# BIAS REDUCTION FOR BOUNDARY-FREE KERNEL ESTIMATORS

リズキー，レザ，ファウジ



出版情報：Kyushu University, 2020, 博士（数理学），課程博士
バージョン：
権利関係：

# BIAS REDUCTION FOR BOUNDARY-FREE KERNEL ESTIMATORS

**Rizky Reza Fauzi**

Supervisor: Yoshihiko Maesono

June 4, 2020

# Contents

# Chapter 1

# Introduction

Nonparametric methods are gradually becoming popular in statistical analysis for analyzing problems in many fields, such as economics, biology, and actuarial science. In most cases, this is because of a lack of information on the variables being analyzed. Smoothing concerning functions, such as density or cumulative distribution, plays a special role in nonparametric analysis. Knowledge on a density function, or its estimate, allows one to characterize the data more completely. We can derive other characteristics of a random variable from an estimate of its density function, such as the probability itself, hazard rate, mean, and variance value. Furthermore from distribution function estimate, we may analyze other probabilistic behaviours such as mean residual life function, or even testing the ruling distribution itself.

## 1.1 Standard kernel methods

Let $X_1, X_2, ..., X_n$ be independently and identically distributed random variables with an absolutely continuous distribution function $F_X$ and a density $f_X$. The simplest nonparametric estimator of $f_X$ is the histogram, which, even though does not enjoy satisfiable properties, can give a preliminary insight before further analysis. On the other hand, we have a quite nice classical estimator of $F_X$, which is the empirical distribution function defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x), \quad x \in \mathbb{R}, \tag{1.1}$$

where $I(A)$ denotes the indicator function of a set $A$. It is obvious that $F_n$ is a step function of height $n^{-1}$ at each observed sample point $x_i$. When considered as a pointwise estimator, $F_n(x)$ is an unbiased and strongly consistent estimator of $F_X(x)$. For the global point of view, the Glivenko-Cantelli Theorem implies that $\sup_{x\in\mathbb{R}} |F_n(x) - F_X(x)| \to_{a.s.} 0$. For details, see section 2.1 of Sterfling (1980). However, given the information that $F_X$ is absolutely continuous, it seems to be more appropriate to use a smooth and continuous estimator of $F_X$ rather than the empirical distribution function $F_n$.

Parzen (1962) and Rosenblatt (1956) introduced the kernel density estimator (we will call it the standard or naive one) as a smooth and continuous estimator of density functions. It is defined as

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}, \tag{1.2}$$

where $K$ is a function called a "kernel", and $h > 0$ is the bandwidth, which is a parameter that controls the smoothness of $\widehat{f}_h$. It is usually assumed that $K$ is a symmetric (about 0) continuous nonnegative function with $\int_{-\infty}^{\infty} K(v)\mathrm{d}v = 1$, as well as $h \to 0$ and $nh \to \infty$ when $n \to \infty$. It is easy to prove that the standard kernel density estimator is continuous and satisfies all the properties of a density function.

Since distribution function is actually an integral of density function, this kernel density estimator gave an idea to define a kernel distribution function estimator. Nadaraya (1964) defined it as

$$\widehat{F}_h(x) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}, \tag{1.3}$$

where $W(v) = \int_{-\infty}^{v} K(w)\mathrm{d}w$. It is easy to prove that this kernel distribution function estimator is continuous and satisfies all the properties of a distribution function. Several properties of $\widehat{F}_h(x)$ are well known. The almost sure uniform convergence of $\widehat{F}_h$ to $F_X$ was proved by Nadaraya (1964), Winter (1973), and Yamato (1973), while Yukich (1989) extended this result to higher dimensions. Watson and Leadbetter (1964) proved the asymptotic normality of $\widehat{F}_h(x)$, and Chung-Smirnov Property was established by Winter (1979) and Degenhardt (1993), i.e.

$$\limsup_{n\to\infty} \sqrt{\frac{2n}{\log\log n}} \sup\left\{|\widehat{F}_h(x) - F_X(x)| \Big| x \in \mathbb{R}\right\} = 1 \quad \text{a.s.}$$

Moreover, Several authors showed that the asymptotic performance of $\widehat{F}_h(x)$ is better than that of $F_n(x)$, see Azzalini (1981), Reiss (1981), Falk (1983), Singh *et al.* (1983), Hill (1985), Swanepoel (1988), Shirahata and Chu (1992), and Abdous (1993).

A typical general measure of the accuracy of $\widehat{f}_h(x)$ is the mean integrated squared error, defined as

$$MISE(\widehat{f}_h) = E\left[\int_{-\infty}^{\infty} \{\widehat{f}_h(x) - f_X(x)\}^2 w(x)\mathrm{d}x\right], \tag{1.4}$$

where $w$ is a weight function (replace $f_X$ with another function and $\widehat{f}_h$ with its estimator for MISE in general). In this thesis, we consider only $w(x) = 1$. For point-wise measures of accuracy, we will use bias, variance, and the mean squared error $MSE[\widehat{f}_h(x)] = E[\{\widehat{f}_h(x) - f_X(x)\}^2]$. It is well known that the MISE and the MSE can be computed with

$$MISE(\widehat{f}_h) = \int_{-\infty}^{\infty} MSE[\widehat{f}_h(x)]\mathrm{d}x \tag{1.5}$$

$$MSE[\widehat{f}_h(x)] = Bias^2[\widehat{f}_h(x)] + Var[\widehat{f}_h(x)]. \tag{1.6}$$

Under the condition that $f_X$ has a continuous second order derivative $f_X''$, it has been proved by the above-mentioned authors that, as $n \to \infty$,

$$Bias[\widehat{f}_h(x)] = \frac{h^2}{2} f_X''(x) \int u^2 K(u)\mathrm{d}u + o(h^2) \tag{1.7}$$

$$Var[\widehat{f}_h(x)] = \frac{f_X(x)}{nh} \int K^2(u)\mathrm{d}u + o\left(\frac{1}{nh}\right) \tag{1.8}$$

and

$$Bias[\widehat{F}_h(x)] = h^2 \frac{f_X'(x)}{2} \int_{-\infty}^{\infty} z^2 K(z)\mathrm{d}z + o(h^2) \tag{1.9}$$

$$Var[\widehat{F}_h(x)] = \frac{1}{n} F_X(x)[1 - F_X(x)] - \frac{2h}{n} r_1 f_X(x) + o\left(\frac{h}{n}\right), \tag{1.10}$$

where $r_1 = \int_{-\infty}^{\infty} yK(y)W(y)\mathrm{d}y \geq 0$.

There have been many proposals in the literature for improving the bias property of the standard kernel density estimator. Typically, under sufficient smoothness conditions placed on the underlying density $f_X$, the bias is reduced from $O(h^2)$ to $O(h^4)$, and the variance remains in the order of $n^{-1}h^{-1}$. Those methods that could potentially have greater impact include bias reduction by geometric extrapolation by Terrel and Scott (1980), variable bandwidth kernel estimators by Abramson (1982), variable location estimators by Samiuddin and El-Sayyad (1990), nonparametric transformation estimators by Ruppert and Cline (1994), and multiplicative bias correction estimators by Jones *et al.* (1995). One also could use, of course, the so-called higher order kernel functions, but this method has a disadvantage in that negative values might appear in the density estimates and distribution function estimates.

Because of the good performances of the method of Terrel and Scott for density estimator, in section 3 we use a similar idea to improve the standard kernel distribution function estimator. However, instead of using a fixed multiplication factor for the bandwidth, we use a general term for that. It can be shown that the proposed estimator, $\widetilde{F}_X$, has a smaller bias in the sense of convergence rate, that is $O(h^4)$. Furthermore, even though the rate of convergence of variance does not change, the variance of our proposed method is smaller up to some constants. Conclusively, our proposed estimator has improved MISE.

## 1.2 Boundary problem and Chen's method

All of the previous explanations implicitly assume that the true density is supported on the entire real line. If we deal with a nonnegative supported distribution, the standard kernel density estimator will suffer the so-called boundary bias problem. In this setting, the interval $[0, h]$ is called a "boundary region", and points greater than $h$ are called "interior points".

In the boundary region, the standard kernel density estimator $\widehat{f}_h(x)$ usually underestimates $f_X(x)$. This is because it does not "feel" the boundary, and it puts weights for the lack of data on the negative axis. To be more precise, if we use a symmetric kernel supported on $[-1, 1]$, we have

$$Bias[\widehat{f}_h(x)] = \left[\int_{-1}^{c} K(u)\mathrm{d}u - 1\right] f_X(x) - hf'_X(x)\int_{-1}^{c} uK(u)\mathrm{d}u + O(h^2)$$

when $x \leq h$, where $c = xh^{-1}$. This means that the standard kernel density is not consistent at $x = 0$ because

$$\lim_{n\to\infty} Bias[\widehat{f}_h(0)] = \left[\int_{-1}^{c} K(u)\mathrm{d}u - 1\right] f_X(0) \neq 0,$$

unless $f_X(0) = 0$.

Several ways of removing the boundary bias problem in density estimator, each with their own advantages and disadvantages, are data reflection (Schuster 1985), simple nonnegative boundary correction (Jones and Foster 1996), boundary kernels (Müller 1991; Müller 1993; Müller and Wang 1994), pseudodata generation (Cowling and Hall 1996), a hybrid method (Hall and Wehrly 1991), empirical transformation (Marron and Ruppert 1994), a local linear estimator (Lejeune and Sarda 1992; Jones 1993), data binning and a local polynomial fitting on the bin counts (Cheng *et al.* 1997), and others.

Most of them use symmetric kernel functions as usual, and then modify their forms or transform the data.

Chen (2000) proposed a simple way to circumvent the boundary bias that appears in the standard kernel density estimation. The remedy consists in replacing symmetric kernels with asymmetric gamma kernels, which never assign a weight outside of the support. In addition to satisfactory asymptotic features, Chen reported good finite sample performances of this cure through a simulation study.

Let $K(y; x, h)$ be an asymmetric function parameterized by $x$ and $h$, called an "asymmetric kernel". Then, the definition of the asymmetric kernel density estimator is

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K(X_i; x, h). \tag{1.11}$$

Since the density of $Gamma(xh^{-1} + 1, h)$,

$$\frac{y^{\frac{x}{h}} e^{-\frac{y}{h}}}{\Gamma\left(\frac{x}{h}+1\right) h^{\frac{x}{h}+1}}, \tag{1.12}$$

is an asymmetric function parameterized by $x$ and $h$, it is natural to use it as an asymmetric kernel. Hence, Chen defined his first gamma kernel density estimator as

$$\widehat{f}_C(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^{\frac{x}{h}} e^{-\frac{X_i}{h}}}{\Gamma\left(\frac{x}{h}+1\right) h^{\frac{x}{h}+1}}. \tag{1.13}$$

The intuitive approach to seeing how Equation (1.13) can be used as a consistent estimator is as follows. Let $Y$ be a $Gamma(xh^{-1} + 1, h)$ random variable with the pdf stated in Equation (1.12); then,

$$E[\widehat{f}_C(x)] = \int_0^\infty f_X(y) K(y; x, h) \mathrm{d}y = E[f_X(Y)].$$

By Taylor expansion,

$$E[f_X(Y)] = f_X(x) + h\left[f'_X(x) + \frac{1}{2} x f''_X(x)\right] + o(h),$$

which will converge to $f_X(x)$ as $n \to \infty$. For a detailed theoretical explanation regarding the consistency of asymmetric kernels, see Bouezmarni and Scaillet (2005).

The bias and variance of Chen's first gamma kernel density estimator are

$$Bias[\widehat{f}_C(x)] = \left[f'_X(x) + \frac{1}{2}xf''_X(x)\right]h + o(h) \tag{1.14}$$

$$Var[\widehat{f}_C(x)] = \begin{cases} \frac{f_X(x)}{2\sqrt{\pi x}n\sqrt{h}}, & \frac{x}{h} \to \infty \\ \frac{\Gamma(2\kappa+1)f_X(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)nh}, & \frac{x}{h} \to c, \end{cases} \tag{1.15}$$

for some $c > 0$. Since the result is quite similar, we do not discuss Chen's second gamma kernel density estimator in this thesis; consult Chen (2000) for reference.

Chen's gamma kernel density estimator obviously solved the boundary bias problem because the gamma pdf is a nonnegative supported function, so no weight will be put on the negative axis. However, it also has some problems; they are:

- The variance depends on a factor $x^{-1/2}$ in the interior, which means the variance becomes much larger quickly when $x$ is small,
- Zhang (2010) showed that the MSE is $O(n^{-2/3})$ when $x$ is close to the boundary (worse than the standard kernel density estimator).

In this thesis, we try to improve Chen's estimator. Using a similar idea but with different parameters of gamma density as a kernel function, we intend to reduce the variance. Then, we strive to reduce the bias by modifying it with expansions of exponential and logarithmic functions. Hence, our modified gamma kernel density estimator is not only free of the boundary bias, but the variance also has smaller orders both in the interior and near the boundary, compared with Chen's method. As a result, the optimal orders of the MSE and the MISE are smaller as well.

## 1.3 Goodness-of-fit tests

Many statistical methods depend on an assumption that the data under consideration are drawn from a certain distribution, or at least from a distribution that is approximately similar to that particular distribution. For example, test of normality for residuals are needed after fitting a linear regression in order to satisfy the normality assumption of the model. Distributional assumption is important because, in most cases, it dictates the methods that can be used to estimate the unknown parameters and also determines the procedures that staticticians may apply. There are some goodness-of-fit tests available to determine whether a sample comes from the assumed distribution. Those popular tests include the Kolmogorov-Smirnov (KS) test, Cramér-von Mises (CvM) test, Anderson-Darling test, and Durbin-Watson test. In this thesis, we will be focusing ourselves to the KS and CvM tests.

In this setting, the Kolmogorov-Smirnov statistic utilizes the empirical distribution function $F_n$ to test the null hypothesis

$$H_0 : F_X = F$$

againsts the alternative hypothesis

$$H_1 : F_X \neq F,$$

where $F$ is the assumed distribution function. The test statistic is defined as

$$KS_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|. \tag{1.16}$$

If under a significance level $\alpha$ the value of $KS_n$ is larger than a certain value from Kolmogorov distribution table, we will reject $H_0$. Likewise, under the same circumstance, the statistic of the Cramér-von Mises test is defined as

$$CvM_n = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \mathrm{d}F(x), \tag{1.17}$$

and we reject the null hypothesis when the value of $CvM_n$ is larger than a certain value from Cramér-von Mises table.

Several discussions regarding those goodness-of-fit tests have been around for decades. The recent articles include the distribution of KS and CvM tests for exponential populations (Evans *et al.* 2017), revision of two-sample KS test (Finner and Gontscharuk 2018), KS test for mixed distributions (Zierk *et al.* 2020), KS test for bayesian ensembles of phylogenies (Antoneli *et al.* 2018), CvM distance for neighbourhood-of-model validation (Baringhaus and Henze 2016), rank-based CvM test (Curry *et al.* 2019), and model selection using CvM distance in a fixed design regression (Chen *et al.* 2018).

Though the standard KS and CvM tests work really well, but it does not mean they bear no problem. The lack of smoothness of $F_n$ causes too much sensitivity near the center of distribution, especially when $n$ is small. Hence, it is not unusual to find the supremum value of $|F_n(x) - F(x)|$ is attained when $x$ is near the center of distribution, or the value of $CvM_n$ gets larger because $[F_n(x) - F(x)]^2$ is large when the data is highly concentrated in one area. Furthermore, given the information that $F_X$ is absolutely continuous, it seems to be more appropriate to use a smooth and continuous estimator rather than the empirical distribution function for testing the goodness-of-fit.

It is natural if one uses the naive kernel distribution function estimator in place of the empirical distribution function to smooth the KS and CvM statistics out. By doing that, we may expect to eliminate the over-sensitivity that standard KS and CvM statistics have. Therefore, the formulas become

$$\widehat{KS} = \sup_{x\in\mathbb{R}} |\widehat{F}_X(x) - F(x)| \tag{1.18}$$

and

$$\widehat{CvM} = n \int_{-\infty}^{\infty} [\widehat{F}(x) - F(x)]^2 \mathrm{d}F(x). \tag{1.19}$$

Omelka *et al.* (2009) proved that under the null hypothesis, the distribution of those statistics converge to the same distributions as the standard ones.

Though both tests are versatile in most settings, but when the support of the data is strictly smaller than the entire real line (let say the support is an interval $\Omega \subset \mathbb{R}$), the naive kernel distribution function estimator also suffers the boundary problem, as in the naive kernel density estimator. Even though in some cases (e.g. $f_X(0) = 0$ when 0 is the boundary point) the boundary effects of $\widehat{F}_X(x)$ is not as severe as in the kernel density estimator, but the problem still occurs. It is because the value of $\widehat{F}_X(x)$ is still larger than 0 (or less than 1) at the boundary points. This phenomena cause large value of $|\widehat{F}_X(x) - F(x)|$ in the boundary regions, and then $\widehat{KS}$ and $\widehat{CvM}$ tend to be larger than they are supposed to be, leading to the rejection of $H_0$ even though $H_0$ is right. To make things worse, chapter 4 will illustrate how this problem enlarges type-2 error by accepting the null hypothesis when it is wrong.

