0.5N(-1, 4/9) + 0.5N(1, 4/9)$, this case is similar with the previous two cases.

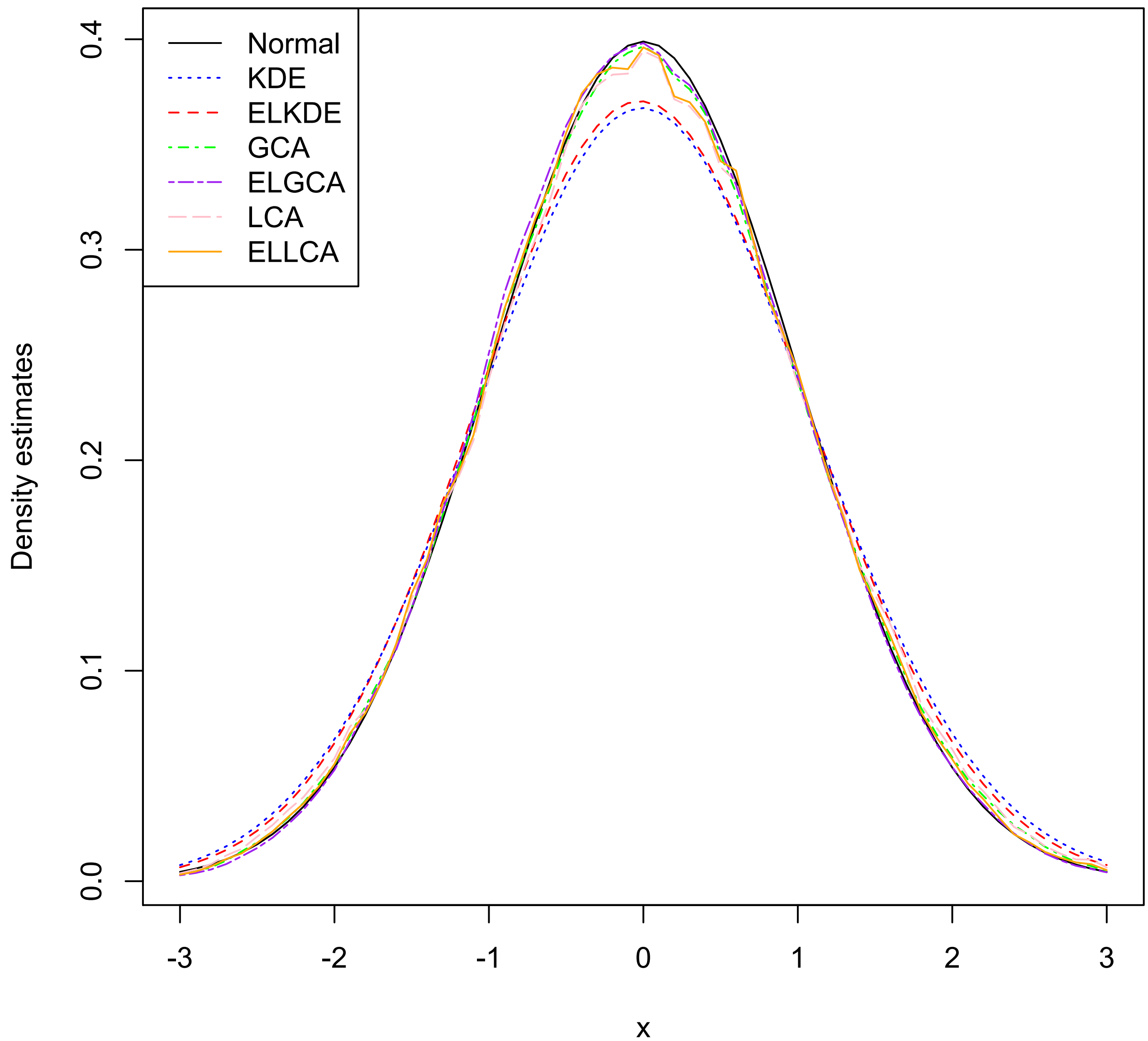
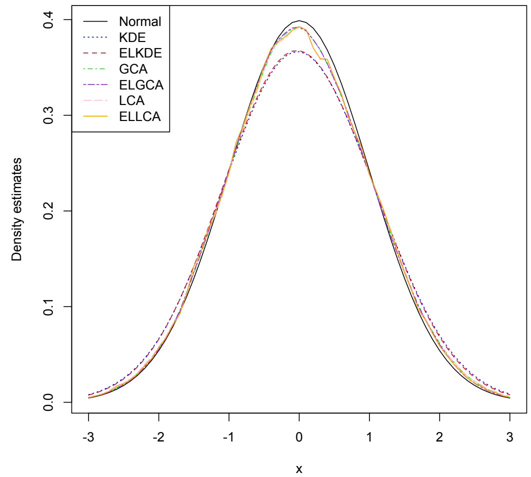
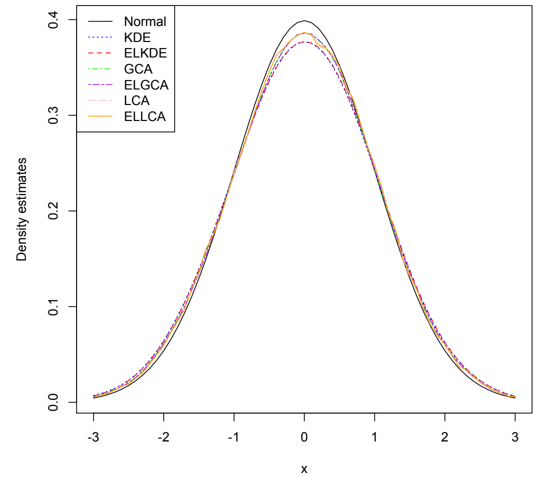


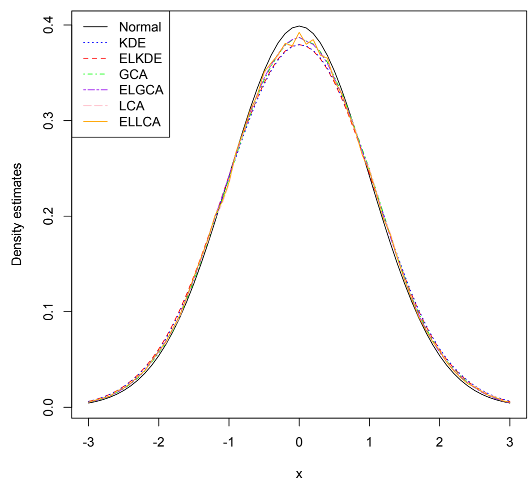

**Figure 7:** Kernel estimates from standard normal distribution for sample size n=15



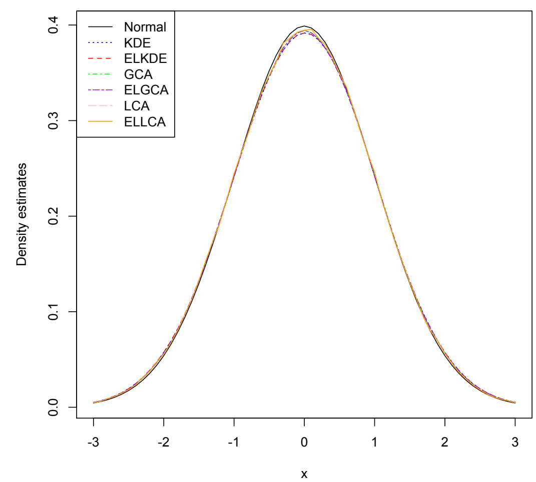

**(a)** n=25




**(b)** n=50




**(c)** n=100




**(d)** n=500

**Figure 8:** Density estimate from normal distribution for different sample size

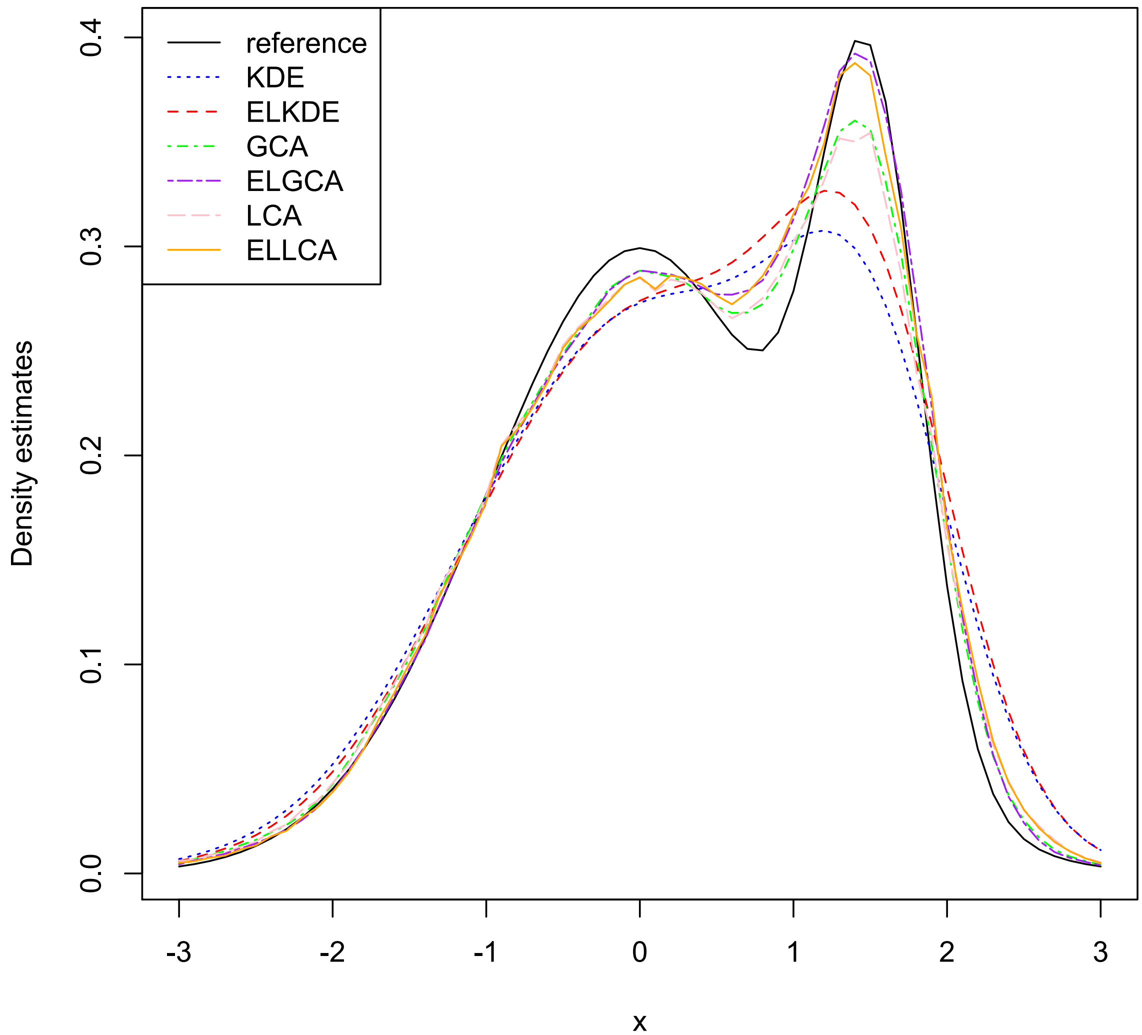
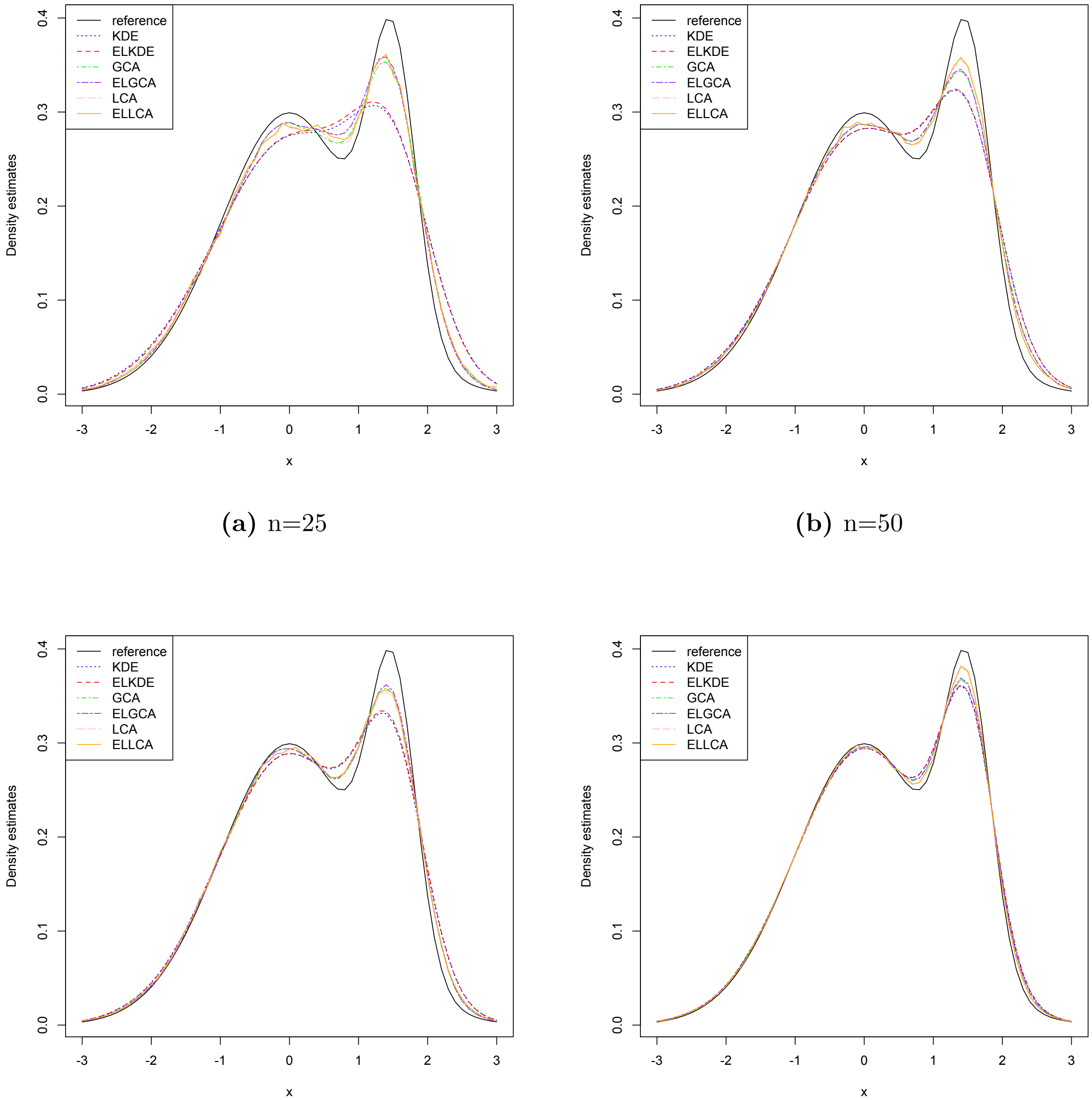



**Figure 9:** Kernel density estimation from $0.75N(0,1) + 0.25N(1.5, 4/9)$ for sample size n=15




**(a)** n=25

**(b)** n=50

**(c)** n=100

**(d)** n=500

**Figure 10:** Kernel density estimation from $0.75N(0,1) + 0.25N(1.5, 4/9)$ for different sample size

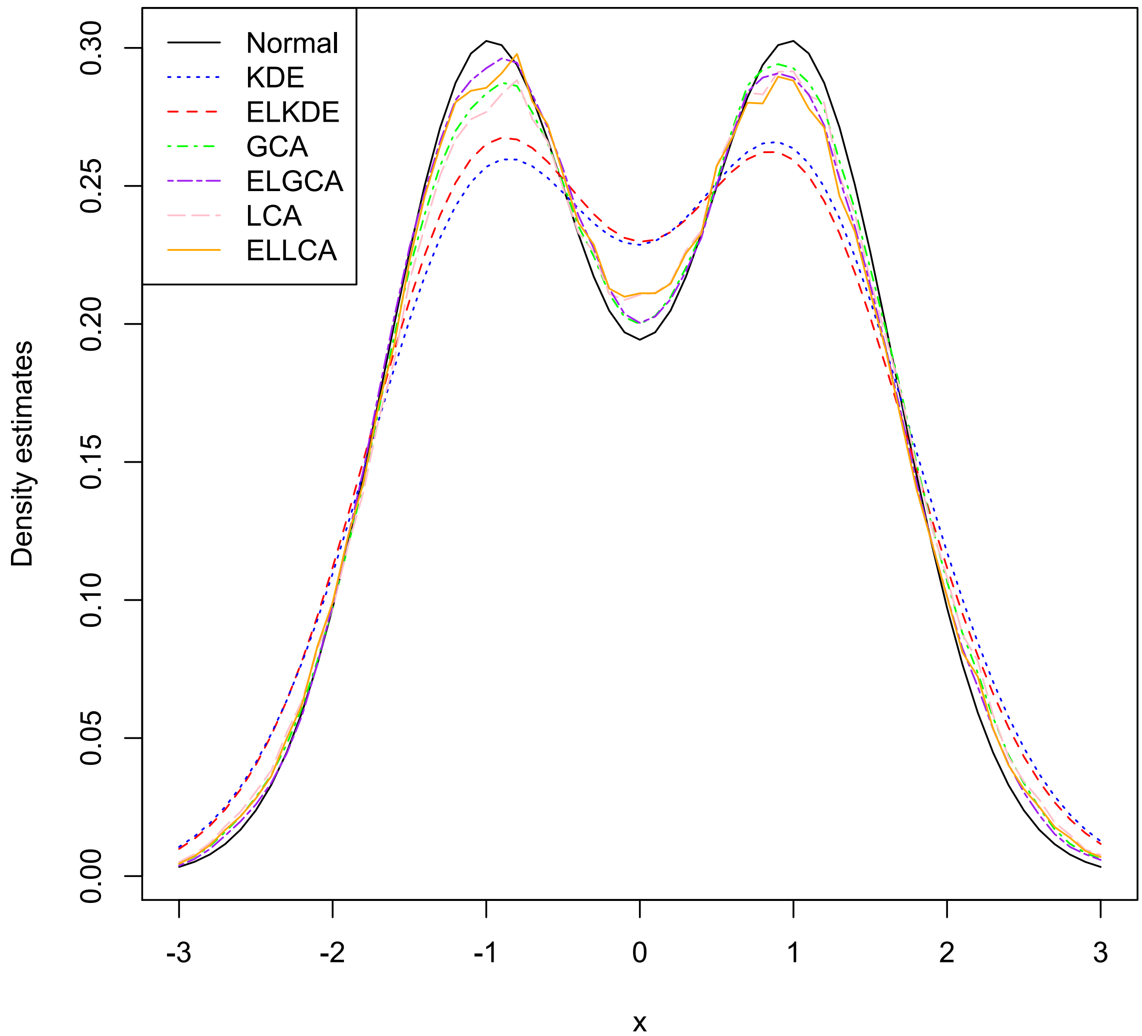
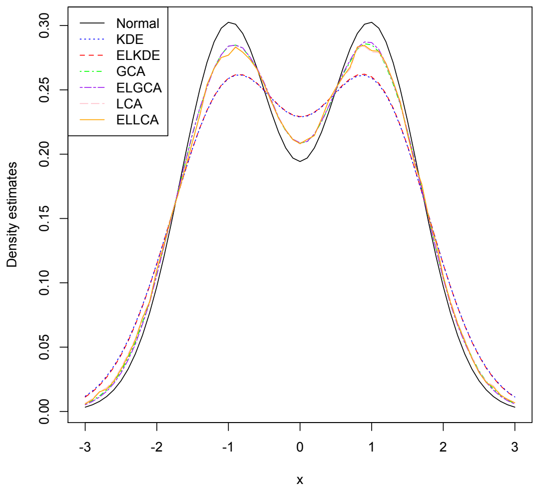
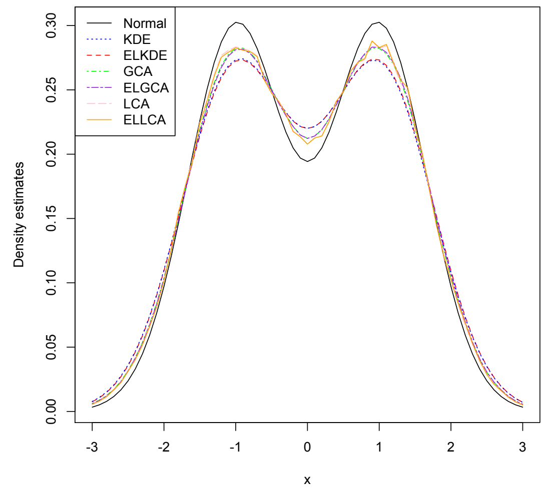


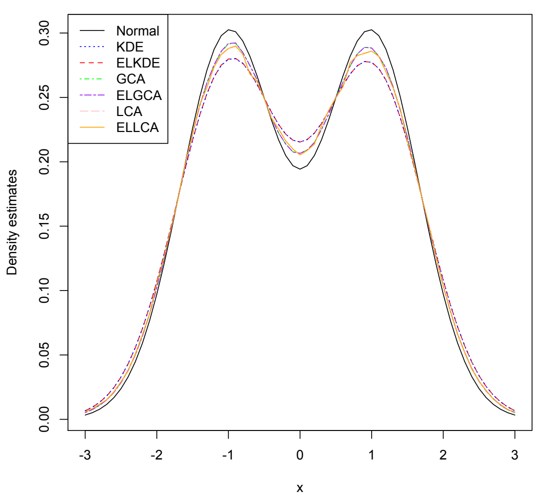

**Figure 11:** Kernel estimations from $0.5N(-1, 4/9) + 0.5N(1, 4/9)$ for sample size n=15



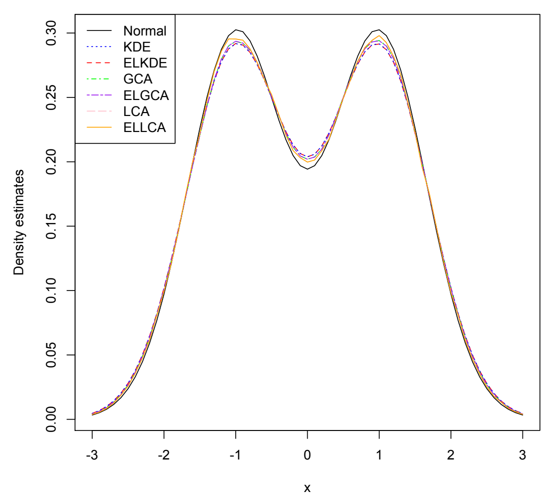

**(a)** n=25




**(b)** n=50




**(c)** n=100




**(d)** n=500

**Figure 12:** Density estimate from normal distribution for different sample size

## 3.6 Appendix

### 3.6.1 Proof Theorem 3.2.1

**Proof** First we show equations (3.3) and (3.4).

$$\begin{aligned}
\mathrm{E}\frac{1}{h}K\left(\frac{x-X_i}{ih}\right) &= \int \frac{1}{h}K\left(\frac{x-y}{ih}\right)f(y)dy \\
&= i\int K(\mu)f(x-\mu ih)d\mu \\
&= i\int K(\mu)\{f(x)-ih\mu f'(x)+\frac{i^2h^2\mu^2}{2}f''(x)+o(i^2h^2)\}d\mu \\
&= if(x)+\frac{i^3h^2}{2}\mu_2(K)f''(x)+o(i^3h^2).
\end{aligned}$$

Then,

$$\begin{aligned}
\mathrm{E}\hat{f}_{GCA}(x) &= \mathrm{E}\frac{2}{n(n+1)h}\sum_{i=1}^{n}K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{2}{n(n+1)}\sum_{i=1}^{n}\mathrm{E}\frac{1}{h}K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{2}{n(n+1)}\sum_{i=1}^{n}\left(if(x)+\frac{i^3h^2}{2}\mu_2(K)f''(x)\right) \\
&= f(x)+\frac{1}{4}n^2h^2\mu_2(K)f''(x)+o(n^2h^2).
\end{aligned}$$

So equation (3.3) is proved.

$$\begin{aligned}
\mathrm{E}\hat{f}_{LCA}(x) &= \mathrm{E}\frac{1}{nh}\sum_{i=1}^{n}\frac{1}{i}K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}\mathrm{E}\frac{1}{h}K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}\{if(x)+\frac{i^3h^2}{2}\mu_2(K)f''(x)+o(i^3h^2)\} \\
&= f(x)+\frac{1}{6}n^2h^2\mu_2(K)f''(x)+o(n^2h^2)
\end{aligned}$$

Equation (3.4) is proved.

Next we calculate the Variance for these two new methods.

First,

$$
\begin{aligned}
&\mathrm{E}\frac{1}{h^2}K^2\left(\frac{x-X_i}{ih}\right)\\
&=\frac{1}{h^2}\int K^2\left(\frac{x-y}{ih}\right)f(y)dy\\
&=\frac{1}{h^2}\int K^2(\mu)f(x-\mu ih)ihd\mu\\
&=\frac{i}{h}\int K^2(\mu)\{f(x)-\mu ihf'(x)+\frac{i^2h^2\mu^2}{2}f''(x)\}d\mu\\
&=\frac{i}{h}\{R(K)f(x)+\frac{i^2}{2}h^2f''(x)\int K^2(\mu)\mu^2d\mu\}.
\end{aligned}
$$

So,

$$
\begin{aligned}
&\mathrm{Var}(\frac{1}{h}K\left(\frac{x-X_i}{ih}\right))\\
&=\mathrm{E}\{\frac{1}{h}K\left(\frac{x-X_i}{ih}\right)\}^2-\{\mathrm{E}\frac{1}{h}K\left(\frac{x-X_i}{ih}\right)\}^2\\
&=\frac{i}{h}\{R(K)f(x)+\frac{i^2}{2}h^2f''(x)\int K^2(\mu)\mu^2d\mu\}-\{if(x)+\frac{i^3h^2}{2}\mu_2(K)f''(x)\}^2\\
&=\frac{i}{h}R(K)f(x)+\frac{i^3}{2}hf''(x)\int K^2(\mu)\mu^2d\mu-i^2f^2(x)-\frac{i^4h^2}{2}f(x)\mu_2(K),
\end{aligned}
$$

hence,

$$
\begin{aligned}
\mathrm{Var}\hat{f}_{GCA}(x)=&\frac{4}{n^2(n+1)^2}\sum_{i=1}^{n}\mathrm{Var}\frac{1}{h}K\left(\frac{x-X_i}{ih}\right)\\
=&\frac{2}{n^2h}R(K)f(x)-\frac{2}{3n}f^2(x)+o(h,n^{-1}),
\end{aligned}
$$

equation (3.5) is proved,

$$\begin{aligned}\text{Var}\{\hat{f}_{LCA}(x)\} &= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}\frac{1}{ih}K\left(\frac{x-X_i}{ih}\right)\\ =&\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{i^2}\{\frac{i}{h}\{R(K)f(x)+\frac{i^3}{2}hf''(x)\int K^2(\mu)\mu^2d\mu - i^2f^2(x) - \frac{i^4h^2}{2}f(x)\mu_2(K)\}\\ =&\frac{\sum_{i=1}^{n}\frac{1}{i}}{n^2h}R(K)f(x) - \frac{1}{n}f^2(x) + o(n^{-1},h),\end{aligned}$$

and equation (3.6) is proved.

This completes the proof of Theorem 3.2.1.

### 3.6.2 Proof Theorem 3.2.3

**Proof** From (3.14), the first term is independent of data. Clearly,

$$\text{E}\hat{\psi}_r(h)_{GCA} = \frac{2K^{(r)}(0)}{n(n+1)h^{r+1}}\sum_{i=1}^{n}(\frac{1}{j})^r + \frac{2}{n^2(n+1)h}\sum_{i\neq j}^{n}\text{E}(\frac{1}{jh})^rK^{(r)}\left(\frac{X_i-X_j}{jh}\right),$$

and

$$\begin{aligned}\text{E}(\frac{1}{jh})^rK^{(r)}\left(\frac{X_i-X_j}{jh}\right) &= \iint(\frac{1}{jh})^rK^{(r)}\left(\frac{x-y}{jh}\right)f(x)f(y)dxdy\\ &= \iint K\left(\frac{x-y}{jh}\right)f^{(r)}(x)f(y)dxdy\\ &= jh\int f^{(r)}(x)f(x)dx + \frac{j^3h^3}{2}\mu_2(K)\int f''(x)f^r(x)dx.\end{aligned}$$

So

$$\text{E}\hat{\psi}_r(h)_{GCA} = \frac{2K^{(r)}(0)}{n(n+1)h^{r+1}}\sum_{i=1}^{n}(\frac{1}{j})^r + \frac{n-1}{n}\psi_r + \frac{n^2h^2}{4}\mu_2(K)\psi_{r+2} + o(n^2h^2),$$

and

$$\mathrm{E}\hat{\psi}_r(h)_{LCA} = \frac{K^{(r)}(0)}{n^2h^{r+1}}\sum_{i=1}^{n}(\frac{1}{j})^{r+1} + \frac{n-1}{n}\psi_r + \frac{n^2h^2}{6}\mu_2(K)\psi_{r+2} + o(n^2h^2),$$

where $\psi_{r+s} = \int f^{(r)}(x)f^{(s)}(x)dx$. The first equation in the theorem is proved. Now we need to find variance of $\hat{\psi}_r(h)_{GCA}$.

$$\mathrm{Var}\hat{\psi}_r(h) = \frac{4}{n^4(n+1)^2h^2}\mathrm{Var}\sum_{i\neq j}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i - X_j}{jh}\right)$$

First, we need to find $\mathrm{E}[\sum_{i\neq j}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i-X_j}{jh}\right)]^2$. There are some different cases:

$$\mathrm{E}[\sum_{i\neq j}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i - X_j}{jh}\right)]^2 = E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + E_7,$$

where

$E_1$ when $i = k \neq j = l$; $E_2$ when $i = l \neq j = k$; $E_3$ when $i = k \neq j \neq l$;

$E_4$ when $i = l \neq j = k$; $E_5$ when $i \neq j = k \neq l$; $E_6$ when $i \neq j = l \neq k$;

$E_7$ when $i \neq j \neq k \neq l$;

$$\begin{aligned}
E_1 =& \mathrm{E}\sum_{k=i\neq j=l}(\frac{1}{jh})^{2r}[K^{(r)}\left(\frac{X_i - X_j}{jh}\right)]^2 \\
=& \sum_{i\neq j}(\frac{1}{jh})^{2r}\mathrm{E}[K^{(r)}\left(\frac{X_i - X_j}{jh}\right)]^2 \\
=& \sum_{i\neq j}(\frac{1}{jh})^{2r}\iint[K^{(r)}\left(\frac{x-y}{jh}\right)]^2 f(x)f(y)dxdy \\
=& \sum_{i\neq j}(\frac{1}{jh})^{2r-1}\iint[K^{(r)}(\mu)]^2 f(y+\mu jh)f(y)dyd\mu \\
=& \sum_{i\neq j}(\frac{1}{jh})^{2r-1}R(f)R(K^{(r)}) \\
=& o(n^{-1}),
\end{aligned}$$

$$
\begin{aligned}
E_2 = \mathrm{E} & \sum_{l=i\neq j=k} (\frac{1}{jh})^r (\frac{1}{ih})^r K^{(r)} \left( \frac{X_i - X_j}{jh} \right) K^{(r)} \left( \frac{X_j - X_i}{ih} \right) \\
= & \sum_{i \neq j} (\frac{1}{jh})^r (\frac{1}{ih})^r \mathrm{E} K^{(r)} \left( \frac{X_i - X_j}{jh} \right) K^{(r)} \left( \frac{X_j - X_i}{ih} \right) \\
= & \sum_{i \neq j} (\frac{1}{jh})^r (\frac{1}{ih})^r \iint K^{(r)} \left( \frac{x - y}{jh} \right) K^{(r)} \left( \frac{y - x}{ih} \right) f(x) f(y) dx dy \\
= & \sum_{i \neq j} (\frac{1}{jh})^r (\frac{1}{ih})^r h \iint K^{(r)} (\frac{\mu}{j}) K^{(r)} (\frac{-\mu}{i}) f(y + \mu h) f(y) dy d\mu \\
= & o(n^{-1}),
\end{aligned}
$$

$$
\begin{aligned}
E_3 = \mathrm{E} & \sum_{k=i\neq j \neq l} (\frac{1}{jh})^r (\frac{1}{lh})^r K^{(r)} \left( \frac{X_i - X_j}{jh} \right) K^{(r)} \left( \frac{X_i - X_l}{lh} \right) \\
= & \sum_{i \neq j \neq l} (\frac{1}{jh})^r (\frac{1}{lh})^r \mathrm{E} K^{(r)} \left( \frac{X_i - X_j}{jh} \right) K^{(r)} \left( \frac{X_i - X_l}{lh} \right) \\
= & \sum_{i \neq j \neq l} (\frac{1}{jh})^r (\frac{1}{lh})^r \iiint K^{(r)} \left( \frac{x - y}{jh} \right) K^{(r)} \left( \frac{x - z}{lh} \right) f(x) f(y) f(z) dx dy dz \\
= & \sum_{i \neq j \neq l} \int f(x) [\int (\frac{1}{jh})^r K^{(r)} \left( \frac{x - y}{jh} \right) f(y) dy \int (\frac{1}{lh})^r K^{(r)} \left( \frac{x - z}{lh} \right) f(z) dz] dx \\
= & \sum_{i \neq j \neq l} \int f(x) [\int K \left( \frac{x - y}{jh} \right) f^{(r)}(y) dy \int K \left( \frac{x - z}{lh} \right) f^{(r)}(z) dz] dx \\
= & \sum_{i \neq j \neq l} jlh^2 \int f(x) [\int K(\mu) f^{(r)}(x - \mu jh) d\mu \int K(\nu) f^{(r)}(x - \nu lh) dz] d\nu \\
= & \frac{1}{12} n^2 (n-1)(n-2)(3n-1) h^2 \int f(x) [f^{(r)}(x)]^2 dx,
\end{aligned}
$$