## 1.4 The mean residual life function

Statistical inference for remaining lifetimes would be intuitively more appealing than the popular hazard rate function, since its interpretation as "the risk of immediate failure" can be difficult to grasp. A function called the mean residual life (or mean excess loss) which represents "the average remaining time before failure" is easier to understand. The mean residual life (or MRL for short) function is of interest in many fields relating to time and finance, such as biomedical theory, survival analysis, and actuarial science.

Let us work under the same settings as in section 1.3, where the distribution is supported on an interval $\Omega \subset \mathbb{R}$, where $\inf \Omega = \omega'$, $\sup \Omega = \omega''$, and $-\infty \leq \omega' < \omega'' \leq \infty$. Also, let $S_X(t) = \Pr(X > t)$ be the survival function, $\mathbb{S}_X(t) = \int_t^\infty S_X(x)\mathrm{d}x$ be the cumulative survival function, and define a new notation $\bar{\mathbb{S}}_X(t) = \int_t^\infty \mathbb{S}_X(x)\mathrm{d}x$. Then

$$m_X(t) = E(X - t|X > t), \quad t \in \Omega \tag{1.20}$$

is the definition of the mean residual life function, or can be written as

$$m_X(t) = \frac{\mathbb{S}_X(t)}{S_X(t)}. \tag{1.21}$$

For a detailed discussion about the MRL function, see Embrechts *et al.* (1997) or Guess and Proschan (1988). Murari and Sujit (1995) and Belzunce *et al.* (1996) discussed the use of the MRL function for ordering and classifying distributions. On the other hand, Cox (1962), Kotz and Shanbhag (1980), and Zoroa *et al.* (1990) proposed how to determine distribution via an inversion formula of $m_X(t)$. Ruiz and Navarro (1994) have considered the problem of characterization of the distribution function through the relationship between the MRL function and the hazard rate function. The MRL functions of finite mixtures and order statistics have been studied as well by Navarro and Hernandez (2008).

Some properties and applications of the MRL concept related to operational research and reliability theory in engineering are interesting topics. While Nanda *et al.* (2010) discussed the properties of associated orderings in the MRL function, Huynh *et al.* (2014) studied the usefulness of the MRL models for maintenance decision-making. Another examples are the utilization of the MRL functions of parallel system by Sadegh (2008), the MRL for records by Raqab and Asadi (2008), the MRL of a $k$-out-of-$n$:G system by Eryilmaz (2012), the MRL of a $(n-k+1)$-out-of-$n$ system by Poursaeed (2010), the MRL in reliability shock models by Eryilmaz (2017), the MRL subjected to Marshall-Olkin type shocks by Bayramoglu and Ozkut (2016), the MRL of coherent systems by Eryilmaz *et al.* (2018) and Kavlak (2017), the MRL for degrading systems by Zhao *et al.* (2018), and the MRL of rail wagon bearings by Ghasemi and Hodkiewicz (2012).

The natural estimator of the MRL function is the empirical one, which is

$$m_n(t) = \frac{\mathbb{S}_n(t)}{S_n(t)} = \frac{\sum_{i=1}^n (X_i - t) I(X_i > t)}{\sum_{i=1}^n I(X_i > t)}, \quad t \in \Omega, \tag{1.22}$$

where $I(A)$ is the usual indicator function on set $A$. Yang (1978), Ebrahimi (1991), and Csörgő and Zitikis (1996) studied the properties of $m_n(t)$. Even though it has several good attributes (e.g. unbiasedness and consistency), the empirical MRL function is just a rough estimate of $m_X(t)$ and lack of smoothness. Estimating is also impossible for large $t$ because $S_n(t) = 0$ for $t > \max\{x_1, x_2, ..., x_n\}$. Though we can just define $m_n(t) = 0$ for such case, it is a major disadvantage as analysing the behaviour of the MRL function when $t \to \infty$ is of an interest.

Various parametric models of MRL have been discussed in literatures, for example the transformed parametric MRL models by Sun and Zhang (2009), the upside-down bathtub-shaped MRL model by Shen *et al.* (2009), the MRL order of convolutions of heterogeneous exponential random variables by Zhao and Balakrishnan (2009), the proportional MRL model by Nanda *et al.* (2006) and Chan *et al.* (2012), and the MRL models with time-dependent coefficients by Sun *et al.* (2012).

Some nonparametric estimators of $m_X(t)$ which are related to the empirical one have been discussed in a fair amount of literature. For example, Ruiz and Guillamón (1996) estimated the numerator in $m_n(t)$ by a recursive kernel estimate and left the empirical survival function unchanged, while Chaubey and Sen (1999) used the Hille's Theorem in Hille (1948) to smooth both the numerator and denominator in $m_n(t)$.

The other maneuver that can be used for estimating the MRL function nonparametrically is the kernel method. We need two other functions derived from the kernel $K(x)$, which are

$$V(x) = \int_x^\infty K(z)\mathrm{d}z \quad \text{and} \quad \mathbb{V}(x) = \int_x^\infty V(z)\mathrm{d}z. \tag{1.23}$$

Hence, the naive kernel MRL function estimator can be defined as

$$\widehat{m}_X(t) = \frac{\widehat{\mathbb{S}}_X(t)}{\widehat{S}_X(t)} = \frac{h\sum_{i=1}^n \mathbb{V}\left(\frac{t-X_i}{h}\right)}{\sum_{i=1}^n V\left(\frac{t-X_i}{h}\right)}, \quad t \in \Omega. \tag{1.24}$$

Guillamón *et al.* (1998) discussed the asymptotic properties of the naive kernel MRL function estimator in detail.

However, as usually $m_X(t)$ is used for time or finance related data, which are on nonnegative real line or bounded interval, the naive kernel MRL function estimator suffers the boundary bias problem as well. In the case of $f_X(\omega') = 0$ (or $f_X(\omega'') = 0$), the boundary effects of $\widehat{m}_X(t)$ when $t \to \omega'$ (or $t \to \omega''$) is not as bad as in the kernel density estimator, but the problems still occur. It is because the term $S_X(\omega')$ and $1 - S_X(\omega'')$ in the $Bias[\widehat{\mathbb{S}}_X(\omega')]$ and $Bias[\widehat{\mathbb{S}}_X(\omega'')]$ can never be 0 since $S_X(\omega') = 1 - S_X(\omega'') = 1$, which means $\widehat{\mathbb{S}}_X(t)$ causes the boundary problems for $\widehat{m}_X(t)$. Moreover, in the case of $f_X(\omega') > 0$ and $f_X(\omega'') > 0$ (e.g. uniform distribution), not only $\widehat{\mathbb{S}}_X(t)$, but $\widehat{S}_X(t)$ also adds its share to the boundary problems for $\widehat{m}_X(t)$.

To make things worse, the naive kernel MRL function estimator does not preserve one of the most important properties of the MRL function, which is $m_X(\omega') + \omega' = E(X)$. It is reasonable if we expect $\widehat{m}_X(\omega') + \omega' \approx \bar{X}$. However, $\widehat{S}_X(\omega')$ is less than 1 and $\widehat{\mathbb{S}}_X(\omega')$ is smaller than the average value of $X_i's$, due to the weight that they still put on the outside of $\Omega$. Accordingly, there is no guarantee of how far or how close $\widehat{m}_X(\omega') + \omega'$ is to $\bar{X}$.

Though Abdous and Berred (2005) successfully adopted the idea of local linear fitting for the MRL function estimation, in this thesis we are going to try bijective transformation idea to remove the boundary effects. In this situation there are no boundary effects at all, as we will not put any weight outside the support.

# Chapter 2

# New Type of Gamma Kernel Density Estimator

In this chapter, we will start our discussion with the formulation of the modified gamma kernel in detail. First we try to use another parameters in gamma density function and then derive its properties. Then, we modify it using some expansions, and calculate further about its new asymptotic properties. At last, we will show the simulation result and compare three kinds of kernel density estimator.

## 2.1 New type of gamma kernel density estimator formulation

Before starting our discussion, we need to impose assumptions; they are:

**A1**. The bandwidth $h > 0$ satisfies $h \to 0$ and $nh \to \infty$ when $n \to \infty$,

**A2**. The density $f_X$ is three times continuously differentiable, and the fourth derivative $f_X^{(4)}$ exists,

**A3**. The following integrals $\int \left[\frac{f'_X(x)}{f_X(x)}\right]^2 \mathrm{d}x$, $\int x^4 \left[\frac{f''_X(x)}{f_X(x)}\right]^2 \mathrm{d}x$, $\int x^2 [f''_X(x)]^2 \mathrm{d}x$, and $\int x^6 [f'''_X(x)]^2 \mathrm{d}x$ are finite.

The first assumption is the usual assumption for the standard kernel density estimator. Since we will use exponential and logarithmic expansions, we need A2 to ensure the validity of our proofs. The last assumption is necessary to make sure we can calculate the MISE.

As we stated before, the modification of the gamma kernel is started by replacing the shape and the scale parameters of the gamma density with suitable functions of $x$ and $h$, and this kernel is defined as a new gamma kernel. Our purpose in doing this is to reduce the variance so that it is smaller than the variance of Chen's method. After trying several combinations of functions, we chose the density of $Gamma(h^{-1/2}, x\sqrt{h}+h)$, which is

$$K(y;x,h) = \frac{y^{\frac{1}{\sqrt{h}}-1}e^{-\frac{y}{x\sqrt{h}+h}}}{\Gamma\left(\frac{1}{\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{1}{\sqrt{h}}}}, \tag{2.1}$$

as a kernel, and we define the new gamma kernel density "estimator" as

$$A_h(x) = \frac{\sum_{i=1}^{n} X_i^{\frac{1}{\sqrt{h}}-1}e^{-\frac{X_i}{x\sqrt{h}+h}}}{n\Gamma\left(\frac{1}{\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{1}{\sqrt{h}}}}, \tag{2.2}$$

where $n$ is the sample size, and $h$ is the bandwidth.

**Remark 2.1.1.** *Even though the formula in Equation* (2.2) *can work as a density estimator properly, it is* ***not our proposed method*** *(that is why we put quotation marks around the word "estimator"). As we will state later, we need another modification for Equation* (2.2) *before our proposed estimator is created.*

After this, we need to derive the bias and the variance formulas of $A_h(x)$. Consult the following theorem.

**Theorem 2.1.2.** *Assuming A1 and A2, for the function $A_h(x)$ in Equation* (2.2), *its bias and variance are*

$$Bias[A_h(x)] = \left[f'_X(x) + \frac{1}{2}x^2 f''_X(x)\right]\sqrt{h} + o(\sqrt{h}) \tag{2.3}$$

*and*

$$Var[A_h(x)] = \begin{cases} \dfrac{R^2\left(\frac{1}{\sqrt{h}}-1\right)f_X(x)}{2(x+\sqrt{h})\sqrt{\pi(1-\sqrt{h})}R\left(\frac{2}{\sqrt{h}}-2\right)nh^{\frac{1}{4}}} + O\left(\frac{h^{\frac{1}{4}}}{n}\right), & \frac{x}{h} \to \infty \\ \dfrac{R^2\left(\frac{1}{\sqrt{h}}-1\right)f_X(x)}{2(c\sqrt{h}+1)\sqrt{\pi(1-\sqrt{h})}R\left(\frac{2}{\sqrt{h}}-2\right)nh^{\frac{3}{4}}} + O\left(\frac{1}{nh^{\frac{1}{4}}}\right), & \frac{x}{h} \to c, \end{cases} \tag{2.4}$$

*for some positive number $c$ and*

$$R(z) = \frac{\sqrt{2\pi}z^{z+\frac{1}{2}}}{e^z\Gamma(z+1)}. \tag{2.5}$$

**Remark 2.1.3.** *The function $R(z)$ (Brown and Chen 1999) monotonically increases with $\lim_{z\to\infty} R(z) = 1$ and $R(z) < 1$, which means $\frac{R^2\left(\frac{1}{h}-1\right)}{R\left(\frac{2}{h}-2\right)} \leq 1$. From these facts, we can conclude that $Var[A_h(x)]$ is $O(n^{-1}h^{-1/4})$ when $x$ is in the interior, and it is $O(n^{-1}h^{-3/4})$ when $x$ is near the boundary. Both of these rates of convergence are faster than the rates of the variance of Chen's gamma kernel estimator for both cases, respectively. Furthermore, instead of $x^{-1/2}$, $Var[A_h(x)]$ depends on $(x + \sqrt{h})^{-1}$, which means the value of the variance will not speed up to infinity when $x$ approaches $0$.*

Even though we have succeeded in reducing the order of the variance, we now encounter a larger bias order. To avoid this problem, we use geometric extrapolation to change the order of bias back to $h$.

**Theorem 2.1.4.** *Let $A_h(x)$ be the function in Equation (2.2). Assuming A1 and A2, if we define $J_h(x) = E[A_h(x)]$, then*

$$[J_h(x)]^2[J_{4h}(x)]^{-1} = f_X(x) + O(h). \tag{2.6}$$

**Remark 2.1.5.** *The function $J_{4h}(x)$ is the expectation of the function in Equation (2.2) with $4h$ as the bandwidth. Furthermore, the term after $f_X(x)$ in Equation (2.6) is in the order $h$, which is the same as the order of bias for Chen's gamma kernel density estimator. This theorem will lead us to the idea to modify $A_h(x)$. We present the explicit asymptotic formula of $O(h)$ in the appendices.*

Theorem 2.1.4 gives us the idea to modify $A_h(x)$ and to define our new estimator. Hence, we propose

$$\widetilde{f}_X(x) = [A_h(x)]^2[A_{4h}(x)]^{-1} \tag{2.7}$$

as the modified gamma kernel density estimator, our proposed method. This idea is actually straightforward. It uses the fact that the expectation of the operation of two statistics is asymptotically equal (in probability) to the operation of the expectation of each statistic. Though we do not use any concept of convergence in probability in our proofs, the idea is still applicable when using Taylor Expansion.

For the bias of our proposed estimator, we have the following theorem.

**Theorem 2.1.6.** *Assuming A1 and A2, the bias of the modified gamma kernel density estimator is*

$$Bias[\widetilde{f}_X(x)] = -2\left[b(x) - \frac{a(x)}{2f_X(x)}\right]h + o(h) + O\left(\frac{1}{nh^{\frac{1}{4}}}\right), \tag{2.8}$$

*where*

$$a(x) = f'_X(x) + \frac{1}{2}x^2 f''_X(x) \tag{2.9}$$

$$b(x) = \left(x + \frac{1}{2}\right) f''_X(x) + x^2 \left(\frac{x}{3} + \frac{1}{2}\right) f'''_X(x). \tag{2.10}$$

As expected, the bias' leading term is actually the same as the explicit form of $O(h)$ in theorem 2.1.4 (see appendices). Its order of convergence changed back to $h$, the same as the bias of Chen's method. This is quite the accomplishment because if we can keep the order of the variance the same as $Var[A_h(x)]$, we can then conclude that the MSE of our modified gamma kernel density estimator is smaller than the MSE of Chen's gamma kernel estimator. However, before jumping into the calculation of variance, we need the following theorem.

**Theorem 2.1.7.** *Assuming A1 and A2, for the function in Equation* (2.2) *with bandwidth* $h$*,* $A_h(x)$*, and with bandwidth* $4h$*,* $A_{4h}(x)$*, the covariance of them is equal to*

$$Cov[A_h(x), A_{4h}(x)] = \frac{R\left(\frac{1}{\sqrt{h}} - 1\right) R\left(\frac{1}{2\sqrt{h}} - 1\right)}{2\sqrt{\pi} R\left(\frac{3}{2\sqrt{h}} - 2\right)\left(3x + 5\sqrt{h}\right)} \frac{\left(\frac{3}{2} - 2\sqrt{h}\right)^{\frac{3}{2\sqrt{h}} - \frac{3}{2}}}{(2 - 2\sqrt{h})^{\frac{1}{\sqrt{h}} - \frac{1}{2}} (1 - 2\sqrt{h})^{\frac{1}{2\sqrt{h}} - \frac{1}{2}}}$$
$$\times \left(\frac{x + \sqrt{h}}{3x + 5\sqrt{h}}\right)^{\frac{1}{2\sqrt{h}} - 1} \left(\frac{2x + 4\sqrt{h}}{3x + 5\sqrt{h}}\right)^{\frac{1}{\sqrt{h}} - 1} \frac{f_X(x)}{nh^{\frac{1}{4}}} + O\left(\frac{h^{\frac{1}{4}}}{n}\right),$$

*when* $xh^{-1} \to \infty$*, and*

$$Cov[A_h(x), A_{4h}(x)] = \frac{R\left(\frac{1}{\sqrt{h}} - 1\right) R\left(\frac{1}{2\sqrt{h}} - 1\right)}{2\sqrt{\pi} R\left(\frac{3}{2\sqrt{h}} - 2\right)\left(3c\sqrt{h} + 5\right)} \frac{\left(\frac{3}{2} - 2\sqrt{h}\right)^{\frac{3}{2\sqrt{h}} - \frac{3}{2}}}{(2 - 2\sqrt{h})^{\frac{1}{\sqrt{h}} - \frac{1}{2}} (1 - 2\sqrt{h})^{\frac{1}{2\sqrt{h}} - \frac{1}{2}}}$$
$$\times \left(\frac{c\sqrt{h} + 1}{3c\sqrt{h} + 5}\right)^{\frac{1}{2\sqrt{h}} - 1} \left(\frac{2c\sqrt{h} + 4}{3c\sqrt{h} + 5}\right)^{\frac{1}{\sqrt{h}} - 1} \frac{f_X(x)}{nh^{\frac{3}{4}}} + O\left(\frac{1}{nh^{\frac{1}{4}}}\right),$$

*when* $xh^{-1} \to c > 0$.