$$E_4 = \mathrm{E}\sum_{l=i\neq j\neq k}(\frac{1}{jh})^r(\frac{1}{ih})^r K^{(r)}\left(\frac{X_i-X_j}{jh}\right)K^{(r)}\left(\frac{X_k-X_i}{ih}\right)$$

$$=\sum_{i\neq j\neq k}(\frac{1}{jh})^r(\frac{1}{ih})^r \mathrm{E}K^{(r)}\left(\frac{X_i-X_j}{jh}\right)K^{(r)}\left(\frac{X_k-X_i}{ih}\right)$$

$$=\sum_{i\neq j\neq k}(\frac{1}{jh})^r(\frac{1}{ih})^r\iiint K^{(r)}\left(\frac{x-y}{jh}\right)K^{(r)}\left(\frac{z-x}{ih}\right)f(x)f(y)f(z)dxdydz$$

$$=\sum_{i\neq j\neq k}\int f(x)[\int(\frac{1}{jh})^r K^{(r)}\left(\frac{x-y}{jh}\right)f(y)dy\int(\frac{1}{ih})^r K^{(r)}\left(\frac{z-x}{ih}\right)f(z)dz]dx$$

$$=\sum_{i\neq j\neq k}\int f(x)[\int K\left(\frac{x-y}{jh}\right)f^{(r)}(y)dy\int K\left(\frac{z-x}{ih}\right)f^{(r)}(z)dz]dx$$

$$=\sum_{i\neq j\neq k} ijh^2\int f(x)[\int K(\mu)f^{(r)}(x-\mu jh)d\mu\int K(\nu)f^{(r)}(x+\nu ih)d\nu]dx$$

$$=\frac{1}{12}n^2(n-1)(n-2)(3n-1)h^2\int f(x)[f^{(r)}(x)]^2dx,$$

$$E_5 = \mathrm{E}\sum_{l\neq i\neq j=k}(\frac{1}{jh})^r(\frac{1}{lh})^r K^{(r)}\left(\frac{X_i-X_j}{jh}\right)K^{(r)}\left(\frac{X_j-X_l}{lh}\right)$$

$$=\sum_{l\neq i\neq j=k}(\frac{1}{jh})^r(\frac{1}{lh})^r \mathrm{E}K^{(r)}\left(\frac{X_i-X_j}{jh}\right)K^{(r)}\left(\frac{X_j-X_l}{lh}\right)$$

$$=\sum_{l\neq i\neq j=k}(\frac{1}{jh})^r(\frac{1}{lh})^r\iiint K^{(r)}\left(\frac{x-y}{jh}\right)K^{(r)}\left(\frac{y-z}{lh}\right)f(x)f(y)f(z)dxdydz$$

$$=\sum_{l\neq i\neq j=k}\int f(y)[\int(\frac{1}{jh})^r K^{(r)}\left(\frac{x-y}{jh}\right)f(x)dx\int(\frac{1}{lh})^r K^{(r)}\left(\frac{y-z}{lh}\right)f(z)dz]dy$$

$$=\sum_{l\neq i\neq j=k}\int f(y)[\int K\left(\frac{x-y}{jh}\right)f^{(r)}(x)dx\int K\left(\frac{y-z}{lh}\right)f^{(r)}(z)dz]dy$$

$$=\sum_{l\neq i\neq j=k} ijh^2\int f(y)[\int K(\mu)f^{(r)}(y+\mu jh)d\mu\int K(\nu)f^{(r)}(y-\nu lh)d\nu]dy$$

$$=\frac{1}{12}n^2(n-1)(n-2)(3n-1)h^2\int f(y)[f^{(r)}(y)]^2dy,$$

$$
\begin{aligned}
E_6 =& \mathrm{E}\sum_{k\neq i\neq j=l}(\frac{1}{jh})^r(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i-X_j}{jh}\right)K^{(r)}\left(\frac{X_k-X_j}{jh}\right)\\
=&\sum_{k\neq i\neq j=l}(\frac{1}{jh})^r(\frac{1}{jh})^r \mathrm{E}K^{(r)}\left(\frac{X_i-X_j}{jh}\right)K^{(r)}\left(\frac{X_k-X_j}{jh}\right)\\
=&\sum_{k\neq i\neq j=l}(\frac{1}{jh})^r(\frac{1}{jh})^r\iiint K^{(r)}\left(\frac{x-y}{jh}\right)K^{(r)}\left(\frac{z-y}{jh}\right)f(x)f(y)f(z)dxdydz\\
=&\sum_{k\neq i\neq j=l}\int f(y)[\int(\frac{1}{jh})^r K^{(r)}\left(\frac{x-y}{jh}\right)f(x)dx\int(\frac{1}{jh})^r K^{(r)}\left(\frac{z-y}{jh}\right)f(z)dz]dy\\
=&\sum_{k\neq i\neq j=l}\int f(y)[\int K\left(\frac{x-y}{jh}\right)f^{(r)}(x)dx\int K\left(\frac{y-z}{jh}\right)f^{(r)}(z)dz]dy\\
=&\sum_{k\neq i\neq j=l}j^2h^2\int f(y)[\int K(\mu)f^{(r)}(y+\mu jh)d\mu\int K(\nu)f^{(r)}(y-\nu jh)d\nu]dy\\
=&\frac{1}{6}n(n-1)^2(n-2)(2n-3)h^2\int f(y)[f^{(r)}(y)]^2dy,
\end{aligned}
$$

and

$$
\begin{aligned}
E_7 =& \mathrm{E}\sum_{i\neq j\neq k\neq l}(\frac{1}{jh})^r(\frac{1}{lh})^r K^{(r)}\left(\frac{X_i-X_j}{jh}\right)K^{(r)}\left(\frac{X_k-X_l}{lh}\right)\\
=&\sum_{i\neq j\neq k\neq l}\mathrm{E}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i-X_j}{jh}\right)\mathrm{E}(\frac{1}{lh})^r K^{(r)}\left(\frac{X_k-X_l}{lh}\right)\\
=&\sum_{i\neq j\neq k\neq l}\iint(\frac{1}{jh})^r K^{(r)}\left(\frac{x-y}{jh}\right)f(x)f(y)dxdy\\
&\iint(\frac{1}{lh})^r K^{(r)}\left(\frac{z-k}{jh}\right)f(z)f(k)dzdk\\
=&\sum_{i\neq j\neq k\neq l}\iint K\left(\frac{x-y}{jh}\right)f(x)f^{(r)}(y)dxdy\iint K\left(\frac{z-k}{lh}\right)f(z)f^{(r)}(k)dzdk\\
=&\sum_{i\neq j\neq k\neq l}[jhK(\mu)f(x)f^{(r)}(x-\mu jh)dxd\mu][lhK(\nu)f(z)f^{(r)}(z-lh\nu)dzd\nu]\\
=&\sum_{k\neq i\neq j\neq l}jlh^2[\int f(x)f^{(r)}(x)dx]^2\\
=&n(n-1)[\frac{1}{4}n^4-2n^3+\frac{23}{4}n^2-7n+3]h^2[\int f(y)f^{(r)}(y)dy]^2
\end{aligned}
$$

Then Var $\sum_{i\neq j}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i-X_j}{jh}\right)$

$$
\begin{aligned}
&\text{Var}\sum_{i\neq j}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i - X_j}{jh}\right)\\
&= E_1 + \cdots + E_7 - [\text{E}\sum_{i\neq j}(\frac{1}{jh})^r K^{(r)}\left(\frac{X_i - X_j}{jh}\right)]^2\\
=&\frac{1}{4}n^2(n-1)(n-2)(3n-1)\int f(y)[f^{(r)}(y)]^2 dy\\
+&n(n-1)[\frac{1}{4}n^4 - 2n^3 + \frac{23}{4}n^2 - 7n + 3]h^2[\int f(y)f^{(r)}(y)dy]^2\\
-&\frac{n^4(n-1)^2}{4}h^2[\int f^{(r)}(x)f(x)dx]^2\\
=&\frac{3}{4}n^5h^2\int f(y)[f^{(r)}(y)]^2 dy - 2n^5h^2[\int f(y)f^{(r)}(y)dy]^2.
\end{aligned}
$$

Then it is

$$
\text{Var}\hat{\psi}_r(h)_{GCA} = \frac{3}{n}\int f(y)[f^{(r)}(y)]^2 dy - \frac{8}{n}[\int f(y)f^{(r)}(y)dy]^2.
$$

Taking the same approach to $\text{Var}\hat{\psi}_r(h)_{LCA}$, we have

$$
\text{Var}\hat{\psi}_r(h)_{LCA} = \frac{2}{nh^{2r+1}}\psi_0 R(K^{(r)}) + \frac{4}{n}\{\int f(y)[f^{(r)}(y)]^2 dy] - \psi_r\}.
$$

### 3.6.3 Proof Theorem 3.4.1

**Proof** Using a similar proof to Chen (1997), we can get

$$\begin{aligned}
\hat{f}_{el.GCA}(x) &= \frac{2}{(n+1)h}\sum_{i=1}^{n} p_i K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{2}{n(n+1)h}\sum_{i=1}^{n}\frac{1}{1+\boldsymbol{\lambda}^T g(X_i)} K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{2}{n(n+1)h}\sum_{i=1}^{n}[1-\boldsymbol{\lambda}^T g(X_i)+\lambda^T g(X_i)g^T(X_i)\boldsymbol{\lambda}+O_p(n^{-1})] \\
& K\left(\frac{x-X_i}{ih}\right) \\
&= \hat{f}_{GCA}(x)-\boldsymbol{\lambda}^T T_1+\boldsymbol{\lambda}^T T_2\boldsymbol{\lambda}+O_p(n^{-1}),
\end{aligned}$$

where the vector $T_1$ and $q \times q$ matrix $T_2$ are defined by

$$\begin{aligned}
T_1 &= \frac{2}{n(n+1)h}\sum_{1}^{n}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right), \\
T_2 &= \frac{2}{n(n+1)h}\sum_{1}^{n}\mathbf{g}(X_i)\mathbf{g}(X_i)^T K\left(\frac{x-X_i}{ih}\right).
\end{aligned}$$

A Taylor expansion for $\boldsymbol{\lambda}$, similar to those given in Chen (1997), is

$$\boldsymbol{\lambda} = \Sigma^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(X_i)+O_p(n^{-1}), \tag{3.34}$$

where $\Sigma = \text{cov}(g_l(X), g_m(X))$.

$$
\begin{aligned}
E(\boldsymbol{\lambda}^T T_1) &= E[\Sigma^{-1}\frac{1}{n}\sum_{j=1}^{n}\mathbf{g}(X_j)]^T\frac{2}{n(n+1)h}\sum_{i=1}^{n}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\\
&= \frac{2}{n^2(n+1)h}E\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{g}(X_j)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\\
&= \frac{2}{n^2(n+1)h}E\sum_{i=1}^{n}\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\\
&= \frac{1}{n}\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)+o(n^{-1}).
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}\left(\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\right)\\
&=\int\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)K\left(\frac{y-x}{ih}\right)f(x)dx\\
&=\int\mathbf{g}(y-\mu ih)^T\Sigma^{-1}\mathbf{g}(y-\mu ih)K(\mu)f(y-\mu ih)ihd\mu\\
&=ih\int\{\mathbf{g}(y)-\mu ih\mathbf{g}'(y)+\frac{\mu^2h^2i^2}{2}\mathbf{g}''(y)\}^T\Sigma^{-1}\\
&\quad\{\mathbf{g}(y)-\mu ih\mathbf{g}'(y)+\frac{\mu^2h^2i^2}{2}\mathbf{g}''(y)\}\\
&\qquad K(\mu)\{f(y)-\mu ihf'(y)+\frac{\mu^2h^2i^2}{2}f''(y)\}d\mu\\
&=ih\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)+o(n^{-1}).
\end{aligned}
$$

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\lambda}^T T_2 \boldsymbol{\lambda}) = &\frac{2}{n^3(n+1)h} \mathrm{E} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{g}(X_j)^T \Sigma^{-1} \mathbf{g}(X_i) \\
&\mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_k) K\left(\frac{x - X_i}{ih}\right) \\
&= \frac{2}{n^3(n+1)h} \sum_{i=1}^{n} \mathrm{E}\left(\mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_i) \mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_i) K\left(\frac{x - X_i}{ih}\right)\right) \\
&= \frac{2}{n^2(n+1)h} \sum_{i=1}^{n} \mathrm{E}\left(\mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_i) K\left(\frac{x - X_i}{ih}\right)\right) \\
&= \frac{1}{n} \mathbf{g}(y)^T \Sigma^{-1} \mathbf{g}(y) f(y) + o(n^{-1}).
\end{aligned}$$

Thus, by delta method,

$$\mathrm{E}(\hat{f}_{el.GCA}) = \mathrm{E}(\hat{f}_{GCA}) + o(n^{-1}). \tag{3.35}$$

To derive the variance of $\hat{f}_{el.GCA}$,

$$\begin{aligned}
&\hat{f}_{el.GCA}(x)^2 \\
&= \frac{4}{n^2(n+1)^2h^2} \sum_{i,j}^{n} \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{g}(X_i)} \frac{1}{1 + \boldsymbol{\lambda}^T \mathbf{g}(X_j)} K\left(\frac{x - X_i}{ih}\right) K\left(\frac{x - X_j}{jh}\right) \\
&= \hat{f}_{GCA}(x)^2 - 2\boldsymbol{\lambda}^T T_1 \hat{f}_{GCA}(x) + 2\boldsymbol{\lambda}^T T_2 \boldsymbol{\lambda} \hat{f}_{GCA}(x) + \boldsymbol{\lambda}^T T_1 T_1 \boldsymbol{\lambda} + o(n^{-1}),
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}(\boldsymbol{\lambda}^T T_1 \hat{f}_{GCA}(x)) \\
&= \frac{4}{n^3(n+1)^2h^2}\mathrm{E}\left(\sum_{i,j,k=1}^{n} \mathbf{g}(X_j)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)K\left(\frac{x-X_k}{kh}\right)\right) \\
&= \frac{4}{n^3(n+1)^2h^2}\sum_{i=j\neq k}^{n}\mathrm{E}\left(\mathbf{g}(X_j)^T\Sigma^{-1}g(X_i)K\left(\frac{x-X_i}{ih}\right)\right)\mathrm{E}\left(K\left(\frac{x-X_k}{kh}\right)\right) \\
&+ \frac{4}{n^3(n+1)^2h^2}\sum_{j=k\neq i}^{n}\mathrm{E}\left(\mathbf{g}(X_j)^T K\left(\frac{x-X_k}{kh}\right)\Sigma^{-1}\right)\mathrm{E}\left(\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\right) \\
&+ \frac{4}{n^3(n+1)^2h^2}\sum_{i=1}^{n}\mathrm{E}\left(\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)^2\right) \\
&= \frac{8}{n^3(n+1)^2}\sum_{i\neq k}^{n} ik\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)^2dy \\
&+ \frac{4}{n^3(n+1)^2h}\sum_{i=1}^{n} i\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)dyR(K) + o(n^{-1}) \\
&= \frac{2}{n}\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)^2dy + o(n^{-1}),
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}(\boldsymbol{\lambda}^T T_2 \boldsymbol{\lambda} \hat{f}_{GCA}(x)) \\
&= \frac{4}{n^4(n+1)^2h^2}\sum_{i,j,l,k}\mathrm{E}\mathbf{g}(X_j)^T\Sigma^{-1}\mathbf{g}(X_i)\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_l) \\
&\quad K\left(\frac{x-X_i}{ih}\right)K\left(\frac{x-X_k}{kh}\right) \\
&= \frac{4}{n^3(n+1)^2h^2}\mathrm{E}\left(\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)K\left(\frac{x-X_k}{kh}\right)\right) \\
&= \frac{1}{n}\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)^2dy + o(n^{-1}),
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\lambda}^T T_1 T_1^T \boldsymbol{\lambda}) =& \frac{4}{n^4(n+1)^2h^2}\sum_{i,j,k,l=1}^{n}\mathrm{E}\mathbf{g}(X_j)^T\Sigma^{-1}\mathbf{g}(X_i)\mathbf{g}(X_k)^T\Sigma^{-1}\mathbf{g}(X_l) \\
& K\left(\frac{x-X_i}{ih}\right)K\left(\frac{x-X_k}{kh}\right) \\
&= \frac{1}{n}\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)^2 + o(n^{-1}).
\end{aligned}$$

Thus, Taylor expansion for $\boldsymbol{\lambda}$, and additional use of the delta method, can get

$$\mathrm{E}(\hat{f}_{el.GCA}(x)^2) = \mathrm{E}(\hat{f}_{GCA}(x)^2) - \frac{1}{n}\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)^2 + o(n^{-1}). \tag{3.36}$$

So $\mathrm{Var}\hat{f}_{el.GCA}(x)$

$$\mathrm{Var}(\hat{f}_{el.GCA}(x)) = \mathrm{Var}(\hat{f}_{GCA}(x)) - \frac{1}{n}\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)^2 + o(n^{-1}), \tag{3.37}$$

equation (3.5) is proved.

Similar proof of (3.3) and (3.5), then

$$\begin{aligned}
\hat{f}_{el.LCA} &= \frac{1}{h}\sum_{i=1}^{n}\frac{p_i}{i}K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{i}\frac{1}{1+\boldsymbol{\lambda}^T\mathbf{g}(X_i)}K\left(\frac{x-X_i}{ih}\right) \\
&= \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{i}\{1-\boldsymbol{\lambda}^T\mathbf{g}(X_i)+\boldsymbol{\lambda}^T\mathbf{g}(X_i)\mathbf{g}(X_i)^T\boldsymbol{\lambda}+o(n^{-1})\}K\left(\frac{x-X_i}{ih}\right) \\
&= \hat{f}_{LCA} - \boldsymbol{\lambda}^T L_1 + \boldsymbol{\lambda}^T L_2 \boldsymbol{\lambda} + o(n^{-1}),
\end{aligned}$$

where the vector $L_1$ and $q \times q$ matrix $L_2$ are defined by

$$\begin{aligned}
L_1 &= \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{i}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right), \\
L_2 &= \frac{1}{nh}\sum_{i=1}^{n}\frac{1}{i}\mathbf{g}(X_i)\mathbf{g}(X_i)^T K\left(\frac{x-X_i}{ih}\right).
\end{aligned}$$

Plug in the $\boldsymbol{\lambda}$ at equation (3.34), then

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\lambda}^T L_1) &= \mathrm{E}\left(\Sigma^{-1}\frac{1}{n}\sum_{j=1}^{n}\mathbf{g}(X_j)]^T\frac{1}{nh}\sum_{i=1}^{n}\frac{1}{i}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\right) \\
&= \frac{1}{n^2h}\mathrm{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{g}(X_j)^T\Sigma^{-1}\frac{1}{i}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\right) \\
&= \frac{1}{n^2h}\mathrm{E}\left(\sum_{i=1}^{n}\frac{1}{i}\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\right) \\
&= \frac{1}{n}\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)f(x)+o(n^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
&\mathrm{E}\left(\frac{1}{ih}g(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)K\left(\frac{x-X_i}{ih}\right)\right) \\
&= \int\frac{1}{i}\mathbf{g}(x)^T\Sigma^{-1}\mathbf{g}(x)K\left(\frac{y-x}{ih}\right)f(x)dx \\
&= \int\frac{1}{ih}\mathbf{g}(y-\mu ih)^T\Sigma^{-1}\mathbf{g}(y-\mu ih)K(\mu)f(y-\mu ih)ihd\mu \\
&= \int\{\mathbf{g}(y)-\mu ih\mathbf{g}'(y)+\frac{\mu^2h^2i^2}{2}\mathbf{g}''(y)\}^T\Sigma^{-1}\{\mathbf{g}(y)-\mu ih\mathbf{g}'(y)+\frac{\mu^2h^2i^2}{2}\mathbf{g}''(y)\} \\
&\qquad K(\mu)\{f(y)-\mu ihf'(y)+\frac{\mu^2h^2i^2}{2}f''(y)\}d\mu \\
&= \mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)+o(n^{-1}),
\end{aligned}$$

and

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{\lambda}^T L_2 \boldsymbol{\lambda}) \\
&= \frac{1}{n^3 h} \mathrm{E}\left( \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g(X_j)^T \Sigma^{-1} \mathbf{g}(X_i) \mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_k) \frac{1}{i} K\left(\frac{x - X_i}{ih}\right)\right) \\
&= \frac{1}{n^3 h} \sum_{i=1}^{n} \mathrm{E}\left( \mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_i) \mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_i) \frac{1}{i} K\left(\frac{x - X_i}{ih}\right)\right) \\
&= \frac{1}{n^2 h} \sum_{i=1}^{n} \mathrm{E}\left( \frac{1}{i} \mathbf{g}(X_i)^T \Sigma^{-1} \mathbf{g}(X_i) K\left(\frac{x - X_i}{ih}\right)\right) \\
&= \frac{1}{n} \mathbf{g}(y)^T \Sigma^{-1} \mathbf{g}(y) f(y) + o(n^{-1}).
\end{aligned}
$$

Thus, by delta method,

$$
\mathrm{E}(\hat{f}_{el.LCA}) = \mathrm{E}(\hat{f}_{LCA}) + o(n^{-1}). \tag{3.38}
$$

To derive the variance of $\hat{f}_{el.LCA}$,

$$
\begin{aligned}
&\hat{f}_{el.LCA}(x)^2 \\
&= \frac{1}{n^2 h^2} \sum_{i,j}^{n} \frac{1}{i(1 + \boldsymbol{\lambda}^T \mathbf{g}(X_i))} \frac{1}{j(1 + \boldsymbol{\lambda}^T \mathbf{g}(X_j))} K\left(\frac{x - X_i}{ih}\right) K\left(\frac{x - X_j}{jh}\right) \\
&= \hat{f}_{LCA}(x)^2 - 2\boldsymbol{\lambda}^T L_1 \hat{f}_{LCA}(x) + 2\boldsymbol{\lambda}^T L_2 \boldsymbol{\lambda} \hat{f}_{LCA}(x) + \boldsymbol{\lambda}^T L_1 L_1 \boldsymbol{\lambda} + o(n^{-1}),
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{\lambda}^T L_1 \hat{f}_{LCA}(x)) \\
&= \frac{1}{n^3h^2}\mathrm{E}\left(\sum_{i,j,k=1}^{n} \mathbf{g}(X_j)^T\Sigma^{-1}\mathbf{g}(X_i)\frac{1}{i}K\left(\frac{x-X_i}{ih}\right)\frac{1}{k}K\left(\frac{x-X_k}{kh}\right)\right) \\
&= \frac{1}{n^3h^2}\sum_{i=j\neq k}^{n}\mathrm{E}\left(\mathbf{g}(X_j)^T\Sigma^{-1}\mathbf{g}(X_i)\frac{1}{i}K\left(\frac{x-X_i}{ih}\right)E\frac{1}{k}K\left(\frac{x-X_k}{kh}\right)\right) \\
&+ \frac{1}{n^3h^2}\sum_{j=k\neq i}^{n}\mathrm{E}\left(\mathbf{g}(X_j)^T\frac{1}{k}K\left(\frac{x-X_k}{kh}\right)\Sigma^{-1}\right)\mathrm{E}\left(\mathbf{g}(X_i)\frac{1}{i}K\left(\frac{x-X_i}{ih}\right)\right) \\
&+ \frac{1}{n^3h^2}\sum_{i=1}^{n}\mathrm{E}\left(\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)\frac{1}{i^2}K\left(\frac{x-X_i}{ih}\right)^2\right) \\
&= \frac{8}{n^3}\sum_{i\neq k}^{n}\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)^2dy \\
&+ \frac{4}{n^3h}\sum_{i=1}^{n}\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)dyR(K)+o(n^{-1}) \\
&= \frac{2}{n}\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)^2dy+o(n^{-1}).
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{\lambda}^T L_2 \boldsymbol{\lambda}\hat{f}_{LCA}(x)) \\
&= \frac{1}{n^4h^2}\sum_{i,j,l,k}\mathrm{E}\mathbf{g}(X_j)^T\Sigma^{-1}\mathbf{g}(X_i)\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_l) \\
&K\left(\frac{x-X_i}{ih}\right)\frac{1}{k}K\left(\frac{x-X_k}{kh}\right) \\
&= \frac{1}{n^3h^2}\mathrm{E}\left(\mathbf{g}(X_i)^T\Sigma^{-1}\mathbf{g}(X_i)\frac{1}{i}K\left(\frac{x-X_i}{ih}\right)\frac{1}{k}K\left(\frac{x-X_k}{kh}\right)\right) \\
&= \frac{1}{n}\mathbf{g}(y)^T\Sigma^{-1}\mathbf{g}(y)f(y)^2dy+o(n^{-1}),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{\lambda}^T L_1 L_1^T \boldsymbol{\lambda}) =& \frac{1}{n^4 h^2} \sum_{i,j,k,l=1}^{n} \mathrm{E}\mathbf{g}(X_j)^T \Sigma^{-1} \mathbf{g}(X_i) \mathbf{g}(X_k)^T \Sigma^{-1} \mathbf{g}(X_l) \\
& \frac{1}{ik} K\left(\frac{x - X_i}{ih}\right) K\left(\frac{x - X_k}{kh}\right) \\
=& \frac{1}{n} \mathbf{g}(x)^T \Sigma^{-1} \mathbf{g}(x) f(x)^2 + o(n^{-1}).
\end{aligned}
$$

Thus, Taylor expansion for $\boldsymbol{\lambda}$, and additional use of the delta method, can get

$$
\mathrm{E}(\hat{f}_{el.LCA}(x)^2) = \mathrm{E}(\hat{f}_{LCA}(x)^2) - \frac{1}{n} \mathbf{g}(x)^T \Sigma^{-1} \mathbf{g}(x) f(x)^2 + o(n^{-1}).
$$

So $\mathrm{var}(\hat{f}_{el.LCA}(x))$

$$
\mathrm{Var}(\hat{f}_{el.LCA}(x)) = \mathrm{Var}(\hat{f}_{LCA}(x)) - \frac{1}{n} \mathbf{g}(x)^T \Sigma^{-1} \mathbf{g}(x) f(x)^2 + o(n^{-1}).
$$

This completes the proof of Theorem 3.4.1.

# 4 GCA and LGA Applications: Regression and CDF Estimation

## 4.1 Introduction

In this chapter we study some applications, namely regression and Cumulative Distribution Function (**CDF**) estimation based on **GCA** and **LGA** estimators. In the regression models, the goal is to estimate the regression function at a particular point by "local" fitting a $p^{th}$ degree polynomial to the data via weighted least squares. In this class, $p = 0$ as a special case with degree zero polynomials, that is local constants, like the Nadaraya-Watson estimator for usual kernel estimator. Another special case is a local linear estimator, corresponding to $p = 1$. We also see the mean squared error properties for $p = 0$ and $p = 1$ analogous to those of these new kernel density estimators. This means that most of the the ideas developed in the context of new density estimations can be easily transported to the context of regression. Another application is the estimate **CDF** that is based on these two new estimators.

## 4.2 Random Design Regression Model

We study a random design regression model when we observe bivariate samples $(X_1, Y_1), \cdots, (X_n, Y_n)$ of random pairs and assume that all are continuously distributed with a joint density $f(y, x)$. Let $f(y|x) = f(y, x)/f(x)$ be conditional density of $Y_i$ given $X_i$. The regression function for $Y_i$ on $X_i$ is

$$m(x) = \mathrm{E}(F_i | X_i = x)$$

in which case the model can be written as

$$Y_i = m(X_i) + \varepsilon_i \quad i = 1, \cdots, n \tag{4.1}$$

where

$$\mathrm{E}(\varepsilon_i | X_i) = 0 \quad \textit{and} \quad \mathrm{E}(\varepsilon_i^2 | X_i) = \sigma^2(x).$$

In the following sections, we introduce a local polynomial based on **GCA** and **LCA** estimators.

### 4.2.1 Local Polynomial Based on GCA Estimators

In this section, we investigate a local polynomial based on a **GCA** estimator that is defined by (3.2). Our task is to derive an explicit expression for the local polynomial based on the **GCA** estimator. For simplicity's sake we will assume that $f$ is supported on $[0, 1]$. Let $p$ be the degree of the polynomial being fit. At a point $x$, the estimator $\hat{m}(x; p, h)$ is obtained by fitting the polynomial

$$\beta_0 + \beta_1(\cdot - x) + \cdots + \beta_p(\cdot - x)^p$$

to the $(X_i, Y_i)$ using weighted least squares with kernel weights $\frac{2}{(n+1)h} K\left(\frac{x - X_i}{ih}\right)$ for the **GCA** estimator. The value of $\hat{m}(x; p, h)$ is the height of the fit $\hat{\beta}_0$, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \cdots, \hat{\beta}_p)$ minimizes (4.2) for the **GCA** estimator.

$$\sum_{i=1}^{n} \{Y_i - \beta_0 - \cdots - \beta_p (X_i - x)^p\}^2 \frac{2}{(n+1)h} K\left(\frac{x - X_i}{ih}\right). \tag{4.2}$$

Assuming the convertibility of $(\mathbf{X_x^T W_{GCA.x} X_x})$, standard weighted least squares theory leads to the solution

$$\hat{\boldsymbol{\beta}}_{GCA} =(\mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{Y}$$

where $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$ is the vector of responses,

$$\mathbf{X}_x = \begin{vmatrix} 1 & X_1 - x & \cdots & (X_1 - x)^p \\ 1 & X_2 - x & \cdots & (X_2 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_n - x & \cdots & (X_n - x)^p \end{vmatrix}$$

is an $n \times (p + 1)$ design matrix, and

$$\mathbf{W}_{GCA.x} = \frac{2}{(n+1)h} \begin{vmatrix} K\left(\frac{X_1 - x}{h}\right) & 0 & \cdots & 0 \\ 0 & K\left(\frac{X_2 - x}{2h}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & K\left(\frac{X_n - x}{nh}\right) \end{vmatrix}$$

is an $n \times n$ diagonal matrix of weights.

Since the estimator of $m(x)$ is the intercept coefficient, we obtain

$$\hat{m}_{GCA}(x; p, h) =\mathbf{e_1^T} \hat{\boldsymbol{\beta}}_{GCA} = \mathbf{e_1^T} (\mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{Y}, \tag{4.3}$$

where $\mathbf{e}_1$ is the $(p + 1) \times 1$ vector, with 1 in the first entry and zero elsewhere.

For special case $p = 0$, the regression based on the **GCA** estimator can be expressed as follows :

$$\hat{m}_{GCA}(x; 0, h) =\frac{\sum_{i=1}^n K\left(\frac{X_i - x}{ih}\right) y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{ih}\right)}, \tag{4.4}$$

and the local linear estimators $p = 1$:

$$\hat{m}_{GCA}(x; 1, h) = n^{-1} \sum_{i=1}^{n} \frac{\{\hat{G}_2(x; h) - \hat{G}_1(x; h)(X_i - x)\} K\left(\frac{X_i - x}{ih}\right) Y_i}{\hat{G}_2(x; h)\hat{G}_0(x; h) - \hat{G}_1(x; h)^2}, \tag{4.5}$$

where $\hat{G}_j(x; h) = \frac{2}{n(n+1)h} \sum_{i=1}^{n} (X_i - x)^j K\left(\frac{X_i - x}{ih}\right)$

**Asymptotic MSE Approximations**

We present the mean and variance calculations for $\hat{m}(x; p, h)$ in the $p = 1$. Suppose that the design is an independent sample, denoted by $X_1, \cdots, X_n$, having density $f$. We make the following assumptions in our analysis:

1. The functions $m''$ are each continuous functions.

2. The kernel $K$ is satisfying,

$$\int K(\mu) d\mu = 1 \quad \textit{and} \quad \int \mu K(\mu) d\mu = 0 \quad \textit{and} \quad \int \mu^2 K(\mu) d\mu < \infty.$$

3. The bandwidth is satisfying $h \to 0$ and $n^2 h \to \infty$.

In this section we analyze the mean and variance of the estimator $\hat{m}(x; p, h)$. It follows directly from (4.3), then

$$\mathrm{E}(\hat{m}_{GCA}(x; 1, h)) = \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{M},$$

where $\mathbf{M} = (m(x_1), \cdots, m(x_n))^T$. And local linear fitting ,design matrix $\mathbf{X}_x$

$$\mathbf{X}_x = \begin{vmatrix} 1 & X_1 - x \\ \vdots & \vdots \\ 1 & X_n - x \end{vmatrix}$$

is an $n \times 2$ matrix. According to a version of Taylor's expansion , for any $x \epsilon [0, 1]$,

$$m(x_i) = m(x) + (x_i - x)m'(x) + \frac{1}{2}(x_i - x)^2 m''(x) + \cdots$$

which implies that

$$\mathbf{M} = \mathbf{X}_x \begin{vmatrix} m(x) \\ m'(x) \end{vmatrix} + \frac{1}{2}m''(x) \begin{vmatrix} (x_1 - x)^2 \\ \vdots \\ (x_n - x)^2 \end{vmatrix} + \cdots$$

The first term in the expression of $\mathrm{E}(\hat{m}_{GCA}(x; 1, h))$ is

$$\mathbf{e}_1^T(\mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x)^{-1}(\mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x) \begin{vmatrix} m(x) \\ m'(x) \end{vmatrix} = \mathbf{e}_1^T \begin{vmatrix} m(x) \\ m'(x) \end{vmatrix} = m(x),$$

Then the bias of $\hat{m}_{GCA}(x; 1, h)$ is

$$\begin{aligned} &\mathrm{E}(\hat{m}_{GCA})(x; 1, h) - m(x) \\ =&\frac{1}{2}m''(x)\mathbf{e}_1^T(\mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{GCA.x} \begin{vmatrix} (X_1 - x)^2 \\ \vdots \\ (X_n - x)^2 \end{vmatrix} + \cdots \end{aligned}$$

If $m^r(x) = 0$ for all $r \geqslant 2$, then $\hat{m}_{GCA}(x; 1, h)$ is an exactly unbiased estimator for linear $m$.