**Theorem 2.1.8.** *Assuming A1 and A2, the variance of the modified gamma kernel density estimator is*

$$Var[\widetilde{f}_X(x)] = 4Var[A_h(x)] + Var[A_{4h}(x)] - 4Cov[A_h(x), A_{4h}(x)] + o\left(\frac{1}{nh^{\frac{1}{4}}}\right),$$

*where its orders of convergence are* $O(n^{-1}h^{-1/4})$ *in the interior and* $O(n^{-1}h^{-3/4})$ *in the boundary region.*

As a conclusion to theorems 2.1.6 and 2.1.8, with the identity of MSE, we have

$$MSE[\widetilde{f}_X(x)] = \begin{cases} O(h^2) + O\left(\frac{1}{nh^{\frac{1}{4}}}\right), & \frac{x}{h} \to \infty \\ O(h^2) + O\left(\frac{1}{nh^{\frac{3}{4}}}\right), & \frac{x}{h} \to c. \end{cases} \tag{2.11}$$

The theoretical optimum bandwidths are $h = O(n^{-4/9})$ in the interior and $O(n^{-4/11})$ in the boundary region. As a result, the optimum orders of convergence are $O(n^{-8/9})$ and $O(n^{-8/11})$, respectively. Both of them are smaller than the optimum orders of Chen's estimator, which are $O(n^{-4/5})$ in the interior and $O(n^{-2/3})$ in the boundary region. Furthermore, since the MISE is just the integration of MSE, it is clear that the orders of convergence of the MISE are the same as of the MSE.

Calculating the explicit formula of $MISE(\widetilde{f}_X)$ is nearly impossible because of the complexity of the formulas of $Bias[\widetilde{f}_X(x)]$ and $Var[\widetilde{f}_X(x)]$. However, there is one thing we would like to discuss regarding this matter. Using a similar argument stated by Chen (2000), the boundary region part of $Var[\widetilde{f}_X(x)]$ is negligible while integrating the variance. Thus, instead of computing $\int_{boundary} Var[\widetilde{f}_X(x)] + \int_{interior} Var[\widetilde{f}_X(x)]$, it is sufficient to just calculate $\int_0^\infty Var[\widetilde{f}_X(x)]\mathrm{d}x$ using the formula of the variance in the interior. With that, computing

$$MISE(\widetilde{f}_X) = \int_0^\infty Bias^2[\widetilde{f}_X(x)]\mathrm{d}x + \int_0^\infty Var[\widetilde{f}_X(x)]\mathrm{d}x$$

can be approximated by using numerical methods (assuming $f_X$ is known).

## 2.2 Simulation studies

In this section, we provide the results of a simulation study we did to show the performances of our proposed method and compare them with other estimators' results. The measures of error we use in this thesis are the MISE, the MSE, bias, and variance. Since we are working under assumptions A1, A2, and A3, the MISE of our proposed estimator is finite. We calculated the average integrated squared error (AISE), the average squared error (ASE), simulated bias, and simulated variance, with a sample size of $n = 50$ and 10000 repetitions for each case.

We compared four gamma kernel density estimators: Chen's gamma kernel density estimator $\widehat{f}_C(x)$, two nonnegative bias-reduced Chen's gamma estimators $\widehat{f}_{KI1}(x)$ and $\widehat{f}_{KI2}(x)$ (Igarashi and Kakizawa 2015, eq. 10 and 11), and our modified gamma kernel density estimator $\widetilde{f}_X(x)$. We generated several distributions for this study; they are exponential distribution $exp(1/2)$, gamma distribution $Gamma(2,3)$, log-normal distribution $log.N(0,1)$, inverse Gaussian distribution $IG(1,2)$, Weibull distribution $Weibull(3,2)$, and absolute normal distribution $abs.N(0,1)$. The least squares cross-validation technique was used to determine the value of the bandwidths.

Table 2.1 compares AISEs, representing the general measure of error. As we can see, the proposed method outperformed the other estimators. Since one of our main concerns is eliminating the boundary bias problem, it is necessary to take our attention to the values of the measures of error in the boundary region. Tables 2.2, 2.3, and 2.4 show the ASE, bias, and variance of those four estimators when $x = 0.01$. Once again, our estimator had the best results. Though the differences among the values of bias were relatively not big (Table 2.3), from Table 2.4, we can witness how our variance reduction has an effect.

As further illustrations, we also provide graphs of point-wise ASE, bias, squared bias, and variance to compare our estimator's performances with those of the others. We generated exponential, gamma, and absolute normal distributions 1000 times to produce Figs. 2.1, 2.2, and 2.3.

In some cases, we found that the bias value of our proposed estimator was away from 0 more than the other estimators (e.g., Fig. 2.1(a) around $x = 1$, Fig. 2.2(a) around $x = 4$, and Fig. 2.3(a) around $x = 0.2$). Though this could reflect poorly on the proposed estimator, from the variance parts (Figs. 2.1(b), 2.2(b), and 2.3(b)), we see that our estimator never failed to give the smallest value of variance, confirming that we succeeded in reducing variance with our method. Moreover, the result of the variance reduction is the reduction of point-wise ASE itself, shown in Figs. 2.1(d), 2.2(d), and 2.3(d). One may take note of Fig. 2.2(d) when $x \in [1,4]$ because the estimators of Igarashi and Kakizawa slightly outperformed the proposed method. However, as $x$ got larger, $ASE[\widehat{f}_{KI1}(x)]$ and $ASE[\widehat{f}_{KI2}(x)]$ failed to get closer to 0 (they will when $x$ is large enough), while $ASE[\widetilde{f}_X(x)]$ approached 0 immediately.

Table 2.1: Comparison of the average integrated squared error ($\times 10^5$)

| Distributions | $\widehat{f}_C(x)$ | $\widehat{f}_{KI1}(x)$ | $\widehat{f}_{KI2}(x)$ | $\widetilde{f}_X(x)$ |
|---|---|---|---|---|
| $exp(1/2)$ | 970 | 1367 | 1304 | **831** |
| $Gamma(2,3)$ | 313 | 2091 | 1913 | **196** |
| $log.N(0,1)$ | 342 | 1845 | 1688 | **206** |
| $IG(1,2)$ | 1002 | 680 | 660 | **297** |
| $Weibull(3,2)$ | 7896 | 4198 | 4120 | **1832** |
| $abs.N(0,1)$ | 8211 | 3785 | 3719 | **2905** |

Table 2.2: Comparison of the average squared error ($\times 10^5$) when $x = 0.01$

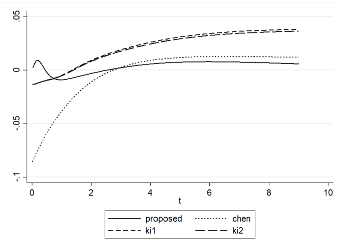
| Distributions | $\widehat{f}_C(x)$ | $\widehat{f}_{KI1}(x)$ | $\widehat{f}_{KI2}(x)$ | $\widetilde{f}_X(x)$ |
|---|---|---|---|---|
| $exp(1/2)$ | 1600 | 1547 | 1553 | **991** |
| $Gamma(2,3)$ | 207 | 384 | 359 | **168** |
| $log.N(0,1)$ | 36 | 178 | 160 | **34** |
| $IG(1,2)$ | 1006 | 829 | 781 | **422** |
| $Weibull(3,2)$ | 1528 | 708 | 643 | **304** |
| $abs.N(0,1)$ | 2389 | 2018 | 1999 | **721** |

Table 2.3: Comparison of the bias ($\times 10^4$) when $x = 0.01$

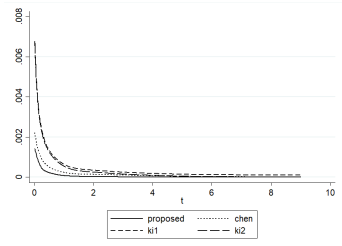
| Distributions | $\widehat{f}_C(x)$ | $\widehat{f}_{KI1}(x)$ | $\widehat{f}_{KI2}(x)$ | $\widetilde{f}_X(x)$ |
|---|---|---|---|---|
| $exp(1/2)$ | $-1054$ | $-1865$ | $-1904$ | $\mathbf{-858}$ |
| $Gamma(2,3)$ | 391 | 583 | 561 | **233** |
| $log.N(0,1)$ | 150 | 417 | 395 | **120** |
| $IG(1,2)$ | 961 | 869 | 840 | **386** |
| $Weibull(3,2)$ | 1215 | 821 | 780 | **342** |
| $abs.N(0,1)$ | $-1383$ | 303 | 297 | **157** |

Table 2.4: Comparison of the variance ($\times 10^5$) when $x = 0.01$

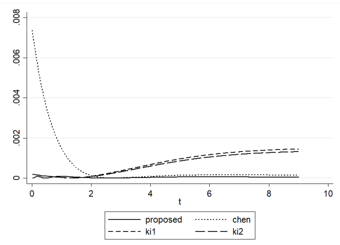
| Distributions | $\widehat{f}_C(x)$ | $\widehat{f}_{KI1}(x)$ | $\widehat{f}_{KI2}(x)$ | $\widetilde{f}_X(x)$ |
|---|---|---|---|---|
| $exp(1/2)$ | 490 | 1465 | 1469 | **244** |
| $Gamma(2,3)$ | 54 | 43 | 44 | **11** |
| $log.N(0,1)$ | 39 | 36 | 36 | **35** |
| $IG(1,2)$ | 835 | 739 | 753 | **273** |
| $Weibull(3,2)$ | 532 | 340 | 343 | **184** |
| $abs.N(0,1)$ | 476 | 1926 | 1910 | **211** |



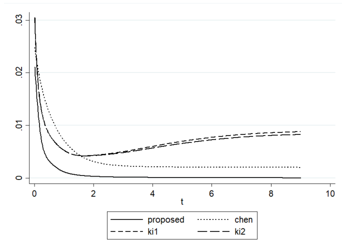

(a) bias




(b) variance

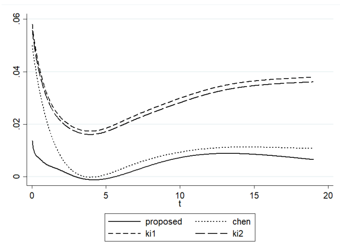



(c) squared bias

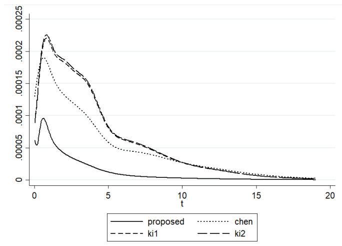



(d) average squared error

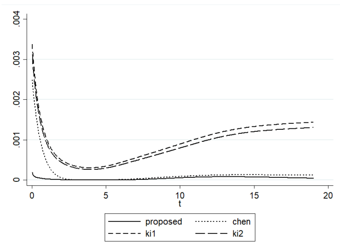
Figure 2.1: Comparison of point-wise bias, variance, and ASE of $\widetilde{f}_X(x)$, $\widehat{f}_C(x)$, $\widehat{f}_{KI1}(x)$, and $\widehat{f}_{KI2}(x)$ for estimating density of $exp(1/2)$ with sample size $150$.



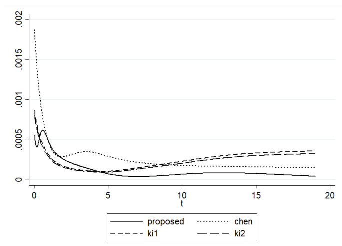

(a) bias




(b) variance

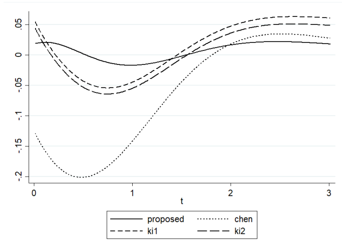



(c) squared bias

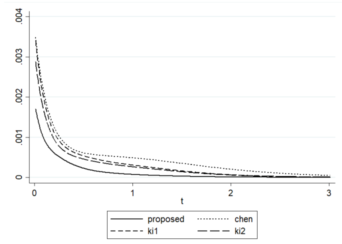



(d) average squared error

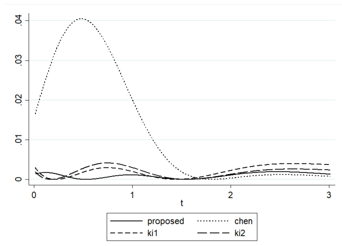
Figure 2.2: Comparison of the point-wise bias, variance, and ASE of $\widetilde{f}_X(x)$, $\widehat{f}_C(x)$, $\widehat{f}_{KI1}(x)$, and $\widehat{f}_{KI2}(x)$ for estimating density of $Gamma(2,3)$ with sample size $150$.



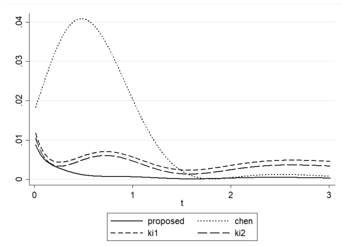

(a) bias




(b) variance




(c) squared bias




(d) average squared error

Figure 2.3: Comparison of the point-wise bias, variance, and ASE of $\widetilde{f}_X(x)$, $\widehat{f}_C(x)$, $\widehat{f}_{KI1}(x)$, and $\widehat{f}_{KI2}(x)$ for estimating density of $abs.N(0,1)$ with sample size $150$.

# Chapter 3

# Modified Kernel Distribution Function Estimator

We will start our discussion in this chapter with the derivation of our proposed distribution function estimator. After that, we present our calculation for the bias and the variance to show our estimator is theoretically better than the standard one. At last, we will show the simulation study.

## 3.1 MISE reduction by geometric extrapolation

In this section, we will apply geometric extrapolation method to the kernel distribution function estimator, in order to reduce bias. The idea of reducing bias by geometric extrapolation is doing a self-elimination technique between two standard kernel distribution function estimators with different bandwidths, with some helps of exponential and logarithmic expansions. By doing that, vanishing the $h^2$ term of the asymptotic bias is possible, and the the order of convergence changes to $h^4$.

Before starting our main purpose, we need to impose some assumptions, they are:

**B1**. The kernel $K$ is a nonnegative continuous function, symmetric about 0, and it integrates to 1,

**B2**. The integral $\int_{-\infty}^{\infty} w^4 K(x) \mathrm{d}w$ is finite,

**B3**. The bandwidth $h > 0$ satisfies $h \to 0$ and $nh \to \infty$ when $n \to \infty$,

**B4**. The density $f_X$ is three times continuously differentiable, and the fourth derivative $f_X^{(4)}$ exists,

**B5**. The integrals $\int_{-\infty}^{\infty} \frac{[f'_X(x)]^2}{F_X(x)} \mathrm{d}x$ and $\int_{-\infty}^{\infty} f'''_X(x) \mathrm{d}x$ are finite.

The first and third ones are the usual assumptions for the standard kernel distribution function estimator. Since we will use exponential and logarithmic expansions, we need B2 and B4 to ensure the validity of our proofs. For the last assumption, it is necessary to make sure we can calculate MISE.

We now ready to begin the explanation about how to modify the standard kernel distribution function estimator and reduce its bias. First, we have this following theorem.

**Theorem 3.1.1.** *Let $j_h(x) = E[\widehat{F}_h(x)]$ and $a(\neq 1)$ be a positive number. Under the assumptions B1-B4, we have*

$$[j_h(x)]^{t_1}[j_{ah}(x)]^{t_2} = F_X(x) + O(h^4), \tag{3.1}$$

*where $t_1 = \frac{a^2}{a^2-1}$ and $t_2 = -\frac{1}{a^2-1}$.*

**Remark 3.1.2.** *The function $j_{ah}(x)$ is an expectation of the standard kernel distribution function estimator with $ah$ as the bandwidth, that is, $j_{ah}(x) = E[\widehat{F}_{ah}(x)]$, where*

$$\widehat{F}_{ah}(x) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{x - X_i}{ah}\right).$$

*Furthermore, the term after $F_X(x)$ in (3.1) is in the order of $h^4$, which is smaller than the order of bias of the standard kernel distribution function estimator. Even though this theorem does not state about a bias of some estimator, it will lead us to the idea to modify the standard kernel distribution function estimator. About the explicit asymptotic formula of $O(h^4)$, we will present it in the appendices.*

The theorem 3.1.1 gives us an idea to modify kernel distribution function estimator which will have, intuitively, similar property for bias. Hence, we propose a new estimator of distribution function as

$$\widetilde{F}_X(x) = [\widehat{F}_h(x)]^{\frac{a^2}{a^2-1}}[\widehat{F}_{ah}(x)]^{-\frac{1}{a^2-1}}. \tag{3.2}$$

**Remark 3.1.3.** *As we can see, the number $a$ acts as the second smoothing parameter here, because it controls the smoothness of $\widehat{F}_{ah}$ (since it is placed inside the function $W$) and determines how much the effect of $\widehat{F}_h$ and $\widehat{F}_{ah}$ as a part of their power. Larger $a$ means the effect of $\widehat{F}_h$ is larger for $\widetilde{F}_X$, and vice versa. Furthermore, when $a \to \infty$, we will find that $\widetilde{F}_X \to \widehat{F}_h$. Oppositely, when $a$ really close to $0$, the effect of $\widehat{F}_h$ is almost vanished. However, different with bandwidth $h$, the number $a$ is purely our choice and does not depend on the sample size $n$. Letting $a$ too close to $0$ is not wise, since it acts as a denominator in the argument of function $W$.*

Now, for the bias of our proposed estimator, we have the theorem below. As expected, its leading term in this formulas is actually the same as the explicit form of $O(h^4)$ in theorem 3.1.1 (see appendices).

**Theorem 3.1.4.** *Under the assumptions B1-B4, the bias of $\widetilde{F}_X(x)$ is given by*

$$Bias[\widetilde{F}_X(x)] = h^4 a^2 \frac{a_2^2(x) - 2a_4(x)F_X(x)}{2F_X(x)} + o(h^4) + O\left(\frac{1}{n}\right), \tag{3.3}$$

*where*

$$a_2(x) = \frac{f_X'(x)}{2} \int_{-\infty}^{\infty} w^2 K(w) \mathrm{d}w \quad \text{and} \quad a_4(x) = \frac{f_X'''(x)}{24} \int_{-\infty}^{\infty} w^4 K(w) \mathrm{d}w.$$

**Remark 3.1.5.** *The factor $\frac{[f_X'(x)]^2}{F_X(x)}$ gives us some uncertain feelings that this bias may be unbounded in some points of real line, especially when $x \to -\infty$. However, even though we did not state it in the theorem, the assumption B5 ensures us that the bias is valid and bounded a.s. on the real line. For a brief example, when the true distribution is the standard normal distribution with $f_X = \phi$, we have, by simple L'Hôpital's rule,*

$$\lim_{x \to -\infty} \frac{[f_X'(x)]^2}{F_X(x)} = \lim_{x \to -\infty} \frac{x^2 \phi^2(x)}{\Phi(x)} = \lim_{x \to -\infty} \frac{2x(1-x^2)\phi^2(x)}{\phi(x)} = 0.$$

**Remark 3.1.6.** *As we expected before, the bias is in the order of $h^4$. This order is same as if we use forth order kernel function for the standard kernel distribution function estimator. However, since in some points in real line we may find negative estimates if we use those kind of kernel function, our proposed estimator is more appealing.*

Next, we discuss about the property of variance. Interestingly enough, there is no differences between the variance of our proposed estimator and the variance of the standard kernel distribution function estimator, in the sense of convergence order, as stated in the theorem below.

**Theorem 3.1.7.** *Under the assumptions B1-B4, the variance of $\widetilde{F}_X(x)$ is*

$$Var[\widetilde{F}_X(x)] = Var[F_n(x)] - \frac{h}{n}\left[\frac{2(a^4+1)}{(a^2-1)^2} r_1 + r_2\right] f_X(x) + o\left(\frac{h}{n}\right), \tag{3.4}$$

*where*

$$r_2 = \int_{-\infty}^{\infty} y \left[K(y)W\left(\frac{y}{a}\right) + \frac{1}{a}W(y)K\left(\frac{y}{a}\right)\right] \mathrm{d}y.$$

**Remark 3.1.8.** *Actually in many cases, we usually omit the $\frac{h}{n}$ term and just denote it as $O\left(\frac{h}{n}\right)$. However, since the dominant term of the variance of the standard kernel distribution function estimator and our proposed method are same, we need the second order term to compare them. It is easy to show that $\frac{a^2}{(a^2-1)^2} r_2 \geq 0$ and $\frac{a^4+1}{(a^2-1)^2} \geq 1$ when $a < 1$ (which is suggested). Hence, up to some constants, the variance of our proposed estimator is smaller than the standard kernel distribution function estimator's variance.*

As we can see, since both of the bias and the variance of our proposed estimator are smaller, then we can conclude that the MISE of our proposed estimator is smaller than the MISE of the standard kernel distribution function estimator.

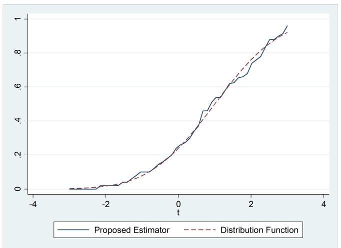
## 3.2 Simulations

In this section, we present the results of our simulation study to support the theoretical discussion. We generated random samples from the standard normal distribution, normal distribution with mean 1 and variance 2, Laplace Distribution with mean 0 and scale parameter 1, and Laplace Distribution with mean 1 and scale parameter 2. The size of each sample is 50, and we did 1000 repetitions for each case. Cross-validation is our method of choice for determining the bandwidths, and the kernel function we used is the Gaussian Kernel function. We calculated AISE (Average Integrated Squared Error) as an estimator for MISE, and compared the standard kernel distribution function estimator $\widehat{F}_h(x)$ and our proposed estimator $\widetilde{F}_X(x)$ with several number of $a$'s. The result can be seen in the Table 3.1.

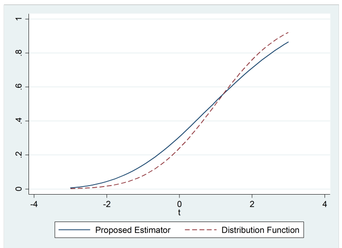
Table 3.1: AISE of Standard and Proposed Method

| Estimators | $N(0,1)$ | $N(1,2)$ | $Lap(0,1)$ | $Lap(1,2)$ |
|---|---|---|---|---|
| standard | 0.06523 | 0.07502 | 0.09098 | 0.08511 |
| a=0.01 | 0.03106 | 0.034894 | 0.03043 | 0.04096 |
| a=0.1 | 0.03127 | 0.035002 | 0.03066 | 0.04149 |
| a=0.25 | 0.03199 | 0.0397 | 0.04353 | 0.04488 |
| a=0.5 | 0.04837 | 0.0469 | 0.04499 | 0.04902 |
| a=0.75 | 0.04917 | 0.04761 | 0.04940 | 0.04947 |
| a=2 | 0.05415 | 0.05017 | 0.06899 | 0.06760 |
| a=3 | 0.05745 | 0.05032 | 0.0695 | 0.06826 |

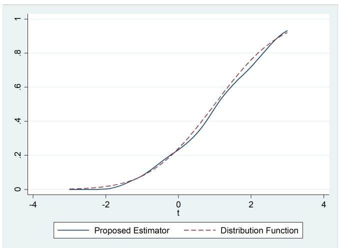
As we can see, the proposed method gives us good results, especially if we use smaller $a$. However, at some point, the differences become smaller and smaller. That is why it is unnecessary to use too small $a$.




(a) $a = 0.01$




(b) $a = 3$




(c) $a = 0.25$

Figure 3.1: $\widetilde{F}_X(x)$ with several $a$'s.

The next study we did is drawing the the graphs of our proposed method with several $a$'s ($a = 0.01$, $a = 0.25$, and $a = 3$), and comparing them with the graph of the true distribution function of the sample. Our purpose of doing this is to see the effect of different $a$'s in our estimations. We generated sample with size 50 from the normal distribution with mean 1 and variance 2. The method of choosing bandwidths and the kernel function we used are same as before.

As we can see, even though Figure 3.1(a) gives us the most accurate estimation, but the graph is not as smooth as it should be and similar to a step function. On the other hand, Figure 3.1(b) gives us the smoothest line, but it is the most inaccurate graph among all three.

In this simulation study, the best one is when $a = 0.25$ (Figure 3.1(c)). It is smooth enough and the accuracy is not so different with Figure 3.1(a). It does not mean in general $a = 0.25$ is the best one, but from this we can conclude that we have to choose $a$ wisely to get the desired result.

Even though in this paper we cannot prove mathematically that $\widetilde{F}_X \to F_n$ when $a \to 0$, but the similarity of Figure 3.1(a) with a step function provides us an intuition that when $a$ is very close to 0, the behaviour of our proposed estimator is becoming more similar to the empirical distribution function.

# Chapter 4

# Kernel-smoothed Goodness-of-fit Tests for Data on General Interval

Some methods mentioned in section 1.2 to eliminate the boundary bias problem of naive kernel density might be reasonably applicable for naive kernel distribution case as well. However, in this chapter we will try another idea, which is utilizing bijective mappings, and extend it to modify the KS and CvM tests. Here we are working under the settings mentioned that the support of the distribution is an interval $\Omega \subset \mathbb{R}$.

## 4.1 Boundary-free kernel distribution function estimator

It is obvious that if we can find an appropiate function $g$ that maps $\mathbb{R}$ to $\Omega$ bijectively, we will not put any weight outside the support. Hence, instead of using $X_1, X_2, ..., X_n$, we will apply the kernel method for $g^{-1}(X_1), g^{-1}(X_2), ...,$ $g^{-1}(X_n)$. To make sure our idea is mathematically applicable, we need to impose some conditions before moving on to our main focus. The conditions we took are:

**C1**. the kernel $K(v)$ is nonnegative, continuous, and symmetric at $v = 0$,

**C2**. the integral $\int_{-\infty}^{\infty} v^2 K(v) \mathrm{d}v$ is finite and $\int_{-\infty}^{\infty} K(v) \mathrm{d}v = 1$,

**C3**. the bandwidth $h > 0$ satisfies $h \to 0$ and $nh \to \infty$ when $n \to \infty$,

**C4**. the increasing function $g$ transforms $\mathbb{R}$ onto $\Omega$,

**C5**. the density $f_X$ and the function $g$ are twice differentiable.

The conditions C1-C3 are standard conditions for kernel method. Albeit it is sufficient for $g$ to be a bijective function, but the increasing property in C4 makes the proofs of our theorems simpler. The last condition is needed to derive the biases and the variances formula.

Under those conditions, we define the boundary-free kernel distribution function estimator as

$$\widetilde{F}_X(x) = \frac{1}{n}\sum_{i=1}^{n} W\left(\frac{g^{-1}(x) - g^{-1}(X_i)}{h}\right), \quad x \in \Omega, \tag{4.1}$$

where $h > 0$ is the bandwidth and $g$ is an appropriate bijective function. As we can see, $\widetilde{F}_X(x)$ is basically just a result of simple subsitution of $g^{-1}(x)$ and $g^{-1}(X_i)$ to the formula of $\widehat{F}_X(x)$. Though it looks simple, but the argument behind this idea is due to the change-of-variable property of distribution function, which cannot always be done to another probability-related functions. Its bias and variance are given in the following theorem.

**Theorem 4.1.1.** *Under the conditions C1-C5, the bias and the variance of $\widetilde{F}_X(x)$ are*

$$Bias[\widetilde{F}_X(x)] = \frac{h^2}{2}c_1(x)\int_{-\infty}^{\infty} v^2 K(v)\mathrm{d}v + o(h^2) \tag{4.2}$$

$$Var[\widetilde{F}_X(x)] = \frac{1}{n}F_X(x)[1 - F_X(x)] - \frac{2h}{n}g'(g^{-1}(x))f_X(x)r_1 + o\left(\frac{h}{n}\right), \tag{4.3}$$

*where*

$$c_1(x) = g''(g^{-1}(x))f_X(x) + [g'(g^{-1}(x))]^2 f'_X(x). \tag{4.4}$$

**Remark 4.1.2.** *Since $g$ is an increasing function, the variance of our proposed estimator will be smaller than $Var[\widehat{F}_X(x)]$ when $g'(g^{-1}(x)) \geq 1$. On the other hand, though it is difficult to conclude in general case, if we carefully take the mapping $g$, the bias of our proposed method is much faster to converge to $0$ than $Bias[\widehat{F}_X(x)]$. For example, when $\Omega = \mathbb{R}^+$ and we choose $g(x) = e^x$, in the boundary region when $x \to 0$ the bias will converge to $0$ faster and $Var[\widetilde{F}_X(x)] < Var[\widehat{F}_X(x)]$.*

Similar to most of kernel type estimators, our proposed estimator attains asymptotic normality, as stated in the following theorem.

**Theorem 4.1.3.** *Under the condition C1-C5, the limiting distribution*

$$\frac{\widetilde{F}_X(x) - F_X(x)}{\sqrt{Var[\widetilde{F}_X(x)]}} \to_D N(0,1) \tag{4.5}$$

*holds.*

Furthermore, we also establish strong consistency of the proposed method.

**Theorem 4.1.4.** *Under the condition C1-C5, the consistency*

$$\sup_{x\in\Omega} |\widetilde{F}_X(x) - F_X(x)| \rightarrow_{a.s.} 0 \tag{4.6}$$

*holds.*

Even though it is not exactly related to our main topic of goodness-of-fit tests, but it is worth to add that from $\widetilde{F}_X$ we can derive another kernel-type estimator. It is clear that the density function $f_X$ is equal to $F'_X$, then we can define a boundary-free kernel density estimator as $\widetilde{f}_X = \frac{\mathrm{d}}{\mathrm{d}x}\widetilde{F}_X$, which is

$$\widetilde{f}_X(x) = \frac{1}{nhg'(g^{-1}(x))} \sum_{i=1}^{n} K\left(\frac{g^{-1}(x) - g^{-1}(X_i)}{h}\right), \quad x \in \Omega. \tag{4.7}$$

As $\widetilde{F}_X$ eliminates boundary bias problem, this new estimator $\widetilde{f}_X$ does the same thing and can be a good competitor for another boundary bias reduction kernel density estimators. The bias and the variance of its are as follow.

**Theorem 4.1.5.** *Under the condition C1-C5, also if $g'''$ exists and $f''_X$ is continuous, then the bias and the variance of $\widetilde{f}_X(x)$ are*

$$Bias[\widetilde{f}_X(x)] = \frac{h^2 c_2(x)}{2g'(g^{-1}(x))} \int_{-\infty}^{\infty} v^2 K(v)\mathrm{d}v + o(h^2) \tag{4.8}$$

$$Var[\widetilde{f}_X(x)] = \frac{f_X(x)}{nhg'(g^{-1}(x))} \int_{-\infty}^{\infty} K^2(v)\mathrm{d}v + o\left(\frac{1}{nh}\right), \tag{4.9}$$

*where*

$$c_2(x) = g'''(g^{-1}(x))f_X(x) + 3g''(g^{-1}(x))g'(g^{-1}(x))f'_X(x) + [g'(g^{-1}(x))]^3 f''_X(x).$$

## 4.2 Boundary-free kernel-smoothed KS and CvM tests

As we discussed before, the problem of the standard KS and CvM statistics is in the over-sensitivity near the center of distribution, because of the lack of smoothness of the empirical distribution function. Since the area around the center of distribution has the highest probability density, most of the realizations of the sample are there. As a result, $F_n(x)$ jumps a lot in those area, and it causes some unstability of estimation especially when $n$ is small.

Conversely, though smoothing $KS_n$ and $CvM_n$ out using kernel distribution function can eliminate the oversensitivity near the center, the value of $\widehat{KS}$ and $\widehat{CvM}$ become larger than it should be when the data we are dealing with is supported on an interval smaller than the entire real line. This phenomenon is caused by the boundary problem.

Therefore, the clear solution to overcome the problems of standard and naive kernel goodness-of-fit tests together is to keep the smoothness of $\widehat{F}_X$ and to get rid of the boundary problem simulateously. One of the idea is by utilizing the boundary-free kernel distribution function estimator in section 4.1. Therefore, we propose boundary-free kernel-smoothed Kolmogorov-Smirnov statistic as

$$\widetilde{KS} = \sup_{x \in \mathbb{R}} |\widetilde{F}_X(x) - F(x)| \tag{4.10}$$

and boundary-free kernel-smoothed Cramér-von Mises statistic as

$$\widetilde{CvM} = n \int_{-\infty}^{\infty} [\widetilde{F}_X(x) - F(x)]^2 \mathrm{d}F(x), \tag{4.11}$$

where $\widetilde{F}_X$ is our proposed estimator with a suitable function $g$.