To compute the leading bias term for general $m$

$$\mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x = \begin{vmatrix} \hat{G}_0(x; h) & \hat{G}_1(x; h) \\ \hat{G}_1(x; h) & \hat{G}_2(x; h) \end{vmatrix}$$

and

$$\mathbf{X}_x^T \mathbf{W}_{GCA.x} \begin{vmatrix} (X_1 - x)^2 \\ \vdots \\ (X_n - x)^2 \end{vmatrix} = \begin{vmatrix} \hat{G}_2(x;h) \\ \hat{G}_3(x;h) \end{vmatrix}$$

$$\begin{aligned}
\mathrm{E}\hat{G}_j(x;h) =& \frac{1}{n} \mathrm{E} \sum_{i=1}^{n} (X_i - x)^j K\left(\frac{X_i - x}{ih}\right) \\
=& \frac{1}{n} \sum_{i=1}^{n} \int (y - x)^j K\left(\frac{y - x}{ih}\right) f(y) dy \\
=& \frac{1}{n} \sum_{i=1}^{n} \int \mu^j (ih)^{j+1} K(\mu) f(x + \mu ih) d\mu \\
=& \frac{1}{n} \sum_{i=1}^{n} (ih)^{j+1} \int \mu^j K(\mu) \{ f(x) + \mu ih f'(x) + \frac{1}{2} (\mu ih)^2 f''(x) \} d\mu.
\end{aligned}$$

So,

$$\mathrm{E}\hat{G}_j(x;h) = \begin{cases} \frac{2}{n(n+1)} \sum_{i=1}^{n} (ih)^{j+1} f(x) \int \mu^j K(\mu) d\mu + O_p(\frac{2}{n(n+1)h} \sum_{i=1}^{n} (ih)^{j+3}) \\ \quad \text{if j is even ;} \\ \frac{2}{n(n+1)} \sum_{i=1}^{n} (ih)^{j+2} f'(x) \int \mu^{j+1} K(\mu) d\mu + O_p(\frac{2}{n(n+1)h} \sum_{i=1}^{n} (ih)^{j+4}) \\ \quad \text{if j is odd .} \end{cases}$$

which leads to

$$\frac{1}{n} \mathbf{X}_x^T \mathbf{W}_{GCA.x} \mathbf{X}_x = \begin{vmatrix} f(x) + O_p(n^2 h^2) & \frac{n^2 h^2}{2} f'(x) \mu_2(K) + O_p(n^3 h^3) \\ \frac{n^2 h^2}{2} f'(x) \mu_2(K) + O_p(n^3 h^3) & \frac{n^2 h^2}{2} f(x) \mu_2(K) + O_p(n^3 h^3) \end{vmatrix},$$

and

$$\frac{1}{n} \mathbf{X}_x^T \mathbf{W}_{GCA.x} \begin{vmatrix} (X_1 - x)^2 \\ \vdots \\ (X_n - x)^2 \end{vmatrix} = \begin{vmatrix} \frac{n^2 h^2}{2} f(x) \mu_2(K) + O_p(n^3 h^3) \\ \frac{n^4 h^4}{3} f'(x) \int \mu^4 K(\mu) d\mu + O_p(n^4 h^4) \end{vmatrix},$$

$$(\frac{1}{n}\mathbf{X}_x^T\mathbf{W}_{GCA.x}\mathbf{X}_x)^{-1} = \begin{vmatrix} f(x)^{-1} & -f'(x)/f^2(x) \\ -f'(x)/f^2(x) & \{\frac{n^2h^2}{2}f(x)\int \mu^2 K(\mu)d\mu\}^{-1} \end{vmatrix}.$$

It follows that the conditional bias is given by

$$\mathrm{E}\{\hat{m}_{GCA}(x;1,h) - m(x)|X_1,\cdots,X_n\} = \frac{n^2h^2}{4}m''(x)\mu_2(K) + o_p(n^2h^2).$$

For the variance approximation, note that

$$\begin{aligned}&\mathrm{Var}\{\hat{m}_{GCA}(x;1,h)\}\\ &= \mathbf{e_1^T}(\mathbf{X}_x^T\mathbf{W}_{GCA.x}\mathbf{X}_x)^{-1}\mathbf{X}_x^T\mathbf{W}_{GCA.x}\mathbf{V}\mathbf{W}_{GCA.x}\mathbf{X}_x(\mathbf{X}_x^T\mathbf{W}_{GCA.x}\mathbf{X}_x)^{-1}\mathbf{e}_1\end{aligned}$$

Using approximations analogous to those used above,

$$\mathbf{X}_x^T\mathbf{W}_{GCA.x}\mathbf{V}\mathbf{W}_{GCA.x} = \begin{vmatrix} \frac{2}{(n+1)h}R(K)\sigma(x) & O(n^{-1}) \\ O(n^{-1}) & \frac{4}{3}h\mu_2(K)\sigma(x) \end{vmatrix},$$

$$\mathrm{Var}\{\hat{m}_{GCA}(x;1,h)\} = \frac{2R(K)\sigma(x)}{n^2h} + o((n^2h)^{-1}).$$

This leads to MSE

$$\begin{aligned}\mathrm{MSE}\{\hat{m}_{GCA}(x;1,h)\} =& \mathrm{Bias}^2\{\hat{m}_{GCA}(x;1,h)\} + \mathrm{Var}\{\hat{m}_{GCA}(x;1,h)\}\\ =&\frac{n^4h^4}{16}\{m''(x)\mu_2(K)\}^2 + \frac{2R(K)\sigma(x)}{n^2h} + o((n^2h)^{-1}).\end{aligned}$$

### 4.2.2 Local polynomial based on LCA Estimator

In this section, we investigate local polynomial **LCA** estimators that are defined by (3.1). Taking a similar approach on the local polynomial based on **GCA** estimator, at a point $x$ the estimator $\hat{m}_{LCA}(x; p, h)$ is obtained by fitting the polynomial

$$\beta_0 + \beta_1(\cdot - x) + \cdots + \beta_p(\cdot - x)^p$$

to the $(X_i, Y_i)$ using weighted least squares with kernel weights $\frac{1}{ih}K\left(\frac{x-X_i}{ih}\right)$ for **LCA**. The value of $\hat{m}_{LGA}(x; p, h)$ is the height of the fit $\hat{\beta}_0$, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \cdots, \hat{\beta}_p)$ minimizes (4.6) for **LCA**.

$$\sum_{i=1}^{n}\{Y_i - \beta_0 - \cdots - \beta_p(X_i - x)^p\}^2 \frac{1}{ih}K\left(\frac{x - X_i}{ih}\right). \tag{4.6}$$

Assuming the convertibility of $(\mathbf{X_x^T W_{LCA.x} X_x})$, standard weighted least squares theory leads to the solution

$$\hat{\boldsymbol{\beta}}_{LCA} = (\mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{Y},$$

where $\mathbf{Y}$ is the same as in the previous section. But a diagonal weight matrix changes to

$$\mathbf{W}_{LCA.x} = \frac{1}{h}\begin{vmatrix} K\left(\frac{X_i - x}{h}\right) & 0 & \cdots & 0 \\ 0 & \frac{1}{2}K\left(\frac{X_i - x}{2h}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n}K\left(\frac{X_i - x}{nh}\right) \end{vmatrix}.$$

Then we can obtain

$$\hat{m}_{LCA}(x; p, h) = \mathbf{e_1^T} \hat{\boldsymbol{\beta}}_{LCA} = \mathbf{e_1^T}(\mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{Y}, \tag{4.7}$$

where $\mathbf{e}_1$ is the $(p + 1) \times 1$ vector, with 1 in the first entry and zero elsewhere.

For special case $p = 0$, the kernel regression **LCA** can be expressed as the following:

$$\hat{m}_{LCA}(x; 0, h) = \frac{\sum_{i=1}^{n} \frac{1}{i} K\left(\frac{X_i - x}{ih}\right) y_i}{\sum_{i=1}^{n} \frac{1}{i} K\left(\frac{X_i - x}{ih}\right)}, \tag{4.8}$$

and local linear estimators $p = 1$:

$$\hat{m}_{LCA}(x; 1, h) = n^{-1} \sum_{i=1}^{n} \frac{\{\hat{L}_2(x; h) - \hat{L}_1(x; h)(X_i - x)\} K\left(\frac{X_i - x}{ih}\right) Y_i}{\hat{L}_2(x; h)\hat{S}_0(x; h) - \hat{L}_1(x; h)^2}, \tag{4.9}$$

where $\hat{L}_j(x; h) = \frac{2}{n(n+1)h} \sum_{i=1}^{n} (X_i - x)^j \frac{1}{ih} K\left(\frac{X_i - x}{ih}\right)$

**Asymptotic** MSE **Approximations**

We present the mean and variance calculations for $\hat{m}(x; p, h)$ in the $p = 1$. Suppose that the design is an independent sample, denoted by $X_1, \cdots, X_n$, having density $f$. We make the same assumptions from the previous section, and analyze the mean and variance of the estimator $\hat{m}_{LCA}(x; p, h)$. It follows directly from (4.7) that

$$\mathrm{E}(\hat{m}_{LCA}(x; 1, h)) = \mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{X}_x)^{-1} \mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{M}.$$

Using Taylor expansion for $\mathbf{M}$ we obtain

$$\mathbf{e}_1^T (\mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{X}_x)^{-1} (\mathbf{X}_x^T \mathbf{W}_{LCA.x} \mathbf{X}_x) \begin{vmatrix} m(x) \\ m'(x) \end{vmatrix} = \mathbf{e}_1^T \begin{vmatrix} m(x) \\ m'(x) \end{vmatrix} = m(x).$$

The bias of $\hat{m}_{LCA}(x;1,h)$ is

$$\begin{aligned}
&\mathrm{E}\hat{m}_{LCA}(x;1,h) - m(x) \\
&= \frac{1}{2}m''(x)\mathbf{e}_1^T(\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{X}_x)^{-1}\mathbf{X}_x^T\mathbf{W}_{LCA.x}\begin{vmatrix}(X_1-x)^2\\ \vdots \\ (X_n-x)^2\end{vmatrix} + \cdots
\end{aligned}$$

If $m^r(x) = 0$ for all $r \geqslant 2$, then $\hat{m}_{LCA}(x;1,h)$ is an exactly unbiased estimator for linear $m$.

To compute the leading bias term for general $m$

$$\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{X}_x = \begin{vmatrix}\hat{L}_0(x;h) & \hat{L}_1(x;h)\\ \hat{L}_1(x;h) & \hat{L}_2(x;h)\end{vmatrix},$$

and

$$\mathbf{X}_x^T\mathbf{W}_{LCA.x}\begin{vmatrix}(X_1-x)^2\\ \vdots \\ (X_n-x)^2\end{vmatrix} = \begin{vmatrix}\hat{L}_2(x;h)\\ \hat{L}_3(x;h)\end{vmatrix}.$$

$$\begin{aligned}
\mathrm{E}\hat{L}_j(x;h) =& \frac{1}{n}\mathrm{E}\sum_{i=1}^{n}(X_i-x)^j\frac{1}{i}K\left(\frac{X_i-x}{ih}\right)\\
=& \frac{1}{nh}\sum_{i=1}^{n}\int(y-x)^j\frac{1}{i}K\left(\frac{y-x}{ih}\right)f(y)dy\\
=& \frac{1}{n}\sum_{i=1}^{n}\int\mu^j(ih)^jK(\mu)f(x+\mu ih)d\mu\\
=& \frac{1}{n}\sum_{i=1}^{n}(ih)^j\int\mu^jK(\mu)\{f(x)+\mu ihf'(x)+\frac{1}{2}(\mu ih)^2f''(x)\}d\mu.
\end{aligned}$$

So,

$$\mathrm{E}\hat{L}_j(x;h) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(ih)^{j+1}f'(x)\int \mu^{j+1}K(\mu)d\mu + O_p(\frac{1}{n}\sum_{i=1}^{n}(ih)^{j+2}) \\ \quad \text{if j is odd ;} \\ \frac{1}{n}\sum_{i=1}^{n}(ih)^{j}f(x)\int \mu^{j}K(\mu)d\mu + O_p(\frac{1}{n}\sum_{i=1}^{n}(ih)^{j+1}) \\ \quad \text{if j is even .} \end{cases}$$

which leads to

$$\frac{1}{n}\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{X}_x = \begin{vmatrix} f(x) + o_p(nh) & \frac{n^2h^2}{3}f'(x)\mu_2(K) + o_p(n^2h^2) \\ \frac{n^2h^2}{3}f'(x)\mu_2(K) + o_p(n^3h^3) & \frac{n^2h^2}{3}f(x)\mu_2(K) + O_p(n^3h^3) \end{vmatrix},$$

and

$$\frac{1}{n}\mathbf{X}_x^T\mathbf{W}_{LCA.x}\begin{vmatrix} (X_1 - x)^2 \\ \vdots \\ (X_n - x)^2 \end{vmatrix} = \begin{vmatrix} \frac{n^2h^2}{3}f(x)\mu_2(K) + O_p(n^3h^3) \\ \frac{n^4h^4}{5}f'(x)\int \mu^4K(\mu)d\mu + O_p(n^4h^4) \end{vmatrix},$$

$$(\frac{1}{n}\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{X}_x)^{-1} = \begin{vmatrix} f(x)^{-1} & -f'(x)/f^2(x) \\ -f'(x)/f^2(x) & \{\frac{n^2h^2}{3}f(x)\mu_2(K)\}^{-1} \end{vmatrix}.$$

It follows that the conditional bias is given by

$$\mathrm{E}\{\hat{m}_{LCA}(x;1,h) - m(x)|X_1,\cdots,X_n\} = \frac{n^2h^2}{6}m''(x)\mu_2(K) + o_p(n^2h^2).$$

For the variance approximation, note that

$$\begin{aligned} &\mathrm{Var}\{\hat{m}_{LCA}(x;1,h) \\ &= \mathbf{e_1^T}(\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{X}_x)^{-1}\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{V}\mathbf{W}_{LCA.x}\mathbf{X}_x(\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{X}_x)^{-1}\mathbf{e}_1. \end{aligned}$$

Using approximations analogous to those used above,

$$\mathbf{X}_x^T\mathbf{W}_{LCA.x}\mathbf{V}\mathbf{W}_{LCA.x} = \begin{vmatrix} \frac{\sum_{i=1}^n \frac{1}{i}}{nh}R(K)\sigma(x) & O(n^{-1}) \\ O(n^{-1}) & \frac{1}{2}h\mu_2(K)\sigma(x) \end{vmatrix},$$

$$\mathrm{Var}\{\hat{m}_{LCA}(x;1,h)\} = \frac{\sum_{i=1}^n \frac{1}{i}R(K)\sigma(x)}{n^2h} + o((n^2h)^{-1}).$$

This leads to MSE

$$\begin{aligned}\mathrm{MSE}\{\hat{m}_{GCA}(x;1,h)\} =&\mathrm{Bias}\{\hat{m}_{GCA}(x;1,h)\} + \mathrm{Var}\{\hat{m}_{GCA}(x;1,h)\} \\ =&\frac{n^4h^4}{36}\{m''(x)\mu_2(K)\}^2 + \frac{\sum_{i=1}^n \frac{1}{i}R(K)\sigma(x)}{n^2h} + o((n^2h)^{-1}).\end{aligned}$$

### 4.2.3 Simulation Study

The true regression function is

$$m(x) = \sin^3(2\pi x^3), \tag{4.10}$$

confined to the interval $[0,1]$, and is represented by the black solid curve. The data $X_1, \cdots, X_n$ are generated by uniform [0,1] and the data $Y_1, \cdots, Y_n$ are generated by

$$Y_i = m(X_i) + 0.1\epsilon_i, \quad i = 1, \cdots, n$$

where $\epsilon_i$ are independent $N(0,1)$ random variables. The $(X_i, Y_i)$ pairs are represented by the circles. The black solid line is the regression function $m(x)$ given by (4.10), the blue solid curve is the usual kernel estimate with $p = 1$, the blue dot curve is the usual kernel estimate with $p = 0$, the red solid curve is **GCA** estimate with $p = 1$, the red dot curve is **GCA** estimate with $p = 0$, the green solid curve is the **LCA** estimate with $p = 1$, and the green dot curve is **GCA** estimate with $p = 0$. Figures 13 and

14 show that there is not too much difference between $p = 0$ and $p = 1$ for different estimators. The **GCA** and **LCA** are better than **KDE**, especially at catching the hump (close to 0.6) and the valley (close to 0.9). The **GCA** and **LCA** are almost the same curve.