**Remark 4.2.1.** *Although the supremum and the integral are evaluated througout the entire real line, but we can just compute them over $\Omega$, as on the outside of the support we have $F_X(x) = \widetilde{F}_X(x)$.*

Although the formulas seem similar, one might expect both proposed tests are totally different with the standard KS and CvM tests. However, these two following theorems explain that the standard ones and our proposed methods turn out to be equivalent in the sense of distribution.

**Theorem 4.2.2.** *Let $F_X$ and $F$ be distribution functions on set $\Omega$. If $KS_n$ and $\widetilde{KS}$ are the standard and the proposed Kolmogorov-Smirnov statistics, respectively, then under the null hypothesis $F_X = F$,*

$$|KS_n - \widetilde{KS}| \to_p 0. \tag{4.12}$$

**Theorem 4.2.3.** *Let $F_X$ and $F$ be distribution functions on set $\Omega$. If $CvM_n$ and $\widetilde{CvM}$ are the standard and the proposed Cramér-von Mises statistics, respectively, then under the null hypothesis $F_X = F$,*

$$|CvM_n - \widetilde{CvM}| \to_p 0. \tag{4.13}$$

Those equivalencies allow us to use the same distribution tables of the standard goodness-of-fit tests for our new statistics. It means, with the same significance level $\alpha$, the critical values are same.

Table 4.1: AISE ($\times 10^5$) comparison of DF estimators

| Distributions | $\widehat{F}_X$ | $\widetilde{F}_{\log}$ | $\widetilde{F}_{\Phi^{-1}\circ\gamma}$ | $\widetilde{F}_{probit}$ | $\widetilde{F}_{logit}$ |
|---|---|---|---|---|---|
| $Gamma(2,2)$ | 2469 | 2253 | **2181** | - | - |
| $Weibull(2,2)$ | 2224 | **1003** | 1350 | - | - |
| $\log .N(0,1)$ | 1784 | 1264 | **1254** | - | - |
| $abs.N(0,1)$ | 2517 | **544** | 727 | - | - |
| $U(0,1)$ | 5074 | - | - | **246** | 248 |
| $Beta(1,3)$ | 7810 | - | - | **170** | 172 |
| $Beta(2,2)$ | 6746 | - | - | **185** | 188 |
| $Beta(3,1)$ | 7801 | - | - | **154** | 156 |

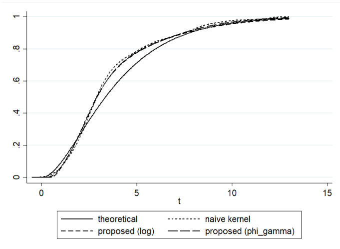
## 4.3 Numerical results

We will show the results of our numerical studies in this section. The studies consist of two parts, the simulations of the proposed estimators $\widetilde{F}_X$ and $\widetilde{f}_X$, and then the results of the new goodness-of-fit tests $\widetilde{KS}$ and $\widetilde{CvM}$.

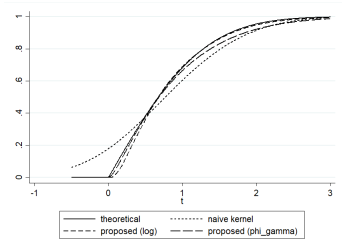
### 4.3.1 boundary-free kernel DF and PDF estimations results

For the simulation to show the performances of the new distribution function estimator, we calculated the average integrated squared error (AISE) and repeated them 1000 times for each case. We compared the naive kernel distribution function estimator $\widehat{F}_X$ and our proposed estimator $\widetilde{F}_X$. In the case of the proposed method, we chose two mappings $g^{-1}$ for each case. When $\Omega = \mathbb{R}^+$, we used the logarithm function $\log(x)$ and a composite of two functions $\Phi^{-1} \circ \gamma$, where $\gamma(x) = 1 - e^x$. However, if $\Omega = [0,1]$, we utilized probit and logit functions. With size 50, the generated samples were drawn from gamma $Gamma(2,2)$, weibull $Weibull(2,2)$, standard log-normal $\log .N(0,1)$, absolute-normal $abs.N(0,1)$, standard uniform $U(0,1)$, and beta distributions with three different sets of parameters ($Beta(1,3)$, $Beta(2,2)$, and $Beta(3,1)$). The kernel function we used here is the Gaussian Kernel and the bandwidths were chosen by cross-validation technique. We actually did the same simulation study using the Epanechnikov Kernel, but the results are quite similar. Graphs of some chosen cases are shown as well in Figure 4.1.

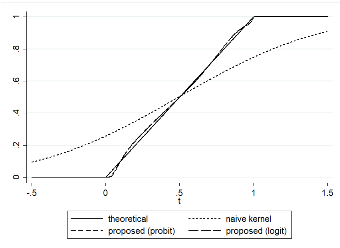
As we can see in Table 4.1, our proposed estimator outperformed the naive kernel distribution function. Though the differences are not so big in the cases of gamma, weibull, and the log-normal distributions, but the



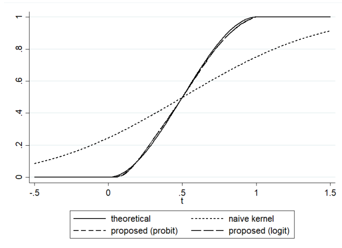

(a) $X \sim Gamma(2,2)$




(b) $X \sim abs.N(0,1)$




(c) $X \sim U(0,1)$




(d) $X \sim Beta(2,2)$

Figure 4.1: Graphs comparisons of $F_X(x)$, $\widehat{F}_X(x)$, and $\widetilde{F}_X(x)$ for several distributions, with sample size $n = 50$.

gaps are glaring in the absolute-normal case or when the support of the distributions is the unit interval. The cause of this phenomena might be seen in Figure 4.1.

Albeit the shapes of $\widetilde{F}_{\log}$ and $\widetilde{F}_{\Phi^{-1}\circ\gamma}$ are more similar to the theoretical distribution in Figure 4.1(a), but we have to admit that the shape of $\widehat{F}_X$ is not so much different with the rests. However in Figure 4.1(b), 4.1(c), and 4.1(d), it is obvious that the naive kernel distribution function is too far-off the mark, particularly in the case of $\Omega = [0,1]$. As the absolute-normal, uniform, and beta distributions have quite high probability density near the boundary point $x = 0$ (also $x = 1$ for unit interval case), the naive kernel estimator spreads this "high density information" around the boundary regions. However, since $\widehat{F}_X$ cannot detect the boundaries, it puts this "high density information" outside the support as well. This is not happening too severely in the case of Figure 4.1(a) because the probability density near $x = 0$ is fairly low. Hence, although the value of $\widehat{F}_X(x)$ might be still positive when $x \approx 0^-$, but it is not so far from 0 and vanishes quickly

**Remark 4.3.1.** *Figure 4.1(c) and 4.1(d) also gave a red-alert if we try to use the naive kernel distribution function estimator in place of empirical distribution function for goodness-of-fit tests. As the shapes of $\widehat{F}_X$ in Figure 4.1(c) and 4.1(d) resemble the normal distribution function a lot, if we test $H_0 : X \sim N(\mu, \sigma^2)$,*

Table 4.2: AISE ($\times 10^5$) comparison of density estimators

| Distributions | $\widehat{f}_X$ | $\widetilde{f}_{\log}$ | $\widetilde{f}_{\Phi^{-1}\circ\gamma}$ | $\widetilde{f}_{probit}$ | $\widetilde{f}_{logit}$ |
|---|---|---|---|---|---|
| $Gamma(2,2)$ | 925 | 744 | **624** | - | - |
| $Weibull(2,2)$ | 6616 | **3799** | 3986 | - | - |
| $\log .N(0,1)$ | 7416 | 3569 | **2638** | - | - |
| $abs.N(0,1)$ | 48005 | 34496 | **14563** | - | - |
| $U(0,1)$ | 36945 | - | - | **14235** | 21325 |
| $Beta(1,3)$ | 109991 | - | - | **18199** | 28179 |
| $Beta(2,2)$ | 52525 | - | - | **5514** | 6052 |
| $Beta(3,1)$ | 109999 | - | - | **17353** | 28935 |

*we will find the tests may not reject the null hypothesis. This shall cause the increment of type-2 error.*

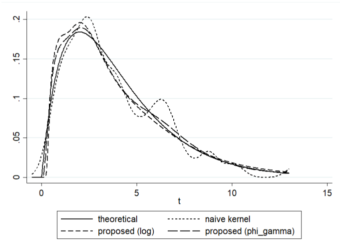
**Remark 4.3.2.** *It is worth to note that in Table 4.1, even though $\widetilde{F}_{probit}$ performed better, but its differences are too little to claim that it outperformed $\widetilde{F}_{logit}$. From here we can conclude that probit and logit functions work pretty much the same for $\widetilde{F}_X$.*

Since we introduced $\widehat{f}_X$ as a new boundary-free kernel density estimator, we also provide some illustrations of its performances in this subsection. Under the same settings as in the simulation study of the distribution function case, we can see the results of its simulation in Table 4.2 and Figure 4.2.

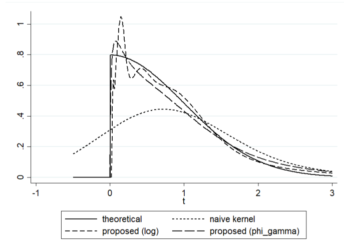
From AISE point of view, once again our proposed estimator outperformed the naive kernel one, and huge gaps happened as well when the support of the distribution is the unit interval. We may take some interests in Figure 4.2(b), 4.2(c), and 4.2(d), as the graphs of $\widehat{F}_X$ are too different with the theoretical ones, and more similar to the gaussian bell shapes instead.

### 4.3.2 boundary-free kernel-type KS and CvM tests simulations

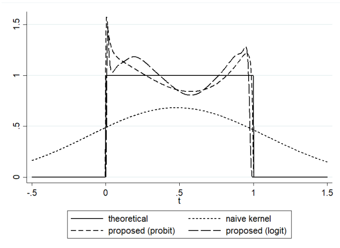
We provide the results of our simulation studies regarding the new Kolmogorov-Smirnov and Cramér-von Mises tests in this part. As a measure of comparison, we calculated the percentage of rejecting several null hypothesis when the samples were drawn from certain distributions. When the actual distribution and the null hypothesis are same, we expect the percentage should be close to $100\alpha\%$ (significance level in percent). However, if the real distribution does not match the $H_0$, we hope to see the percentage is as large



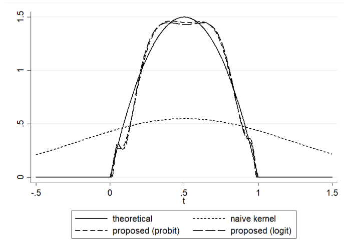

(a) $X \sim Gamma(2,2)$




(b) $X \sim abs.N(0,1)$




(c) $X \sim U(0,1)$



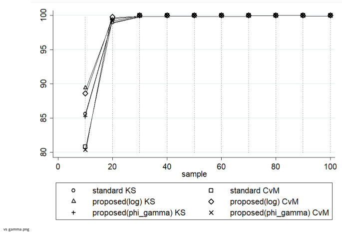

(d) $X \sim Beta(2,2)$

Figure 4.2: Graphs comparisons of $f_X(x)$, $\widehat{f}_X(x)$, and $\widetilde{f}_X(x)$ for several distributions, with sample size $n = 50$.

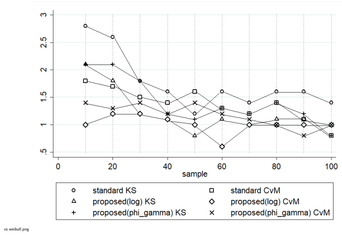
as possible. To illustrate how the behaviours of the statistics change, we generated a sequential numbers of sample sizes, started from 10 until 100, with 1000 repetitions for each case. We chose level of significance $\alpha = 0.01$, and we compared the standard KS and CvM tests with our proposed tests.

From Figure 4.3, we see that the modified KS and CvM tests outperformed the standard ones, especially the proposed KS test with logarithm as the bijective transformation. From Figure 4.3(a), 4.3(c), and 4.3(d), KS test with log function has the highest percentage of rejecting $H_0$ even when the sample sizes were still 10. However, even though the new CvM test with logarithmic function was always the second highest in the beginning, $\widetilde{CvM}_{\log}$ was also the first one that reached 100%. On the other hand, based on Figure 4.3(b) we can say all statistical tests (standard and proposed) were having similar stable behaviours, as their numbers were still in the interval $0.5\% - 2\%$. However at this time, $\widetilde{CvM}_{\log}$ performed slightly better than the others, because its numbers in general were the closest to 1%.

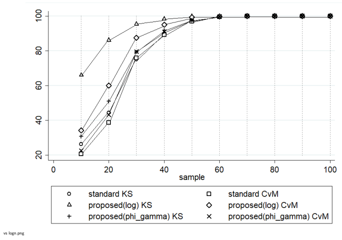
Similar things happened when we drew the samples from the standard log-normal distribution, which our proposed methods outperformed the standard ones. However this time, the modified KS test with $g^{-1} = \log$ always gave the best results. Yet, we may take some notes from Figure 4.4. First, although when $n = 10$ all the percentages were far from 1% in Figure 4.4(c), but



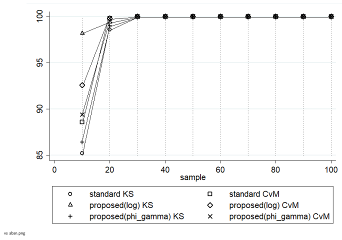

(a) $H_0 : Gamma(2, 2)$




(b) $H_0 : Weibull(2, 2)$

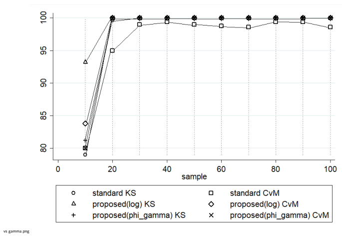



(c) $H_0 : \log .N(0, 1)$

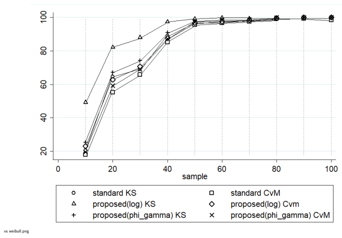



(d) $H_0 : abs.N(0, 1)$

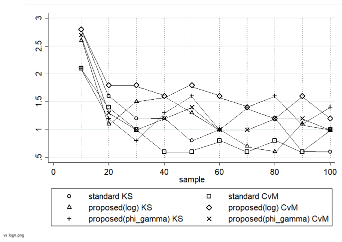
Figure 4.3: Simulated percentage (%) of rejecting null hypothesis when the samples were drawn from $Weibull(2, 2)$.



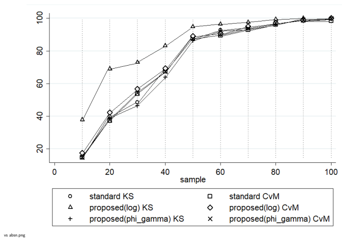

(a) $H_0 : Gamma(2, 2)$




(b) $H_0 : Weibull(2, 2)$

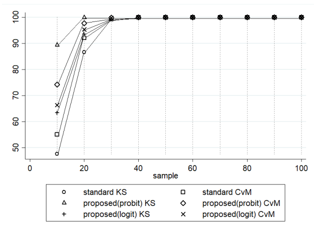



(c) $H_0 : \log .N(0, 1)$

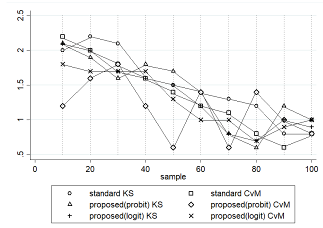



(d) $H_0 : abs.N(0, 1)$

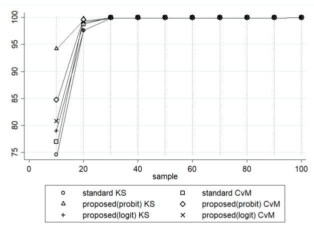
Figure 4.4: Simulated percentage (%) of rejecting null hypothesis when the samples were drawn from $\log .N(0, 1)$.