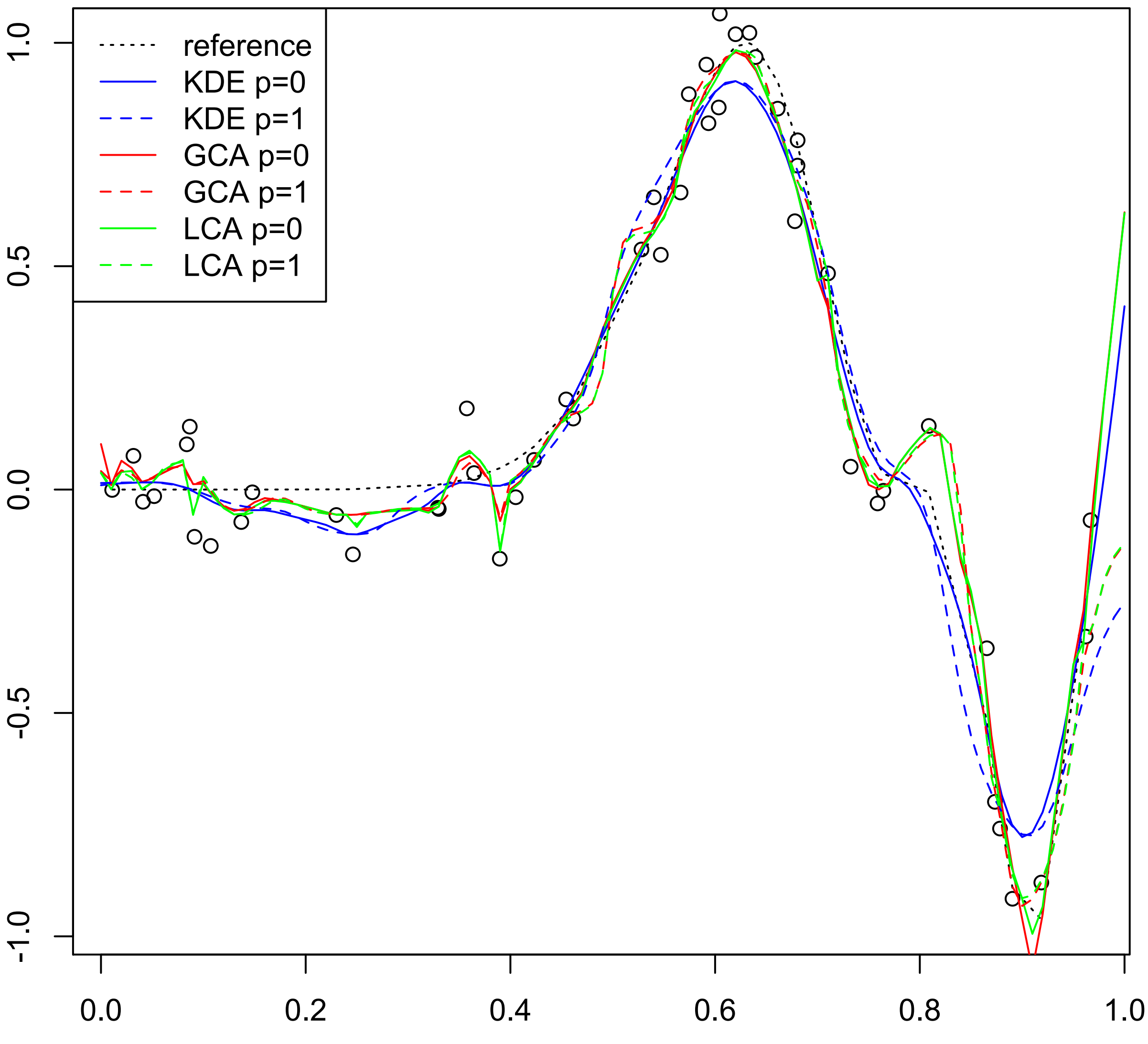



**Figure 13:** Estimated regression function with sample size n=50

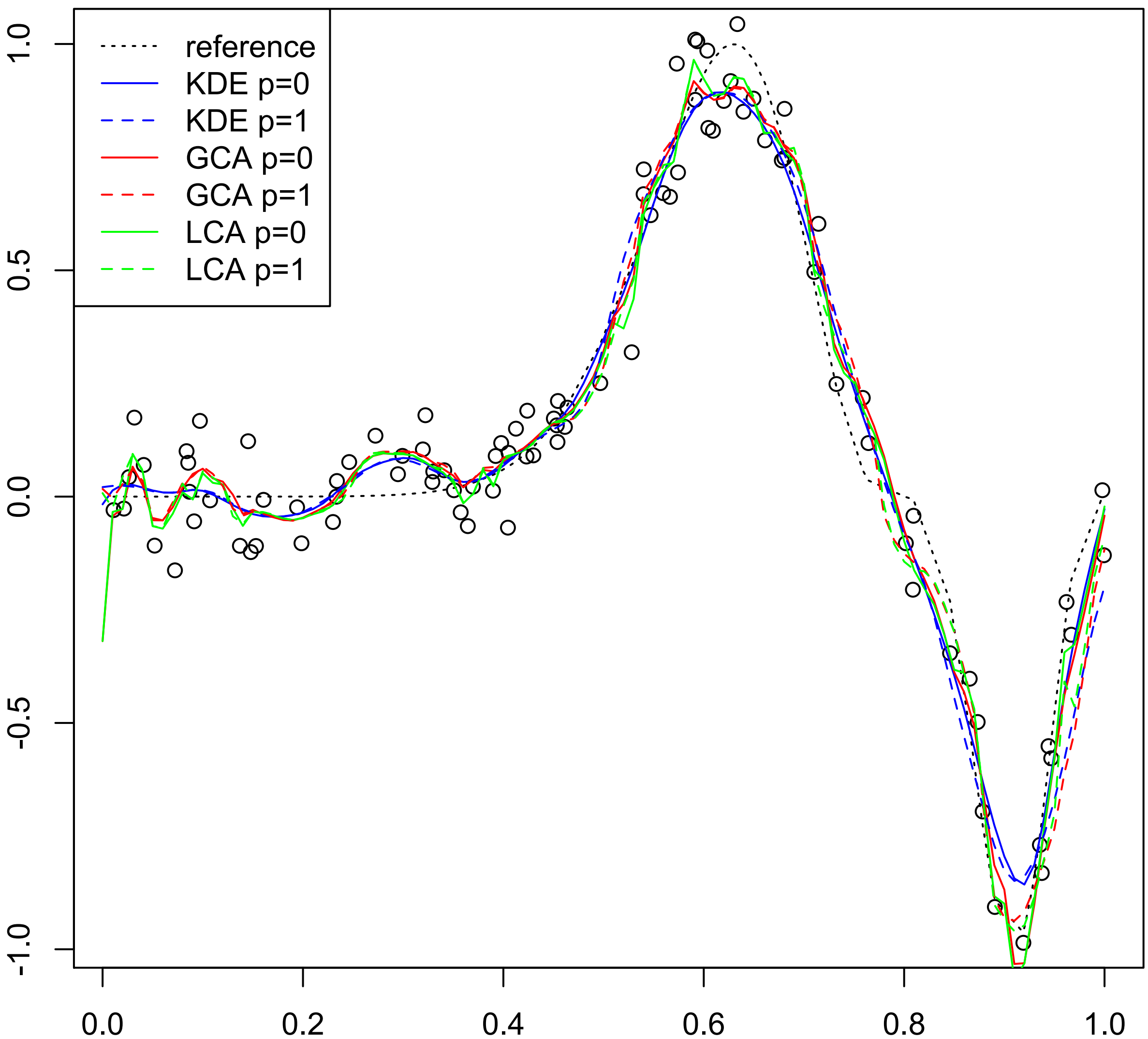



**Figure 14:** Estimated regression function with sample size n=100

### 4.3 Cumulative Distribution Function Estimation

Let $X_1, \cdots, X_n$ be independent and identically distributed random vectors in $\mathbb{R}$, with absolutely continuous distribution function $F$ and corresponding probability density function $f$. The traditional estimator of the **CDF** is the empirical distribution function, which is given by

$$\hat{F}_n(t) = \frac{1}{n}\sum_{j=1}^{n} I(X_i \leqslant t)$$

where $I(\cdot)$ represents the indicator function. The smoothed estimation of the **CDF**, $\hat{F}_n(x)_{GCA}$ and $\hat{F}_n(x)_{GCA}$ are constructed by integrating $\hat{f}_{GCA}$ and $\hat{f}_{LCA}$, which are defined as (3.2) and (3.1). So, smooth estimators $\hat{F}_n(x)_{GCA}$ and $\hat{F}_n(x)_{GCA}$ can be expressed as:

$$\begin{aligned}
\hat{F}_n(x)_{GCA} =& \frac{2}{n(n+1)h}\int_{-\infty}^{t}\sum_{i=1}^{n} K\left(\frac{x-X_i}{ih}\right)dx = \frac{2}{n(n+1)}\sum_{i=1}^{n} i\bar{\mathbb{K}}\left(\frac{x-X_i}{ih}\right),\\
\hat{F}_n(x)_{LCA} =& \frac{1}{nh}\int_{-\infty}^{t}\sum_{i=1}^{n}\frac{1}{i}K\left(\frac{X_i-x}{ih}\right) = \frac{1}{n}\sum_{i}^{n}\bar{\mathbb{K}}\left(\frac{x-X_i}{ih}\right),
\end{aligned}$$

and the estimator based on the standard kernel density estimation is

$$\hat{F}_n(x)_{KDE} = \frac{1}{nh}\int_{-\infty}^{t}\sum_{i=1}^{n} K\left(\frac{X_i-x}{h}\right) = \frac{1}{n}\sum_{i}^{n}\bar{\mathbb{K}}\left(\frac{x-X_i}{h}\right),$$

where $\bar{\mathbb{K}}(t) = \int_{-\infty}^{t} K(\mu)d\mu$. In the next section, we calculate the mean and variance of smooth estimators $\hat{F}_n(x)_{GCA}$ and $\hat{F}_n(x)_{GCA}$.

### 4.3.1 Properties

**Theorem 4.3.1** *Assume that $F(x)$ is twice continuously differentiable, $K$ is bounded, symmetric and compactly supported. Then as $n \to \infty$*

$$\mathrm{E}(\hat{F}_n(t)_{GCA}) = F(t) + \frac{n^2h^2}{2}\mu_2(K)f'(t) + o(n^2h^2), \tag{4.11}$$

$$\mathrm{E}(\hat{F}_n(t)_{LCA}) = F(t) + \frac{n^2h^2}{6}\mu_2(K)f'(t) + o(n^2h^2), \tag{4.12}$$

$$\mathrm{Var}(\hat{F}_n(t)_{GCA}) = \frac{4}{3n}(F(t) - F^2(t)) - 2h\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(nh), \tag{4.13}$$

$$\mathrm{Var}(\hat{F}_n(t)_{LCA}) = \frac{1}{n}(F(t) - F^2(t)) - h\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(nh), \tag{4.14}$$

*where* $f'(t) = F^{(2)}(t)$.

For proof see Appendix Section 4.4.

By following Theorem 4.3.1, we can immediately obtain the following results for the MSE of $\hat{F}_n(t)_{GCA}$ and $\hat{F}_n(t)_{LCA}$:

$$\begin{aligned}
\text{MSE}\hat{F}_n(t)_{GCA} =& \frac{n^4h^4}{4}\mu_2^2(K)(f'(t))^2 + \frac{4}{3n}(F(t) - F^2(t)) \\
&- 2h\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(n^4h^4), \qquad (4.15) \\
\text{MSE}\hat{F}_n(t)_{LCA} =& \frac{n^4h^4}{16}\mu_2^2(K)(f'(t))^2 + \frac{1}{n}(F(t) - F^2(t)) \\
&- h\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(n^4h^4). \qquad (4.16)
\end{aligned}$$

### 4.3.2 Simulation Study

We generate the data from standard normal distribution, mixture normal $0.5N(-1,1) + 0.5N(1,1)$, and mixture normal $0.75N(0,1) + 0.25N(1.5,4/9)$ from difference sample sizes n=15, 25, 50, 100 and 500. From the figure 16, the green dashed line is best, followed by the red dotted line, with the last being the blue dotdash line. So the performance of $\hat{F}_n(t)_{LCA}$ is slightly better than $\hat{F}_n(t)_{GCA}$, and these two estimators are significantly better than standard $\hat{F}_n(t)_{KDE}$, especially at the extreme value. Even for the sample size n=500, the new estimators are better than a standard kernel estimator. Mixture normal $0.5N(-1,1) + 0.5N(1,1)$ and $0.75N(0,1) + 0.25N(1.5,4/9)$ have the same results.

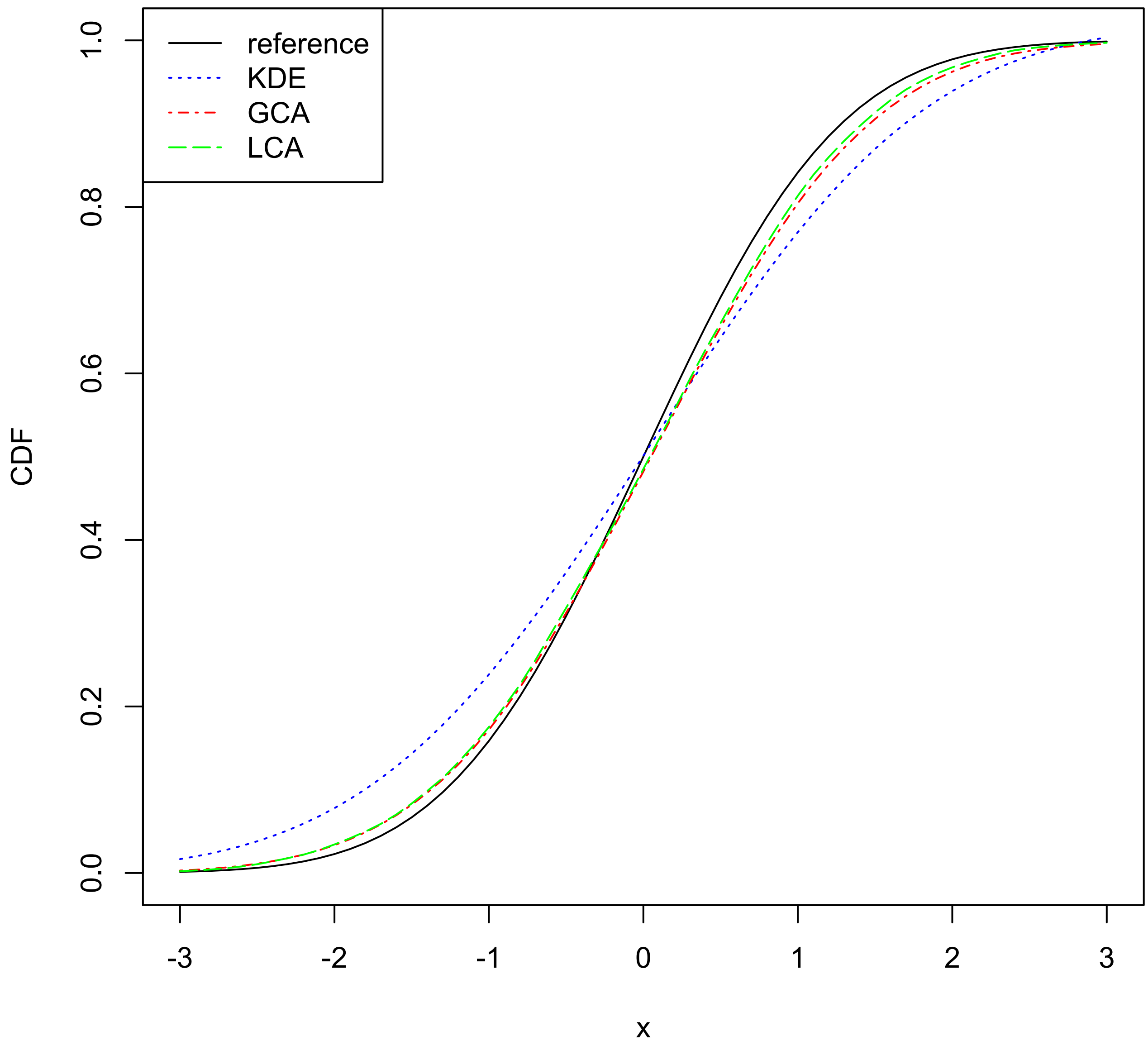
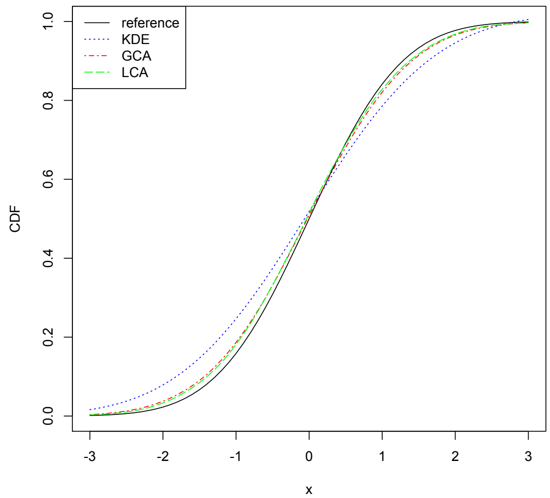
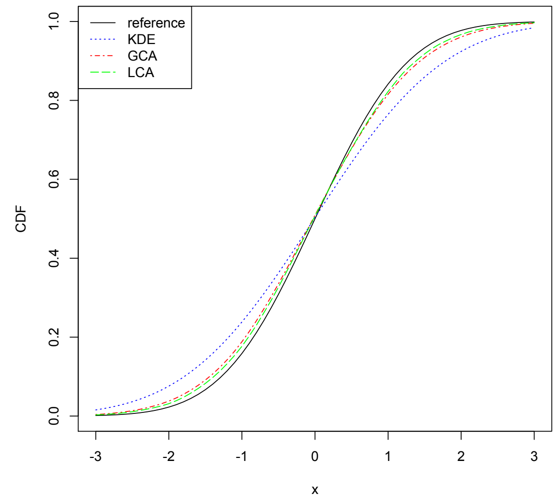