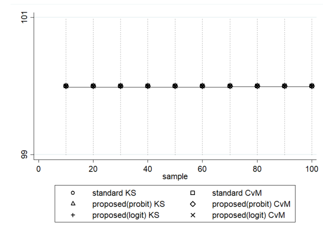

(a) $H_0 : U(0, 1)$




(b) $H_0 : Beta(1, 3)$




(c) $H_0 : Beta(2, 2)$




(d) $H_0 : Beta(3, 1)$

Figure 4.5: Simulated percentage (%) of rejecting null hypothesis when the samples were drawn from $Beta(1, 3)$.

after $n = 20$ every tests went stable inside $0.5\% - 2\%$ interval. Second, as seen in Figure 4.4(d), it seems difficult to reject $H_0 : abs.N(0, 1)$ when the actual distribution is $\log.N(0, 1)$, even $\widetilde{KS}_{\log}$ could only reach 100% rejection after $n = 80$. While, on the other hand, it was quite easy to reject $H_0 : Gamma(2, 2)$ as most of the tests already reached 100% rejection when $n = 20$ (similar to Figure 4.3(a)).

Something more extreme happened in Figure 4.5, as all of the tests could reach 100% rejection rate since $n = 30$, even since $n = 10$ in Figure 4.5(d). Though seems strange, the cause of this phenomenon is obvious. The shape of the distribution function of $Beta(1, 3)$ is so different with other three distributions in this study, especially with $Beta(3, 1)$. Hence, even with small sample size, the tests could reject the false null hypothesis. However, we still are able to claim that our proposed tests worked better than the standard goodness-of-fit tests, because before all the tests reached 100% point, the standard KS and CvM tests had the lowest percentages.

From this numerical studies, we can conclude that both the standard and the proposed KS and CvM tests will give the same result when the sample size is large. However, if the sample size is small, our proposed methods will give better and more reliable results.

# Chapter 5

# Kernel-type Mean Residual Life Function Estimators for Data on General Interval

In this chapter we are going to remove the boundary effects on the naive kernel MRL function estimator by utilizing transformations that map $\Omega$ to $\mathbb{R}$ bijectively (similar to chapter 4). However, even though the idea is easy to understand, we cannot just substitute $t$ with $g^{-1}(t)$ and $X_i$ with $Y_i$ in the formula of $\widehat{m}_X(t)$, due to avoiding nonintegrability. We need to modify the naive kernel MRL function estimator before substituting $g^{-1}(t)$ and $Y_i$ in order to preserve the integrability and to ensure that the new formulas are good estimators of the mean excess loss function.

Before moving on to our main focus, we need to impose some conditions:

**D1**. The kernel $K(x)$ is a continuous nonnegative function and symmetric at $x = 0$ with $\int_{-\infty}^{\infty} K(x)\mathrm{d}x = 1$

**D2**. The bandwidth $h > 0$ satisfies $h \to 0$ and $nh \to \infty$ when $n \to \infty$

**D3**. The function $g : \mathbb{R} \to \Omega$ is continuous and strictly increasing

**D4**. The density $f_X$ and the function $g$ are continuously differentiable at least twice

**D5**. The integrals $\int_{-\infty}^{\infty} g'(ux)K(x)\mathrm{d}x$ and $\int_{-\infty}^{\infty} g'(ux)V(x)\mathrm{d}x$ are finite for all $u$ in an $\varepsilon$-neighbourhood of the origin

**D6**. The expectations $E(X)$, $E(X^2)$, and $E(X^3)$ exist.

The first and the second conditions are standard assumptions for kernel methods, and D3 is needed for the bijectivity and the simplicity of the transformation. Since we will use some expansions of the survival and the cumulative survival functions, D4 is important to ensure the validity of our proofs. The last two conditions are necessary to make sure we can derive the bias and the variance formulas.

## 5.1 Estimators of the survival function and the cumulative survival function

Before jumping into the estimation of the mean residual life function, we will first discuss on the estimations of each component, which are the survival function $S_X(t)$ and the cumulative survival function $\mathbb{S}_X(t)$. In this thesis, we proposed two sets of estimators using the idea of transformation. Based on those two sets of estimators, we will propose two estimators of the MRL function in section 5.2.

The first idea came from equation (4.7), which is a boundary-free transformed kernel density estimator. Then, by doing simple subtitution technique on $\int_t^{\omega''} \widetilde{f}_X(x)\mathrm{d}x$, the first proposed survival function estimator is

$$\widetilde{S}_{X,1}(t) = \frac{1}{n}\sum_{i=1}^{n} V_{1,h}(g^{-1}(t), g^{-1}(X_i)), \quad t \in \Omega, \tag{5.1}$$

where

$$V_{1,h}(x,y) = \frac{1}{h}\int_x^\infty K\left(\frac{z-y}{h}\right)\mathrm{d}z. \tag{5.2}$$

Using the same approach, we define the first proposed cumulative survival function estimator as

$$\widetilde{\mathbb{S}}_{X,1}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{V}_{1,h}(g^{-1}(t), g^{-1}(X_i)), \quad t \in \Omega, \tag{5.3}$$

where

$$\mathbb{V}_{1,h}(x,y) = \int_x^\infty g'(z) V\left(\frac{z-y}{h}\right)\mathrm{d}z. \tag{5.4}$$

Their biases and variances are given in the following theorem.

**Theorem 5.1.1.** *Under the condition D1-D6, the biases and the variances of $\widetilde{S}_{X,1}(t)$ and $\widetilde{\mathbb{S}}_{X,1}(t)$ are*

$$Bias[\widetilde{S}_{X,1}(t)] = -\frac{h^2}{2}b_1(t)\int_{-\infty}^{\infty} y^2K(y)\mathrm{d}y + o(h^2) \tag{5.5}$$

$$Var[\widetilde{S}_{X,1}(t)] = \frac{1}{n}S_X(t)F_X(t) - \frac{h}{n}g'(g^{-1}(t))f_X(t)\int_{-\infty}^{\infty} V(y)W(y)\mathrm{d}y + o\left(\frac{h}{n}\right)$$

*and*

$$Bias[\widetilde{\mathbb{S}}_{X,1}(t)] = \frac{h^2}{2}b_2(t)\int_{-\infty}^{\infty} y^2K(y)\mathrm{d}y + o(h^2) \tag{5.6}$$

$$Var[\widetilde{\mathbb{S}}_{X,1}(t)] = \frac{1}{n}[2\bar{\mathbb{S}}_X(t) - \mathbb{S}_X^2(t)] + o\left(\frac{h}{n}\right), \tag{5.7}$$

*where*

$$b_1(t) = g''(g^{-1}(t))f_X(t) + [g'(g^{-1}(t))]^2 f'_X(t) \tag{5.8}$$

$$b_2(t) = [g'(g^{-1}(t))]^2 f_X(t) + \int_{g^{-1}(t)}^{\infty} g''(x)g'(x)f_X(g(x))\mathrm{d}x. \tag{5.9}$$

*Furthermore, the covariance of them is*

$$Cov[\widetilde{\mathbb{S}}_{X,1}(t), \widetilde{S}_{X,1}(t)] = \frac{1}{n}\mathbb{S}_X(t)F_X(t) + o\left(\frac{h}{n}\right). \tag{5.10}$$

**Remark 5.1.2.** *Because $\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{\mathbb{S}}_{X,1}(t) = -\widetilde{S}_{X,1}(t)$, it means that our first set of estimators preserves the relationship between the theoretical $\mathbb{S}_X(t)$ and $S_X(t)$.*

We have utilized the relationship among density, survival, and cumulative survival functions to construct the first set of estimators, now we are going to use another maneuver to build our second set of estimators. The second proposed survival function estimator is defined as

$$\widetilde{S}_{X,2}(t) = \frac{1}{n}\sum_{i=1}^{n} V_{2,h}(g^{-1}(t), g^{-1}(X_i)), \quad t \in \Omega, \tag{5.11}$$

where

$$V_{2,h}(x, y) = V\left(\frac{x-z}{h}\right). \tag{5.12}$$

As we can see, $\widetilde{S}_{X,2}(t)$ is basically just a result of a simple subsitution of $g^{-1}(t)$ and $g^{-1}(X_i)$ to the formula of $\widehat{S}_X(t)$. This can be done due to the change-of-variable property of the survival function (for a brief explanation of the change-of-variable property, see lemma .0.2). Though it is a bit trickier, the change-of-variable property of the cumulative survival function leads us to the construction of our second proposed cumulative survival function estimator, which is

$$\widetilde{\mathbb{S}}_{X,2}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{V}_{2,h}(g^{-1}(t), g^{-1}(X_i)), \quad t \in \Omega, \tag{5.13}$$

where

$$\mathbb{V}_{2,h}(x,y) = \int_{-\infty}^{y} g'(z) V\left(\frac{x-z}{h}\right) \mathrm{d}z. \tag{5.14}$$

In the above formula, multiplying $V$ with $g'$ is necessary to make sure that $\widetilde{\mathbb{S}}_{X,2}(t)$ is an estimator of $\mathbb{S}_X(t)$ (see equation (35)). Now, with $\widetilde{\mathbb{S}}_{X,2}(t)$ and $\widetilde{S}_{X,2}(t)$, their biases and variances are as follows.

**Theorem 5.1.3.** *Under the condition D1-D6, the biases and the variances of $\widetilde{S}_{X,2}(t)$ and $\widetilde{\mathbb{S}}_{X,2}(t)$ are*

$$Bias[\widetilde{S}_{X,2}(t)] = -\frac{h^2}{2} b_1(t) \int_{-\infty}^{\infty} y^2 K(y) \mathrm{d}y + o(h^2) \tag{5.15}$$

$$Var[\widetilde{S}_{X,2}(t)] = \frac{1}{n} S_X(t) F_X(t) - \frac{h}{n} g'(g^{-1}(t)) f_X(t) \int_{-\infty}^{\infty} V(y) W(y) \mathrm{d}y + o\left(\frac{h}{n}\right)$$

*and*

$$Bias[\widetilde{\mathbb{S}}_{X,2}(t)] = \frac{h^2}{2} b_3(t) \int_{-\infty}^{\infty} y^2 K(y) \mathrm{d}y + o(h^2) \tag{5.16}$$

$$Var[\widetilde{\mathbb{S}}_{X,2}(t)] = \frac{1}{n} [2\bar{\mathbb{S}}_X(t) - \mathbb{S}_X^2(t)] + o\left(\frac{h}{n}\right), \tag{5.17}$$

*where*

$$b_3(t) = [g'(g^{-1}(t))]^2 f_X(t) - g''(g^{-1}(t)) S_X(t). \tag{5.18}$$

*Furthermore, the covariance of them is*

$$Cov[\widetilde{\mathbb{S}}_{X,2}(t), \widetilde{S}_{X,2}(t)] = \frac{1}{n} \mathbb{S}_X(t) F_X(t) + o\left(\frac{h}{n}\right). \tag{5.19}$$

**Remark 5.1.4.** *As we can see in theorem 5.1.1 and theorem 5.1.3, a lot of similarities are possessed by both sets of estimators. For example, both of them have the same covariances, which means the statistical relationship between $\widetilde{S}_{X,2}(t)$ and $\widetilde{\mathbb{S}}_{X,2}(t)$ is same to the one of $\widetilde{S}_{X,1}(t)$ and $\widetilde{\mathbb{S}}_{X,1}(t)$.*

**Remark 5.1.5.** *We can prove that both $\widetilde{S}_{X,1}(\omega')$ and $\widetilde{S}_{X,2}(\omega')$ are always equal to $1$, and it is obvious that both $\widetilde{S}_{X,1}(\omega'')$ and $\widetilde{S}_{X,2}(\omega'')$ are $0$. Hence, it is clear that their variances are $0$ when $t$ approaches the boundaries. This is one of the reasons our proposed methods outperform the naive kernel estimator.*

## 5.2 Estimators of the mean residual life function

In this section, we will discuss the estimation of the mean residual life function. As we already have defined the survival function and the cumulative survival function estimators, we just need to plug them into the MRL function formula. Hence, our proposed estimators of the mean excess loss function are

$$\widetilde{m}_{X,1}(t) = \frac{\widetilde{\mathbb{S}}_{X,1}(t)}{\widetilde{S}_{X,1}(t)} = \frac{h\sum_{i=1}^{n}\int_{g^{-1}(t)}^{\infty} g'(z)V\left(\frac{z-g^{-1}(X_i)}{h}\right)\mathrm{d}z}{\sum_{i=1}^{n}\int_{g^{-1}(t)}^{\infty} K\left(\frac{z-g^{-1}(X_i)}{h}\right)\mathrm{d}z}, \quad t \in \Omega \tag{5.20}$$

and

$$\widetilde{m}_{X,2}(t) = \frac{\widetilde{\mathbb{S}}_{X,2}(t)}{\widetilde{S}_{X,2}(t)} = \frac{\sum_{i=1}^{n}\int_{-\infty}^{g^{-1}(X_i)} g'(z)V\left(\frac{g^{-1}(t)-z}{h}\right)\mathrm{d}z}{\sum_{i=1}^{n} V\left(\frac{g^{-1}(t)-g^{-1}(X_i)}{h}\right)}, \quad t \in \Omega. \tag{5.21}$$

At first glance, $\widetilde{m}_{X,1}(t)$ seems more representative to the theoretical $m_X(t)$, since the mathematical relationship between $\widetilde{S}_{X,1}(t)$ and $\widetilde{\mathbb{S}}_{X,1}(t)$ are same as the relationship between the numerator and the denumerator of $m_X(t)$, as stated in Remark 5.1.2. This is not a major problem for $\widetilde{m}_{X,2}(t)$, as we stated in Remark 5.1.4 that the relationship between $\widetilde{S}_{X,2}(t)$ and $\widetilde{\mathbb{S}}_{X,2}(t)$ is statistically same to the relationship between $\widetilde{S}_{X,1}(t)$ and $\widetilde{\mathbb{S}}_{X,1}(t)$. However, when a statistician wants to keep the mathematical relationship between the survival and the cumulative survival functions in their estimates, it is suggested to use $\widetilde{m}_{X,1}(t)$ instead.

**Theorem 5.2.1.** *Under the condition D1-D6, the biases and the variances of $\widetilde{m}_{X,i}(t)$, $i = 1, 2$, are*

$$Bias[\widetilde{m}_{X,1}(t)] = \frac{h^2}{2S_X(t)}[b_2(t) + m_X(t)b_1(t)]\int_{-\infty}^{\infty} y^2K(y)\mathrm{d}y + o(h^2) \tag{5.22}$$

$$Bias[\widetilde{m}_{X,2}(t)] = \frac{h^2}{2S_X(t)}[b_3(t) + m_X(t)b_1(t)]\int_{-\infty}^{\infty} y^2K(y)\mathrm{d}y + o(h^2) \tag{5.23}$$

$$Var[\widetilde{m}_{X,i}(t)] = \frac{1}{n}\frac{b_4(t)}{S_X^2(t)} - \frac{h}{n}\frac{b_5(t)}{S_X^2(t)}\int_{-\infty}^{\infty} V(y)W(y)\mathrm{d}y + o\left(\frac{h}{n}\right), \tag{5.24}$$

*where*

$$b_4(t) = 2\bar{\mathbb{S}}_X(t) - S_X(t)m_X^2(t) \quad \text{and} \quad b_5(t) = g'(g^{-1}(t))f_X(t)m_X^2(t). \tag{5.25}$$

Similar to most of kernel type estimators, our proposed estimators attain asymptotic normality, as stated in theorem 5.2.2.

**Theorem 5.2.2.** *Under the condition D1-D6, the limiting distribution*

$$\frac{\widetilde{m}_{X,i}(t) - m_X(t)}{\sqrt{Var[\widetilde{m}_{X,i}(t)]}} \to_D N(0,1) \tag{5.26}$$

*holds for* $i = 1, 2$.

Furthermore, we also establish strong consistency of the proposed estimators in the form of the following theorem.

**Theorem 5.2.3.** *Under the condition D1-D6, the consistency*

$$\sup_{t \in \Omega} |\widetilde{m}_{X,i}(t) - m_X(t)| \to_{a.s.} 0 \tag{5.27}$$

*holds for* $i = 1, 2$.

The last property that we would like to discuss is the behaviour of our proposed estimators when $t$ is in the boundary regions. As stated in section 1.4, we want our estimators to preserve the behaviour of the theoretical MRL function, specifically the property of $m_X(\omega') = E(X) - \omega'$. If we can prove this, then not only will our proposed methods be free of boundary problems, but also superior in the sense of them preserving the key property of the MRL function.