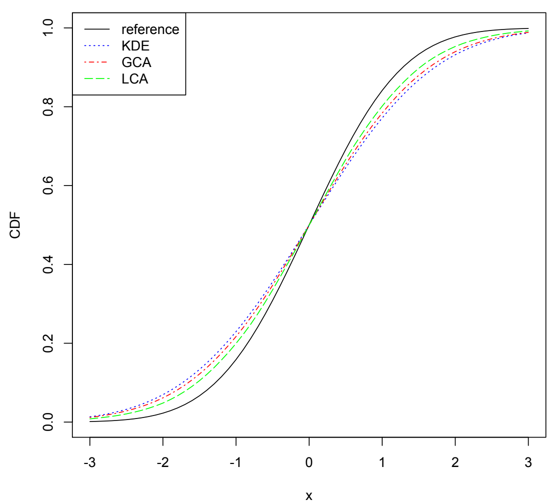

**Figure 15:** CDF estimation from normal distribution for sample size n=15



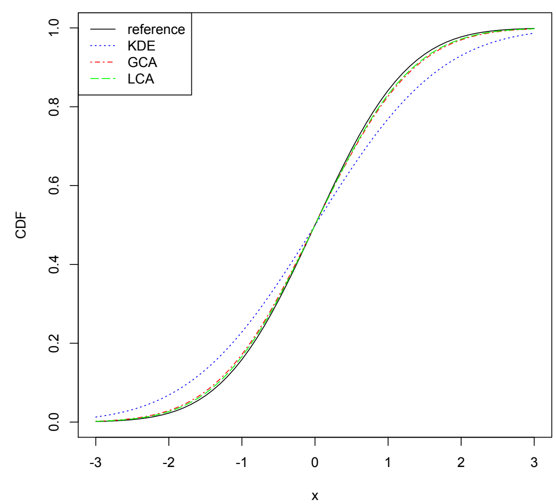

**(a)** n=25




**(b)** n=50




**(c)** n=100




**(d)** n=500

**Figure 16:** CDF estimation from normal distribution for different sample sizes

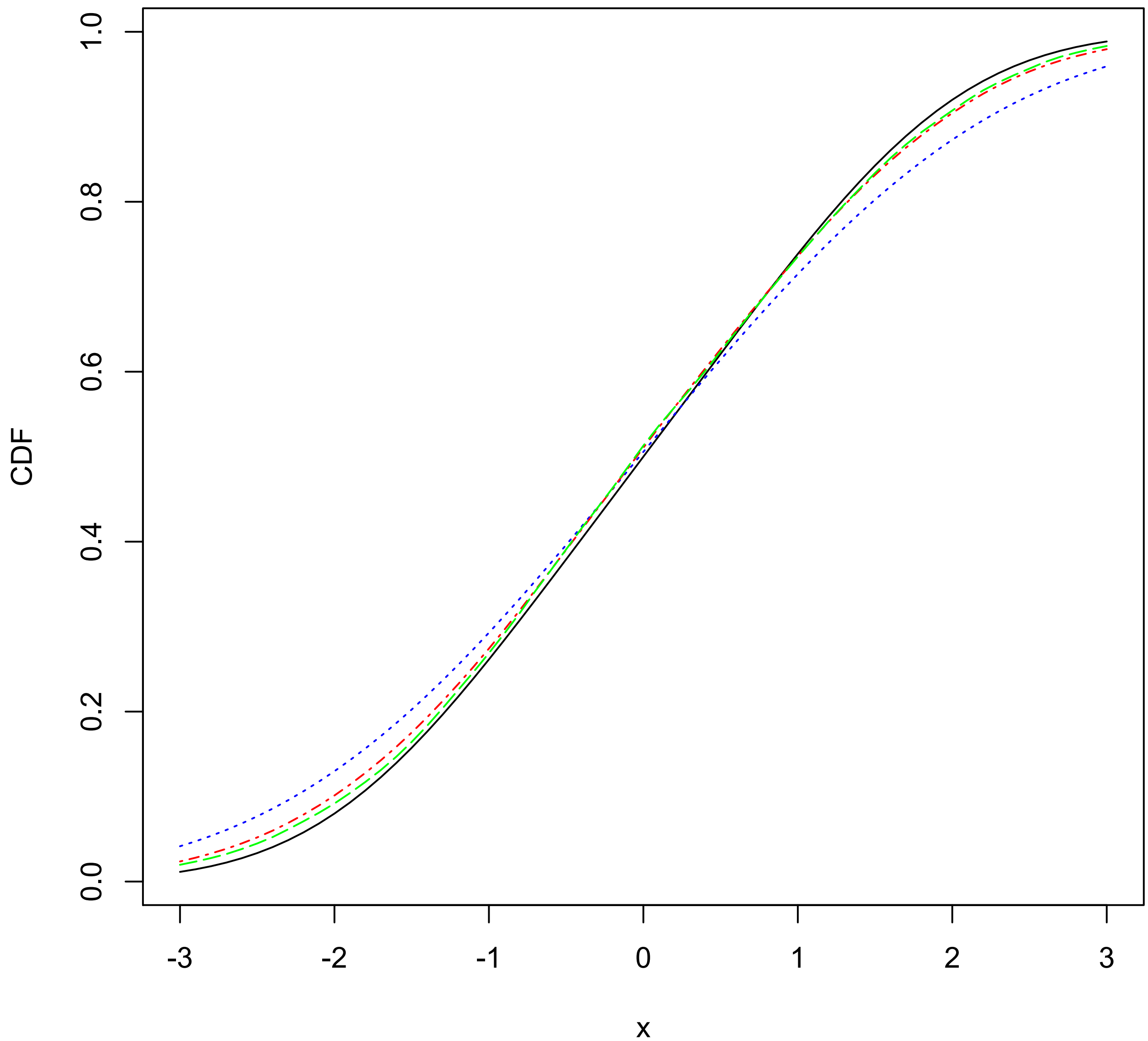
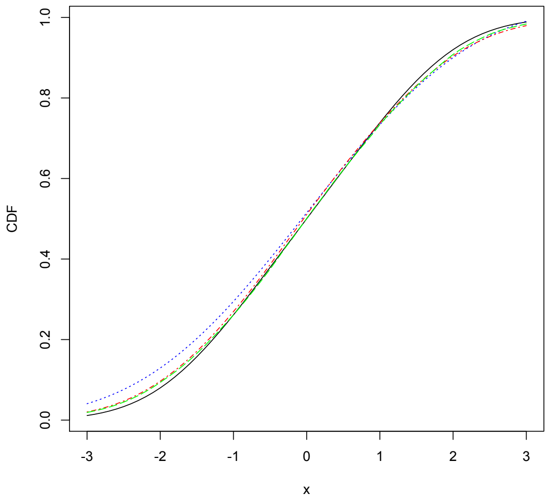
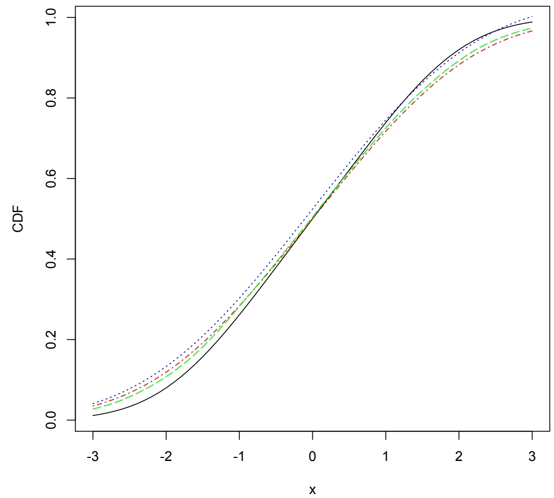


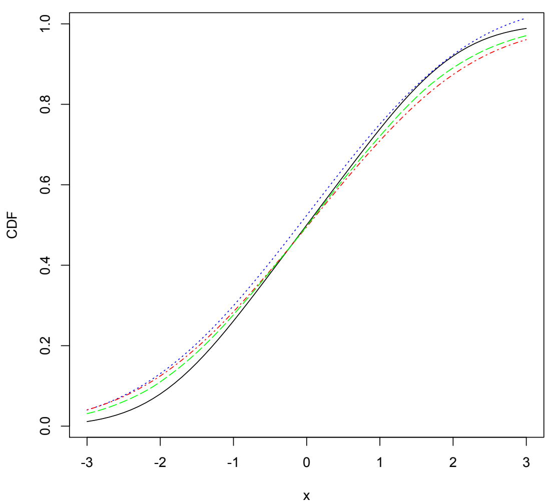

**Figure 17:** CDF estimation from $0.5N(-1,1) + 0.5N(1,1)$ for sample size n=15



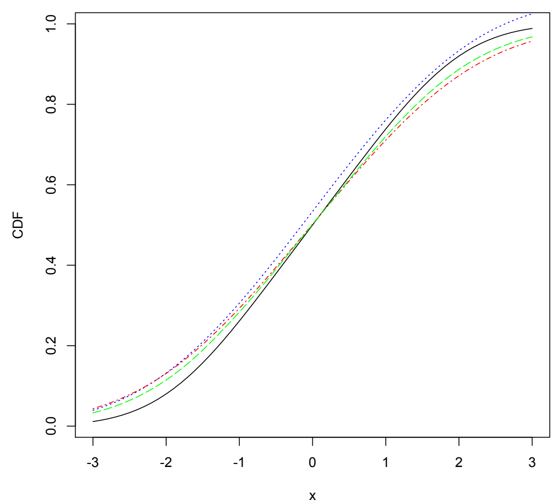

**(a)** n=25




**(b)** n=50




**(c)** n=100




**(d)** n=500

**Figure 18:** CDF estimation from $0.5N(-1, 1) + 0.5N(1, 1)$ for different sample size

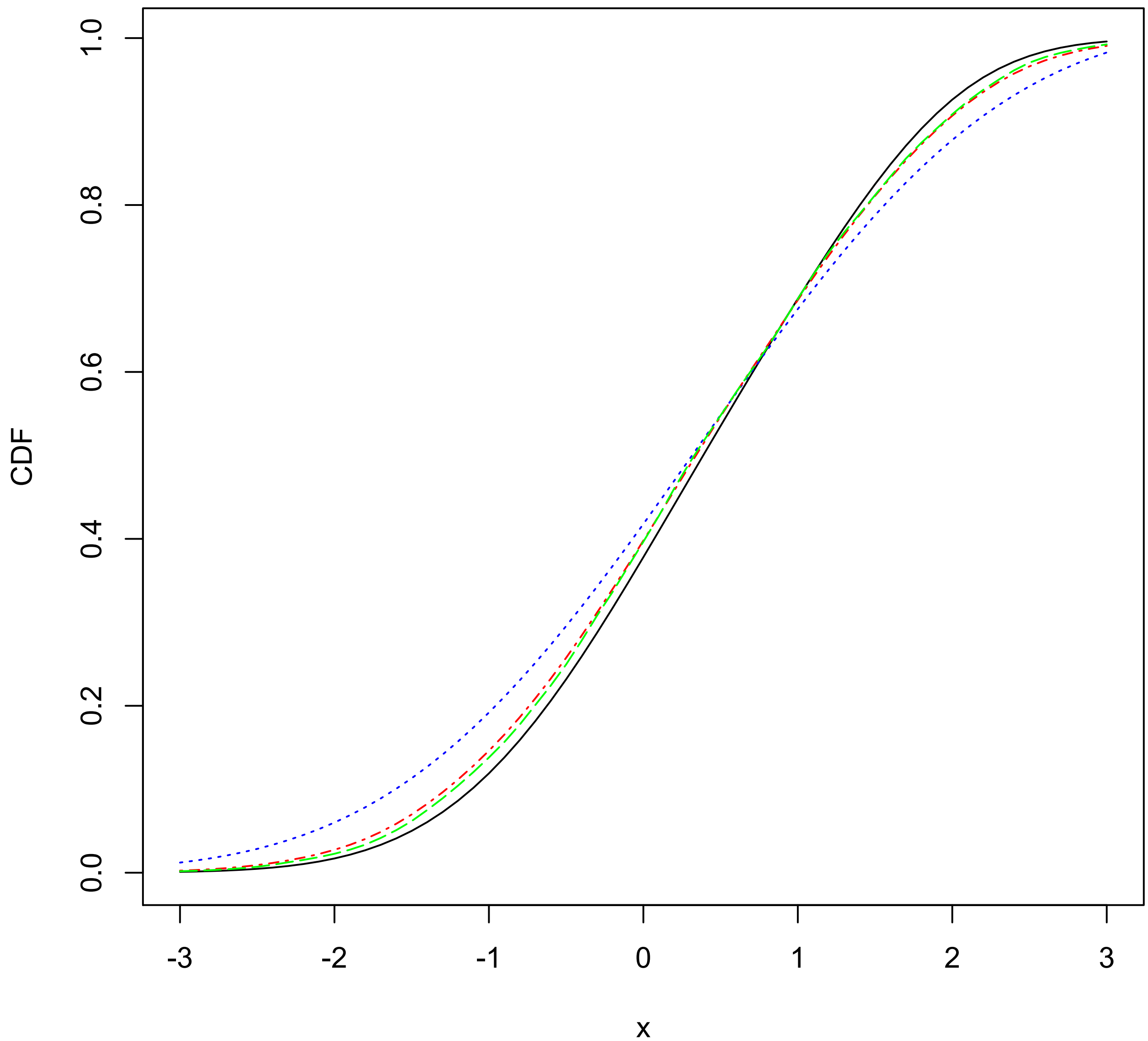
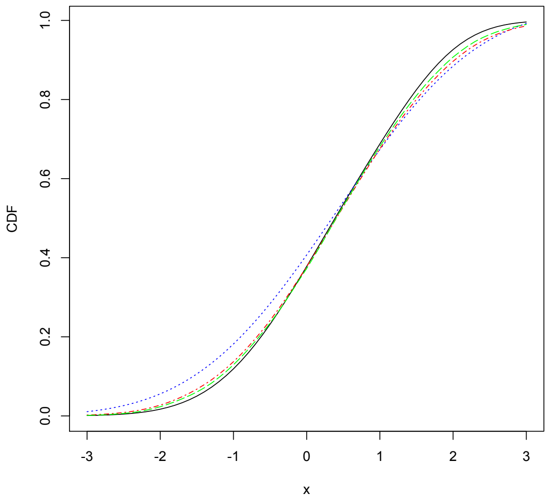
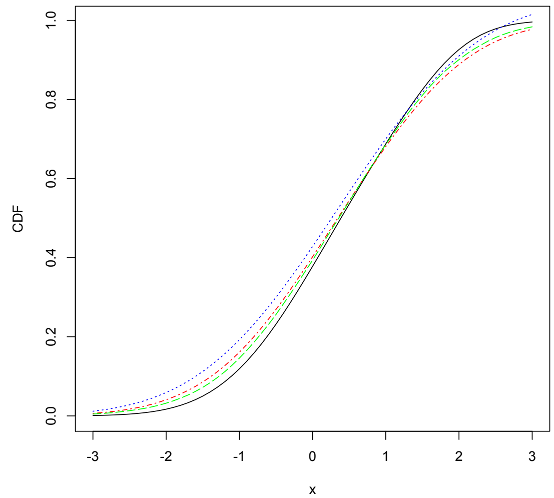


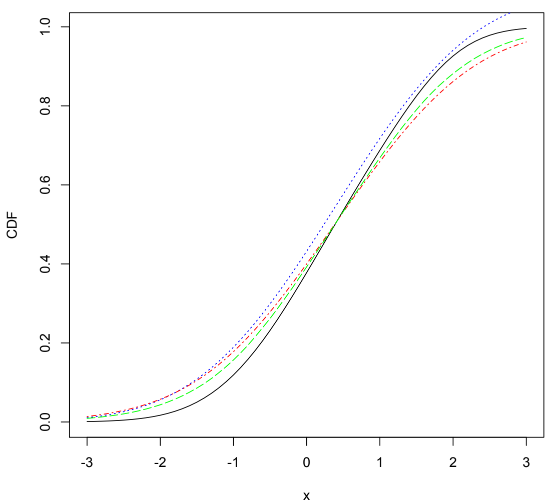

**Figure 19:** CDF estimation from $0.75N(0,1) + 0.25N(1.5, 4/9)$ for sample size n=15



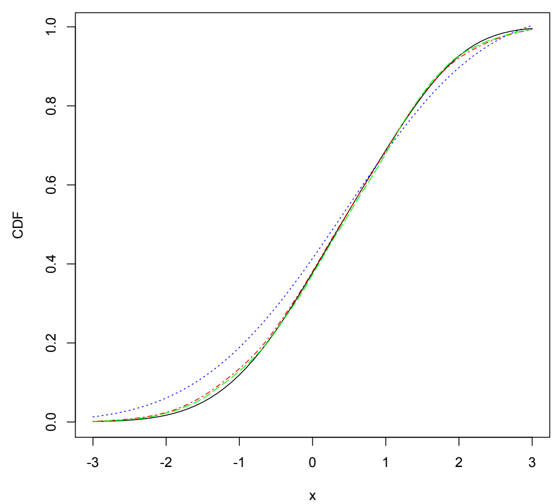

**(a)** n=25




**(b)** n=50




**(c)** n=100




**(d)** n=500

**Figure 20:** CDF estimation from $0.75N(0,1) + 0.5N(1.5, 4/9)$ for different sample size

### 4.4 Appendix

First look at the $\mathrm{E}\{\hat{F}_n(x)_{GCA}\}$.

$$\begin{aligned}
\mathrm{E}\{\hat{F}_n(t)_{GCA}\} =& \mathrm{E}\{\frac{2}{n(n+1)}\sum_{i=1}^{n} i\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)\} \\
=& \frac{2}{n(n+1)}\sum_{i=1}^{n} i\mathrm{E}\{\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)\}.
\end{aligned}$$

Now working with

$$\begin{aligned}
\mathrm{E}\{\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)\} &= \int_{-\infty}^{\infty} \bar{\mathbb{K}}\left(\frac{t-x}{ih}\right) f(x)dx \\
&= \int_{-\infty}^{\infty} \bar{\mathbb{K}}\left(\frac{t-x}{ih}\right) dF(x) \\
&= \underbrace{\bar{\mathbb{K}}\left(\frac{t-x}{ih}\right) F(x)\Big|_{x=-\infty}^{x=\infty}}_{=0} + \frac{1}{ih}\int F(x)K\left(\frac{t-x}{ih}\right) dx \\
&= \frac{1}{ih}\int F(x)K\left(\frac{t-x}{ih}\right) dx \\
&= \frac{1}{ih}\int F(t-\mu ih)K(\mu)(ih)d\mu \\
&= \int \{F(t) - \mu ih F^{(1)}(t) + \frac{(\mu ih)^2}{2}F^{(2)}(t)\}K(\mu)d\mu \\
&= F(t) + \frac{i^2h^2}{2}\mu_2(K)F^{(2)}(t) + o(n^2h^2).
\end{aligned}$$

Then

$$\begin{aligned}
\mathrm{E}\{\hat{F}_n(t)_{GCA}\} =& \frac{2}{n(n+1)}\sum_{i=1}^{n} i\mathrm{E}\{\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)\} \\
=& F(t) + \frac{n^2h^2}{2}\mu_2(K)F^{(2)}(t) + o(n^2h^2),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\{\hat{F}_n(t)_{LCA}\} =& \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\{\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)\} \\
=& F(t) + \frac{n^2h^2}{6}\mu_2(K)F^{(2)}(t) + o(n^2h^2).
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}\{\bar{\mathbb{K}}^2\left(\frac{t-X_i}{ih}\right)\} &= \int_{-\infty}^{\infty} \bar{\mathbb{K}}^2\left(\frac{t-x}{ih}\right) f(x)dx \\
&= \int_{-\infty}^{\infty} \bar{\mathbb{K}}^2\left(\frac{t-x}{ih}\right) dF(x) \\
&= \underbrace{\bar{\mathbb{K}}^2\left(\frac{t-x}{ih}\right) F(x)\Big|_{x=-\infty}^{x=\infty}}_{=0} + \frac{2}{ih}\int F(x)\bar{\mathbb{K}}\left(\frac{t-x}{ih}\right) K\left(\frac{t-x}{ih}\right) dx \\
&= 2\int \bar{\mathbb{K}}(\mu)K(\mu)F(t-\mu ih)d\mu \\
&= 2\int \bar{\mathbb{K}}(\mu)K(\mu)\{F(t)-\mu ihF^{(1)}(t)\}d\mu + o(h^2n^2) \\
&= F(t)\int d\bar{\mathbb{K}}(t) - 2ih\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(n^2h^2) \\
&= F(t) - 2ih\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(n^2h^2),
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}\{\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)\} &= \mathrm{E}\{\bar{\mathbb{K}}^2\left(\frac{t-X_i}{ih}\right)\} - [\mathrm{E}\{\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)\}]^2 \\
&= F(t) - 2ih\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu - [F(t) + \frac{i^2h^2}{2}\mu_2(K)F^{(2)}(t)]^2 \\
&= F(t) - F^2(t) - 2ih\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(nh).
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathrm{Var}\hat{F}_n(t)_{GCA} &= \mathrm{Var}[\frac{2}{(n+1)n}\sum_{i=1}^{n} i\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)] \\
&= \frac{4}{n^2(n+1)^2}\mathrm{Var}\sum_{i=1}^{n} i\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right) \\
&= \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n}\mathrm{Var}[i\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)] \\
&= \frac{4}{3n}F(t)(1-F(t)) - 2h\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(nh),
\end{aligned}
$$

and

$$\begin{aligned}
\text{Var}\hat{F}_n(t)_{LCA} =& \text{Var}[\frac{1}{n}\sum_{i=1}^{n} i\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)] \\
=& \frac{1}{n^2}\text{Var}\sum_{i=1}^{n}\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right) \\
=& \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}[\bar{\mathbb{K}}\left(\frac{t-X_i}{ih}\right)] \\
=& \frac{1}{n}F(t)(1-F(t)) - h\int \bar{\mathbb{K}}(\mu)K(\mu)\mu d\mu + o(nh).
\end{aligned}$$

This completes the proof of Theorem 4.3.1.

# 5 Future Research

In the future, I will concentrate on several issues pertaining to the **GCA** and **LCA** density estimation, including their empirical likelihood version. Specifically, I propose to:

1. Study the empirical likelihood versions of regression and **CDF**, based on the new estimates **GCA** and **LCA**.

2. Extend the settings of **GCA** and **LCA** to multivariate density estimation; study the properties of high dimension versions of **GCA** and **LCA**; investigate how dimensions affect MISE or MSE and compare with standard multivariate **KDE**; also, address data-driven bandwidth selection methods, such as cross-validation, plug-in, and contrast methods.

3. Apply **GCA** and **LCA** estimations in some hypothesis testing, such as testing goodness of fit, symmetry, and independence; evaluate power gains for both these estimates and their empirical likelihood versions over standard **KDE**.

4. Generalize the rank versions of **GCA** and **LCA** estimate settings; for example, let $X_1, \cdots, X_n$ be a random sample from unknown distribution $F$ with **PDF** $f$, and let $R_1, \cdots, R_n$ be the rank of $X_i$, then the rank versions of **GCA** and

**LCA** estimators can be defined as:

$$\hat{f}_{RGCA}(x) = \frac{2}{n(n+1)h} \sum_{i=1}^{n} K\left(\frac{x - X_i}{R_i h}\right) \tag{5.1}$$

$$\hat{f}_{RLCA}(x) = \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{R_i} K\left(\frac{x - X_i}{R_i h}\right) \tag{5.2}$$

5. Address the bias reduction methods for the **GCA** and **LCA** estimates using random transformation, both in the univariate and multivariate; study the density estimations **GCA** and **LCA** at boundaries for densities with finite support.

# BIBLIOGRAPHY


Ahmad, I. A. (1982). Nonparametric estimation of the location and scale parameters based on density estimation. *Annals of the Institute of Statistical Mathematics 34*(1), 39–53.

Ahmad, I. A. (2000). Testing exponentiality against positive ageing using kernel methods. *Sankhy: The Indian Journal of Statistics, Series A 62*, 244–257.

Ahmad, I. A. (2002). On moment inequalities of the supremum of empirical processes with applications to kernel estimation. *Statistics and Probability Letters 57*(3), 215 – 220.

Ahmad, I. A. and M. Amezziane (2011). Estimation of location and scale parameters based on kernel functional estimators. In D. Hunter, D. Richards, and J. Rosenberger (Eds.), *Nonparametric Statistics And Mixture Models* (1 ed.)., pp. 1–14. World Scientific Publishing Company.

Ahmad, I. A. and Q. Li (1997a). Testing independence by nonparametric kernel method. *Statistics and Probability Letters 34*, 201 – 210.

Ahmad, I. A. and Q. Li (1997b). Testing symmetry of an unknown density function by kernel method. *Journal of Nonparametric Statistics 7*, 279–293.

Ahmad, I. A. and I. S. Ran (2004). Data based bandwidth selection in kernel density estimation with parametric start via kernel contrasts. *Journal of Nonparametric Statistics 16*(6), 841–877.

Alez, G. (2012). *Kernel Density Estimation: Introduction, Bandwidth Selection, Statistical Implementations, and More.* Webster's Digital Services.

Aubuchon, J. C. and T. P. Hettmansperger (1984). A note on the estimation of the integral of 2(x). *Journal of Statistical Planning and Inference 9*, 321 – 331.

Bickel, P. J. and M. Rosenblatt (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics 1*, 1071–1095.

Birge, L. and P. Massart (1995). Estimation of integral functionals of a density. *The Annals of Statistics 23*, 11–29.

Bowman, A. and A. Azzalini (1997). *Applied Smoothing Techniques for Data Analysis : The Kernel Approach with S-Plus Illustrations: The Kernel Approach with S-Plus Illustrations*. Oxford Science Publications. OUP Oxford.

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

Cao, R. and I. v. Keilegom (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique 34*, 61–77.

Chen, S. (1994a). Comparing empirical likelihood and bootstrap hypothesis tests. *Journal of Multivariate Analysis 51*(2), 277–293.

Chen, S. and I. V. Keilegom (2009). A review on empirical likelihood methods for regression. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research 18*, 415–447.

Chen, S. X. (1994b). Empirical likelihood confidence intervals for linear regression coefficients. *Journal of Multivariate Analysis 49*(1), 24–40.

Chen, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika 83*(2), 329–341.

Chen, S. X. (1997). Empirical likelihood-based kernel density estimation. *Australian Journal of Statistics 39*, 47–56.

Chen, S. X. and P. Hall (1993). Smoothed empirical likelihood confidence intervals for quantiles. *The Annals of Statistics 21*(3), 1166–1181.

Chen, S. X. and Y. S. Qin (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika 87*(4), 946–953.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association 74*(368), pp. 829–836.

Devroye, D., J. Beirlant, R. Cao, R. Fraiman, P. Hall, M. Jones, G. Lugosi, E. Mammen, J. Marron, C. Snchez-Sellero, J. Ua, F. Udina, and L. Devroye (1997). Universal smoothing factor selection in density estimation: theory and practice. *Test 6*(2), 223–320.

Devroye, L. and G. Lugosi (2001). *Combinatorial Methods in Density Estimation*. Springer Series in Statistics.

Devroye, Luc; Gyorfi, L. and G. Lugosi (1996). *A probabilistic theory of pattern recognition*. New York;Berlin: Springer.

DiCiccio, T., P. Hall, and J. Romano (1991). Empirical likelihood is bartlett-correctable. *The Annals of Statistics 19*(2), 1053–1061.

Gasser, T. and H.-G. Mller (1979). Kernel estimation of regression functions. In T. Gasser and M. Rosenblatt (Eds.), *Smoothing Techniques for Curve Estimation*, Volume 757 of *Lecture Notes in Mathematics*, pp. 23–68. Springer Berlin Heidelberg.

Hall, P., J. S. Marron, and B. U. Park (1992). Smoothed cross-validation. *Probability Theory and Related Fields 92*, 1–20.

Hardle, W. (1990). *Applied nonparametric regression*, Volume 27. Cambridge Univ Press.

Heidenreich, N.-B., A. Schindler, and S. Sperlich (2013, October). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA Advances in Statistical Analysis 97*(4), 403–433.

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

Jones, M. C., J. S. Marron, and B. U. Park (1991). A simple root n bandwidth selector. *The Annals of Statistics 19*, 1919–1932.

Klemelä, J. (2009). *Smoothing of multivariate data: density estimation and visualization*, Volume 737. Wiley.

Li, Q. and J. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Nadaraya, E. (1964). On estimating regression. *Theory of Probability and Its Applications 9*, 141–142.

Otsu, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression models. *Journal of Econometrics 142*(1), 508 – 538.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*(1), 90–120.

Owen, A. (2001). *Empirical likelihood*. New York: Chapman and Hall.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237–249.

Pagan, A. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge University Press.

Park, B. U. and J. S. Marron (1992). On the use of pilot estimators in bandwidth selection. *J. Nonparametr. Statist. 1*(3), 231–240.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics 33*, 1065–1076.

Pitman, E. J. G. (1939). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika 30*(3/4), 391–421.

Pons, O. (2011). *Functional estimation for density, regression models and processes*. World Scientific Publishing Company Incorporated.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*(1), 300–325.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics 27*, 832–837.

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *The Annals of Statistics 3*, 1–14.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics 9*(2), pp. 65–78.

Scott, D. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley.

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

Sheather, S. J. and M. C. Jones (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. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability.

Simonoff, J. (1996). *Smoothing Methods in Statistics*. Springer Series in Statistics. Springer.

Stone, C. J. (1977). Consistent nonparametric regression. *The annals of statistics 5*(4), 595–620.

Tate, R. F. (1959). Unbiased estimation: Functions of location and scale parameters. *The Annals of Mathematical Statistics 30*(2), 341–366.

Terrell, G. R. (1990). The maximal smoothing principle in density estimation. *Journal of the American Statistical Association 85*(410), pp. 470–477.

Wand, M. and M. Jones (1994). *Kernel Smoothing*. Chapman and Hall/CRC.

Watson, G. S. (1964). Smooth regression analysis. *Sankhy: The Indian Journal of Statistics, Series A (1961-2002) 26*, 359–372.

Watson, G. S. and M. R. Leadbetter (1963). On the estimation of the probability density, 1. *The Annals of Mathematical Statistics 34*, 480–491.

Wertz, W. (1978). *Statistical Density Estimation: A Survey*. Angewandte Statistik und Ökonometrie. Vandenhoeck und Ruprecht.

# VITA

Ningning Wang

Candidate for the Degree of

Doctor of Philosophy/Education

Thesis: NONPARAMETRIC EMPIRICAL DENSITY FUNCTIONALS ESTIMATION AND APPLICATIONS

Major Field: Statistics

Biographical:

Completed the requirements for the Doctor of Philosophy in Statistics at Oklahoma State University, Stillwater, Oklahoma in December, 2013.

Completed the requirements for the Master of Science in Mathematics at Huazhong University Science and Technology, Wuhan, Hubei, China, July, 2004.

Completed the requirements for the Bachelor of Science in Mathematics at Yantai Normal University (Ludong University), Yantai, Shandong, July, 2001.

Experience:

Teaching Assistant at Oklahoma State University, Aug. 2009–Dec. 2013

Instructor of record at Oklahoma State University

Elementary Statistics (STAT 2013) Spring 2013

Intermediate Statistical Analysis (STAT 3013) Summer 2013

Professional Memberships:

American Statistical Association (ASA) 2013-present