**Theorem 5.2.4.** *Let* $\widetilde{m}_{X,1}(t)$ *and* $\widetilde{m}_{X,2}(t)$ *be the transformed kernel mean residual life function estimators. Then*

$$\widetilde{m}_{X,1}(\omega') + \omega' = \bar{X} + O_p(h^2) \tag{5.28}$$

*and*

$$\widetilde{m}_{X,2}(\omega') + \omega' = \bar{X}. \tag{5.29}$$

**Remark 5.2.5.** *Please note that, although for convenience it is written as* $\widetilde{m}_{X,i}(\omega')$ *(or* $\widetilde{m}_{X,i}(\omega'')$*), but we actually mean it as* $\lim_{t \to \omega'^+} \widetilde{m}_{X,i}(t)$ *(or* $\lim_{t \to \omega''^-} \widetilde{m}_{X,i}(t)$*), since* $g^{-1}(\omega')$ *(or* $g^{-1}(\omega'')$*) might be undefined.*

**Remark 5.2.6.** *From equation (5.29), because*

$$E[\widetilde{m}_{X,2}(\omega')] = E(X) - \omega' = m_X(\omega').$$

*we can say that* $\widetilde{m}_{X,2}(\omega')$ *is unbiased. In other words, its bias is exactly* $0$. *On the other hand, even though* $\widetilde{m}_{X,1}(\omega')$ *is not exactly the same as* $\bar{X} - \omega'$, *we can at least say they are close enough, and the rate of* $h^2$ *error is relatively small. However, from this we may take a conclusion that* $\widetilde{m}_{X,2}(t)$ *is superior than* $\widetilde{m}_{X,1}(t)$ *in the aspect of preserving behaviour of the theoretical MRL function near the boundary.*

## 5.3 Numerical studies

In this section, we show the results of our numerical studies. The studies are divided into two parts, the simulations and the real data analysis.

### 5.3.1 simulation results

In this study, we calculated the average integrated squared error (AISE) and the average squared error (ASE) with $n = 50$ as the sample size, and repeated them 10000 times for each case. We compared four estimators: empirical $m_n(t)$; naive kernel $\widehat{m}_X(t)$; and our two proposed estimators $\widetilde{m}_{X,1}(t)$ and $\widetilde{m}_{X,2}(t)$. The distributions which we generated are standard uniform $U(0,1)$, beta $Beta(3,2)$, gamma $Gamma(2,3)$, Weibull $Weibull(3,2)$, and absolute-normal $abs.N(0,1)$ distributions. For $U(0,1)$ and $Beta(3,2)$, we took $g = \Phi$, the standard normal distribution function; and we chose $g^{-1} = \log$ for the rests. The kernel function we used here is the Epanechnikov Kernel and the bandwidths were chosen by cross-validation technique. We actually did the same simulation study using Gaussian Kernel, but the results are quite similar. That being the case, we do not show those results in this thesis.

(Table 5.1) compares the AISE in order to illustrate the general measure of error among the estimators. (Table 5.2) compares the ASE of each estimate when $t = 0.001$, as a representation of the error when $t$ is in the boundary region. For (Table 5.3), the ASE at $t = E(X)$ represents the error when the point of evaluation is moderate. The last table represents the error of the estimators when $t$ is large enough.

As we can see in the tables, our proposed estimators gave the best results for all cases. This is particularly true for our second proposed estimator in most cases. Though our first proposed estimator's performances are not as good as the second one, it is still fairly comparable because the differences are not huge. Furthermore, the first proposed estimator is better than the empirical and the naive kernel estimators in most cases.

We may take interest in (Table 5.2) as the empirical MRL function gave similar results as our second proposed estimator did. However, this is reasonable due to the fact that $m_n(0) = \bar{X}$, same as $\widetilde{m}_{X,2}(0)$ according to theorem 5.2.4. In (Table 5.3), even though our second estimator still outperformed the others, the margins of difference with the other estimators are not big. This can be explained as $t = E(X)$ has high density, neither it has boundary problems nor lack of data as in the tail. However, (Table 5.4) showed another story. As the tail of distribution has lesser density of data, the empirical and naive kernel estimators dropped to 0 quickly.

As further illustrations, we also provide some graphs to compare our

Table 5.1: Average integrated squared error comparison

| Distributions | Empirical | Naive | Proposed 1 | Proposed 2 |
|---|---|---|---|---|
| $U(0,1)$ | $12.62201$ | $12.92260$ | $6.43667$ | $\mathbf{6.06429}$ |
| $Beta(3,2)$ | $31.49324$ | $35.09813$ | $14.27382$ | $\mathbf{10.09131}$ |
| $Gamma(2,3)$ | $58.86433$ | $66.06180$ | $29.01258$ | $\mathbf{24.51671}$ |
| $Weibull(3,2)$ | $0.14887$ | $0.27073$ | $0.11483$ | $\mathbf{0.04073}$ |
| $abs.N(0,1)$ | $0.13255$ | $0.09379$ | $0.07700$ | $\mathbf{0.04943}$ |

Table 5.2: Average squared error comparison when $t = 0.001$

| Distributions | Empirical | Naive | Proposed 1 | Proposed 2 |
|---|---|---|---|---|
| $U(0,1)$ | $\mathbf{0.08082}$ | $0.27639$ | $0.08728$ | $\mathbf{0.08082}$ |
| $Beta(3,2)$ | $0.23834$ | $0.77734$ | $0.28031$ | $0.23833$ |
| $Gamma(2,3)$ | $\mathbf{0.35739}$ | $0.82174$ | $0.37886$ | $\mathbf{0.35739}$ |
| $Weibull(3,2)$ | $\mathbf{0.00741}$ | $0.05175$ | $0.00958$ | $\mathbf{0.00741}$ |
| $abs.N(0,1)$ | $\mathbf{0.00728}$ | $0.07433$ | $0.00803$ | $\mathbf{0.00728}$ |

Table 5.3: Average squared error comparison when $t = E(X)$

| Distributions | Empirical | Naive | Proposed 1 | Proposed 2 |
|---|---|---|---|---|
| $U(0,1)$ | $0.20656$ | $0.26846$ | $0.20237$ | $\mathbf{0.18102}$ |
| $Beta(3,2)$ | $0.42396$ | $0.61622$ | $0.42069$ | $\mathbf{0.31329}$ |
| $Gamma(2,3)$ | $0.76834$ | $1.01913$ | $0.73909$ | $\mathbf{0.56705}$ |
| $Weibull(3,2)$ | $0.00806$ | $0.02740$ | $0.03183$ | $\mathbf{0.00665}$ |
| $abs.N(0,1)$ | $0.01176$ | $0.02295$ | $0.01127$ | $\mathbf{0.00566}$ |

Table 5.4: Average squared error comparison when $t = E(X) + \sigma$

| Distributions | Empirical | Naive | Proposed 1 | Proposed 2 |
|---|---|---|---|---|
| $U(0,1)$ | $3.09158$ | $3.06731$ | $\mathbf{1.33397}$ | $1.34459$ |
| $Beta(3,2)$ | $5.89501$ | $6.03789$ | $1.85105$ | $\mathbf{1.51738}$ |
| $Gamma(2,3)$ | $10.61608$ | $11.18569$ | $4.26919$ | $\mathbf{4.00378}$ |
| $Weibull(3,2)$ | $0.13199$ | $0.12829$ | $0.06907$ | $\mathbf{0.01307}$ |
| $abs.N(0,1)$ | $0.08860$ | $0.08057$ | $0.03929$ | $\mathbf{0.02293}$ |

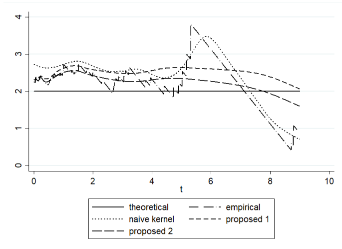
proposed estimators' performances with the other estimators. (Figure 5.1) is about the graphs comparison of the empirical, the naive kernel, and our two proposed estimators. By (Figure 5.2), we compare the point-wise simulated bias of the same estimators. From those, we can say that our proposed estimators outperformed the empirical and the naive kernel estimators.

There are three things that we want to emphasize from these figures. First, instead of resembling the theoretical shape, the graphs of the naive kernel estimator are more like a smoothed version of the graphs of the empirical estimator, especially in (Figure 5.1(a)) and (Figure 5.1(b)). This is somewhat interesting, as even though lack of smoothness, empirical type estimators (e.g. empirical distribution function) usually quite resemble the shape of the theoretical ones. However, in this MRL function case, the empirical MRL function cannot be used as a reference, because its shape is too unstable and too different to the theoretical shape (see (Figure 5.1(a)), (Figure 5.1(b)), and (Figure 5.1(d))). Same goes for the naive kernel MRL function estimator. Even though (Figure 5.1(c)) and (Figure 5.1(d)) showed the naive kernel estimator has nice graphs, it performed fairly poorly in (Figure 5.1(a)) and (Figure 5.1(b)). On the other hand, the graphs of our proposed estimators resemble the theoretical ones. The difference is quite striking in (Figure 5.1(a)), where the empirical and naive kernel estimators are jumpy, but the proposed estimators gave almost straight-line graphs.

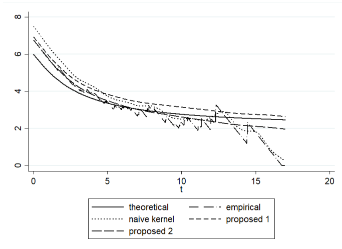
The second thing we want to emphasize is, from all figures we can see that the boundary bias problems affect naive kernel estimator severely, as (Figure 5.2) shows the simulated bias values of $\widehat{m}_X(t)$ near $t = 0$ are the farthest from 0. We can also conclude that the empirical MRL function does not suffer from the boundary bias problems, as its bias is almost 0 near $t = 0$. However, as $t$ goes larger, the bias drops to negative value quickly, especially in (Figure 5.2(a)). In contrast, our estimators, especially the second one, gave almost straight line at 0-ordinate in (Figure 5.2(b)), which means its simulated bias is almost always 0. And at last, we can conclude that though all of the graphs of the estimators presented here will fade to 0 when $t$ is large enough, our proposed estimators are more stable and fading to 0 much slower than the other two estimators.

### 5.3.2 real data analysis

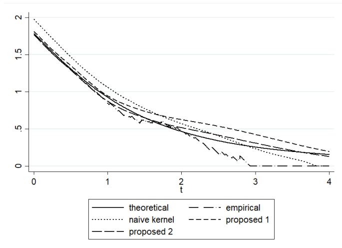
In this analysis, we used the UIS Drug Treatment Study Data from Hosmer and Lemeshow (1998) to show the performances our proposed methods for real data. The data set records the result of an experiment about how long someone who got drug treatment to relapse (reuse) the drug again. The variable we used in the calculation is the "time" variable, which represents



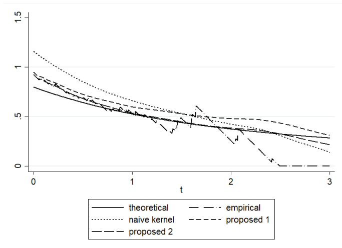

(a) $X \sim exp\left(\frac{1}{2}\right)$




(b) $X \sim Gamma(2, 3)$

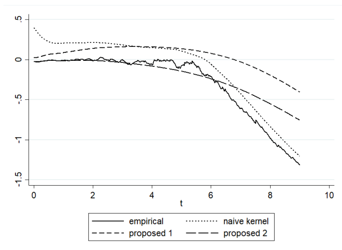



(c) $X \sim Weibull(3, 2)$

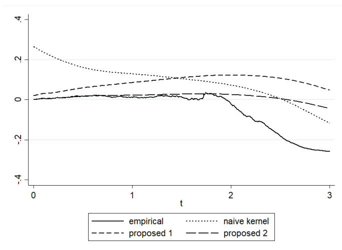



(d) $X \sim abs.N(0, 1)$

Figure 5.1: Graphs comparisons of $m_X(t)$, $m_n(t)$, $\widehat{m}_X(t)$, $\widetilde{m}_{X,1}(t)$, and $\widetilde{m}_{X,2}(t)$ for several distributions, with sample size $n = 50$.




(a) $X \sim exp\left(\frac{1}{2}\right)$




(b) $X \sim abs.N(0, 1)$

Figure 5.2: Simulated bias comparisons of $m_n(t)$, $\widehat{m}_X(t)$, $\widetilde{m}_{X,1}(t)$, and $\widetilde{m}_{X,2}(t)$ for several distributions, with sample size $n = 50$ and $500$ repetitions.

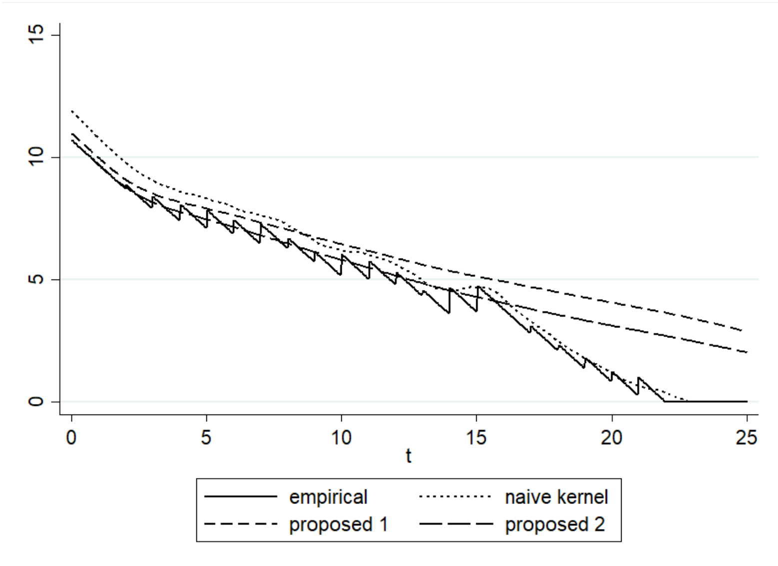



Figure 5.3: Comparison of $m_n(t)$, $\widehat{m}_X(t)$, $\widetilde{m}_{X,1}(t)$, and $\widetilde{m}_{X,2}(t)$ for UIS data

the number of days after the admission to drug treatment until drug relapse.

(Figure 5.3) shows that, once again, the naive kernel estimator is just a smoothed version of the empirical MRL function. Furthermore, soon after $m_n(t)$ touches 0, $\widehat{m}_X(t)$ also reaches 0. Conversely, though our proposed estimators are decreasing as well, but the speeds are much slower than the other two.

# Acknowledgement

This thesis would not be finished without any support, suggestion, and help from many people. The author feels deeply grateful for kind guidance of Professor Yoshihiko Maesono as his supervisor. Even until the last time, Professor Maesono never gave up to give his wise support for the author. For many helps and supports he received, the author is thankful for his senior, Taku Moriyama. Professor Masuda, Professor Hirose, and Professor Tsukuda also gave a lot of suggestions, and because of that the author feels grateful.

# Bibliography

[1] Abdous, B. (1993) Note on the minimum mean integrated squared error of kernel estimates of a distribution function and its derivatives. *Communications in Statistics - Theory and Methods* Vol. 22, 603-609.

[2] Abdous, B. and Berred, A. (2005) Mean residual life estimation. *Journal of Statistical Planning and Inference* Vol. 132, 3-19.

[3] Abramson, I. S. (1982) On bandwidth variation in kernel estimates - a square root law. *Annals of Statistics* Vol. 10, 1217-1223.

[4] Antoneli, F. *et al.* (2018) A Kolmogorov-Smirnov test for the molecular clock based on bayesian ensembles of phylogenies. *PLoS ONE* Vol. 13(1), e0190826.

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

[6] Baringhaus, L. and Henze, N. (2017) Cramér-von Mises distance: Probabilistic interpretation, confidence intervals, and neighbourhood-of-model validation. *Journal of Nonparametric Statistics* Vol. 29, 167-188.

[7] Bayramoglu, I. and Ozkut, M. (2016) Mean residual life and inactivity time of a coherent system subjected to Marshall–Olkin type shocks. *Journal of Computational and Applied Mathematics* Vol. 298, 190-200.

[8] Belzunce, F. *et al.* (1996) The dilation order, the dispersion order, and orderings of residual lifes. *Statistics and Probability Letters* Vol. 33, 263-275.

[9] Bouezmarni, T. and Scaillet, O. (2005) Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* Vol. 21, 390-412.

[10] Brown, B. M. and Chen, S. X. (1999) Beta-Bernstein smoothing for regression curves with compact supports. *Scandinavian Journal of Statistics* Vol. 26, 47-59.

[11] Chan, K. C. G., Chen, Y. Q., and Di, C. Z. (2012) Proportional mean residual life model for right-censored length-biased data. *Biometrika* Vol. 99(4), 995-1000.

[12] Chaubey, Y. P. and Sen, P. K. (1999) On smooth estimation of mean residual life. *Journal of Statistical Planning and Inference* Vol. 75, 223-236.

[13] Chen, H., Döring, M., and Jensen, U. (2018) Test for model selection using Cramér-von Mises distance in a fixed design regression setting. *AStA Advances in Statistical Analysis* Vol. 102, 505-535.

[14] Chen, S. X. (2000) Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics* Vol. 52, 471-480.

[15] Cheng, M. Y., Fan, J., and Marron, J. S. (1997) On automatic boundary corrections. *Annals of Statistics* Vol. 25, 1691-1708.

[16] Cowling, A. and Hall, P. (1996) On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society B* Vol. 58, 551-563.

[17] Cox, D. R. (1962) *Renewal Theory.* London: Methuen.

[18] Csörgő, M. and Zitikis, R. (1996) Mean residual life processes. *Annals of Statistics* Vol. 24, 1717-1739.

[19] Curry, J., Dang, X., and Sang, H. (2019) A rank-based Cramér-von-Mises-type test for two samples. *Brazilian Journal of Probability and Statistics* Vol. 33(3), 425-454.

[20] Degenhardt, H. J. A. (1993) Chung-Smirnov Property for perturbed empirical distribution functions. *Statistics and Probability Letters* Vol. 16, 97-101.

[21] Ebrahimi N. (1991) On estimating change point in a mean residual life function. *Sankhya Series A* Vol. 53, 206-219.

[22] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997) *Modelling extremal events.* Berlin: Springer.

[23] Eryilmaz, S. (2012) On the mean residual life of a $k$-out-of-$n$:G system with a single cold standby component. *European Journal of Operational Research* Vol. 222, 273-277.

[24] Eryilmaz, S. (2017) Computing optimal replacement time and mean residual life in reliability shock models. *Computers and Industrial Engineering* Vol. 103, 40-45.

[25] Eryilmaz S., Coolen, F. P. A., and Coolen-Maturi, T. (2018) Mean residual life of coherent systems consisting of multiple types of dependent components. *Naval Research Logistics* Vol. 65(1), 86-97.

[26] Evans, D. L.,Drew, J. H., and Leemis, L. M. (2017) The distribution of the Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling test statistics for exponential populations with estimated parameters. *Computational Probability Applications* Vol. 247, 165-190.

[27] Falk, M. (1983) Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statistica Neerlandica* Vol. 37, 73-83.

[28] Finner, H. and Gontscharuk, V. (2018) Two-sample Kolmogorov-Smirnov-type tests revisited: Old and new tests in terms of local levels. *Annals of Statistics* Vol. 46(6A), 3014-3037.

[29] Ghasemi, A. and Hodkiewicz, M. R. (2012) Estimating mean residual life for a case study of rail wagon bearings. *IEEE Transactions on Reliability* Vol. 61(3), 719–730.

[30] Guess, F. and Proschan, F. (1998) Mean residual life. In: Krishnaiah PR, Rao CR, editors. *Handbook of statistics* Vol. 7, Amsterdam: North-Holland; 215-224.

[31] Guillamón, A., Navarro, J., and Ruiz, J. M. (1998) Nonparametric estimator for mean residual life and vitality function. *Statistical Papers* Vol. 39, 263-276.

[32] Hall, P. and Wehrly, T. E. (1991) A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association* Vol. 86, 665-672.

[33] Hill, P. D. (1985) Kernel estimation of a distribution function. *Communications in Statistics - Theory and Methods* Vol. 14, 605-620.

[34] Hille, E. (1948) *American mathematical society colloquium publications* Vol. 31. New York: American Mathematical Society; Functional analysis and semigroups.

[35] Hosmer, D. W. and Lemeshow, S. (1998) *Applied survival analysis: Regression modeling of time to event data*. New York: John Wiley and Sons.

[36] Huynh, K.T. et al. (2014) On the use of mean residual life as a condition index for condition-based maintenance decision-making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* Vol. 44(7), 877-893.

[37] Igarashi, G. and Kakizawa, Y. (2015) Bias corrections for some asymmetric kernel estimators. *Journal of Statistical Planning and Inference* Vol. 159, 37-63.

[38] Jones, M. C. (1993) Simple boundary correction for kernel density estimation. *Statistics and Computing* Vol. 3, 135-146.

[39] Jones, M. C. and Foster, P. J. (1996) A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica* Vol. 6, 1005-1013.

[40] Jones, M. C., Linton, O., and Nielsen, J. P. (1995) A simple bias reduction method for density estimation. *Biometrika* Vol. 82, 327-338.

[41] Kavlak, K.B. (2017) Reliability and mean residual life functions of coherent systems in an active redundancy. Naval Research Logistics Vol. 64(1), 19–28.

[42] Kotz, S. and Shanbhag, D. (1980) Some new approaches to probability distributions. *Advances in Applied Probability* Vol. 12, 903-921.

[43] Lejeune, M. and Sarda, P. (1992) Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis* Vol. 14, 457-471.

[44] Loeve, M (1963) *Probability Theory.* New Jersey: Van Nostrand-Reinhold.

[45] Marron, J. S. and Ruppert, D. (1994) Transformation to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society B* Vol. 56, 653-671.

[46] Müller, H. G. (1991) Smooth optimum kernel estimators near endpoints. *Biometrika* Vol. 78, 521-530.

[47] Müller, H. G. (1993) On the boundary kernel method for nonparametric curve estimation near endpoints. *Scandinavian Journal of Statistics* Vol. 20, 313-328.

[48] Müler, H. G. and Wang, J. L. (1994) Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* Vol. 50, 61-76.

[49] Murari, M. and Sujit, K. (1995) Change point estimation in non-monotonic aging models. *Annals of the Institute of Statistical Mathematics* Vol. 3, 483-491.

[50] Nadaraya, E. A. (1964) Some new estimates for distribution functions. *Theory Probability and Applications* Vol. 15, 497-500.

[51] Nanda, A. K., Bhattacharjee, S., and Alam, S. S. (2006) Properties of proportional mean residual life model. *Statistics and Probability Letters* Vol. 76, 880-890.

[52] Nanda, A. K., Bhattacharjee, S., and Balakrishnan, N. (2010) Mean residual life function, associated orderings, and properties. *IEEE Transactions on Reliability* Vol. 59(1), 55–65.

[53] Navarro, J. and Hernandez, P. J. (2008) Mean residual life functions of finite mixtures, order statistics, and coherent systems. *Metrika* Vol. 67, 277-298.

[54] Omelka, M., Gijbels, I., and Veraverbeke, N. (2009) Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Annals of Statistics* Vol. 37, 3023-3058.

[55] Parzen, E. (1962) On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* Vol. 32, 1065-1076.

[56] Poursaeed, M.H. (2010) A note on the mean past and the mean residual life of a $(n-k+1)$-out-of-$n$ system under multi monitoring. *Statistical Papers* Vol. 51, 409-419.

[57] Raqab, M.Z. and Asadi, M. (2008) On the mean residual life of records. *Journal of Statistical Planning and Inference* Vol. 138, 3660-3666.

[58] Reiss, R. D. (1981) Nonparametric estimation of smooth distribution functions. *Scandinavian Journal of Statistics* Vol. 8, 116-119.

[59] Rosenblatt, M. (1956) Remarks on some non-prametric estimates of a density function. *The Annals of Mathematical Statistics* Vol. 27, 832-837.

[60] Ruiz, J. M. and Guillamón, A. (1996) Nonparametric recursive estimator for mean residual life and vitality function under dependence conditions. *Communications in Statistics - Theory and Methods* Vol. 25, 1997-2011.

[61] Ruiz, J. M. and Navarro, J. (1994) Characterization of distributions by relationships between the failure rate and the mean residual life. *IEEE Transactions on Reliability* Vol. 43, 640-644.

[62] Ruppert, D. R. and Cline, D. B. H. (1994) Bias reduction in kernel density estimation by smoothed empirical transformations. *Annals of Statistics* Vol. 22, 185-210.

[63] Sadegh, M. K. (2008) Mean past and mean residual life functions of a parallel system with nonidentical components. *Communications in Statistics - Theory and Methods* Vol. 37(7), 1134-1145.

[64] Samiuddin, M. and El-Sayyad, G. M. (1990) On nonparametric kernel density estimates. *Biometrika* Vol. 77, 865-874.

[65] Schuster, E. F. (1985) Incorporating support constraints into nonparametric estimators of densities. *Communication in Statistics - Theory and Methods* Vol. 14, 1123-1136.

[66] Shen, Y., Tang, L. C., and Xie, M. (2009) A model for upside-down bathtub-shaped mean residual life and its properties. *IEEE Transactions on Reliability* Vol. 58(3), 425-431.

[67] Shirahata, S. and Chu, I. S. (1992) Integrated squared error of kernel type estimator of distribution function. *Annals of the Institute of Statistical Mathematics* Vol. 44, 579-591.

[68] Singh, R. S., Gasser, T., and Prasad, B. (1983) Nonparametric estimates of distribution functions. *Communications in Statistics - Theory and Methods* Vol. 12, 2095-2108.

[69] Sun, L., Song, X., and Zhang, Z. (2012) Mean residual life models with time-dependent coefficients under right censoring. *Biometrika* Vol. 99(1), 185-197.

[70] Sun, L. and Zhang, Z. (2009) A class of transformed mean residual life models with censored survival data. *Journal of the American Statistical Association* Vol. 104(486), 803-815.

[71] Sterfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics.* Wiley.

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

[73] Terrel, G. R. and Scott, D. W. (1980) On improving convergence rates for non-negative kernel density estimation. *The Annals of Statistics* Vol. 8, 1160-1163.

[74] Watson, G. S. and Leadbetter, M. R. (1964) Hazard analysis II. *Sankhya Series A* Vol. 26, 101-106.

[75] Winter, B. B. (1973) Strong uniform consistency of integrals of density estimators. *Canadian Journal of Statistics* Vol. 1, 247-253.

[76] Winter, B. B. (1979) Convergence rate of perturbed empirical distribution functions. *Journal of Applied Probability* Vol. 16, 163-173.

[77] Yamato, H. (1973) Uniform convergence of an estimator of a distribution function. *Bulletin of Mathematical Statistics* Vol. 15, 69-78.

[78] Yang, G. L. (1978) Estimation of a biometric function. *Annals of Statistics* Vol. 6, 112-116.

[79] Yukich, J. E. (1989) A note on limit theorems for pertubed empirical processes. *Stochastic Process and Applications* Vol. 33, 163-173.

[80] Zhang, S. (2010) A note on the performance of the gamma kernel estimators at the boundary. *Statistics and Probability Letters* Vol. 80, 548-557.

[81] Zhao, P. and Balakrishnan, N. (2009) Mean residual life order of convolutions of heterogeneous exponential random variables. *Journal of Multivariate Analysis* Vol. 100, 1792-1801.

[82] Zhao, S. *et al.* Evaluation of reliability function and mean residual life for degrading systems subject to condition monitoring and random failure. *IEEE Transactions on Reliability* Vol. 67(1), 13-25.

[83] Zierk, J. *et al.* (2020) Reference interval estimation from mixed distributions using truncation points and the Kolmogorov-Smirnov distance (kosmic). *Scientific Reports* Vol. 10, 1704.

[84] Zoroa, P., Ruiz, J. M., and Marlin, J. (1990) A characterization based on conditional expectations. *Communications in Statistics - Theory and Methods* Vol. 19, 3127-3135.

# Appendix

## Proof of theorem 2.1.2

First, by usual reasoning of i.i.d. random variables, we have

$$E[A_h(x)] = \int_0^\infty \frac{w^{\frac{1}{\sqrt{h}}-1} e^{-\frac{w}{x\sqrt{h}+h}}}{\Gamma\left(\frac{1}{\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{1}{\sqrt{h}}}} f_X(w)\mathrm{d}w.$$

If we define a random variable $W \sim Gamma(h^{-1/2}, x\sqrt{h}+h)$ with mean

$$\mu_W = h^{-1/2}(x\sqrt{h}+h),$$

$Var(W) = h^{-1/2}(x\sqrt{h}+h)^2$, and $E[(W-\mu_W)^3] = 2h^{-1/2}(x\sqrt{h}+h)^3$, we can see the integral as an expectation of $f_X(W)$, and we are then able to use Taylor expansion twice, first around $\mu_W$, and next around $x$. This results in

$$\begin{aligned} E[f_X(W)] &= E\left[f_X(\mu_W) + f'_X(\mu_W)(W-\mu_W) + \frac{f''_X(\mu_W)}{2}(W-\mu_W)^2 + ...\right] \\ &= f_X(x+\sqrt{h}) + \frac{1}{2}f''_X(x+\sqrt{h})\frac{1}{\sqrt{h}}(x\sqrt{h}+h)^2 + ... \\ &= f_X(x) + \left[f'_X(x) + \frac{1}{2}x^2 f''_X(x)\right]\sqrt{h} + o(\sqrt{h}). \end{aligned}$$

Hence, we have the bias is in the order of $\sqrt{h}$.

Next, we derive the formula of the variance, which is

$$Var[A_h(x)] = n^{-1}E[K^2(X_1; x, h)] + O(n^{-1}).$$

First, we take a look at the expectation part,

$$
\begin{aligned}
E[K^2(X_1;x,h)] &= \int_0^\infty \frac{v^{\frac{2}{\sqrt{h}}-2} e^{-\frac{2v}{x\sqrt{h}+h}}}{\Gamma^2\left(\frac{1}{\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{2}{\sqrt{h}}}} f_X(v)\mathrm{d}v \\
&= \frac{\Gamma\left(\frac{2}{\sqrt{h}}-1\right)\left(\frac{x\sqrt{h}+h}{2}\right)^{\frac{2}{\sqrt{h}}-1}}{\Gamma^2\left(\frac{1}{\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{2}{\sqrt{h}}}} \\
&\quad \times \int_0^\infty \frac{v^{\left(\frac{2}{\sqrt{h}}-1\right)-1} e^{-\frac{2v}{x\sqrt{h}+h}}}{\Gamma\left(\frac{2}{\sqrt{h}}-1\right)\left(\frac{x\sqrt{h}+h}{2}\right)^{\frac{2}{\sqrt{h}}-1}} f_X(v)\mathrm{d}v \\
&= B(x,h)E[f_X(V)],
\end{aligned}
$$

where $V$ is a $Gamma(2h^{-1/2}-1,(x\sqrt{h}+h)/2)$ random variable, $B(x,h)$ is a factor outside the integral, and the integral itself can be considered as $E[f_X(V)]$. Similar as before, the random variable $V$ has mean $\mu_V = (2h^{-1/2}-1)(x\sqrt{h}+h/2)$ and

$$
Var(V) = (2h^{-1/2}-1)(x\sqrt{h}+h)^2/4.
$$

In the same fashion as in $E[f_X(W)]$ before, we have

$$
\begin{aligned}
E[f_X(V)] &= f_X\left(x+\sqrt{h}-\frac{x\sqrt{h}+h}{2}\right)+\frac{1}{2}f_X''\left(x+\sqrt{h}-\frac{x\sqrt{h}+h}{2}\right) \\
&\quad \times \left(\frac{2}{\sqrt{h}}-1\right)\left(\frac{x\sqrt{h}+h}{2}\right)^2 + \dots \\
&= f_X(x)+O(\sqrt{h}).
\end{aligned}
$$

Now, let $R(z)=\frac{\sqrt{2\pi}z^{z+\frac{1}{2}}}{e^z\Gamma(z+1)}$; then, $B(x,h)$ can be rewritten to become

$$
\begin{aligned}
B(x,h) &= \frac{\sqrt{2\pi}\left(\frac{2}{\sqrt{h}}-2\right)^{\frac{2}{\sqrt{h}}-\frac{3}{2}}}{e^{\frac{2}{\sqrt{h}}-2}R\left(\frac{2}{\sqrt{h}}-2\right)} \frac{e^{\frac{2}{\sqrt{h}}-2}R^2\left(\frac{1}{\sqrt{h}}-1\right)}{2\pi\left(\frac{1}{\sqrt{h}}-1\right)^{\frac{2}{\sqrt{h}}-1}} \frac{1}{2^{\frac{2}{\sqrt{h}}-1}(x\sqrt{h}+h)} \\
&= \frac{R^2\left(\frac{1}{\sqrt{h}}-1\right)}{2(x+\sqrt{h})\sqrt{\pi(1-\sqrt{h})}R\left(\frac{2}{\sqrt{h}}-2\right)h^{\frac{1}{4}}}.
\end{aligned}
$$

Thus, we obtain Equation 2.4, and the proof is completed.

## Proof of theorem 2.1.4

We have already expanded $J_h(x)$ until the $\sqrt{h}$ term. Now, extending it until the $h$ term results in

$$\begin{aligned}
J_h(x) &= f_X(x) + \sqrt{h} f'_X(x) + o(\sqrt{h}) + \frac{1}{2}(x^2\sqrt{h} + 2xh + h\sqrt{h}) \\
&\quad \times [f''_X(x) + \sqrt{h} f'''_X(x) + o(\sqrt{h})] + ... \\
&= f_X(x)\left[1 + \left\{f'_X(x) + \frac{1}{2}x^2 f''_X(x)\right\}\frac{\sqrt{h}}{f_X(x)}\right. \\
&\quad \left. + \left\{\left(x + \frac{1}{2}\right) f''_X(x) + x^2\left(\frac{x}{3} + \frac{1}{2}\right) f'''_X(x)\right\}\frac{h}{f_X(x)} + o(h)\right] \\
&= f_X(x)\left[1 + \frac{a(x)}{f_X(x)}\sqrt{h} + \frac{b(x)}{f_X(x)}h + o(h)\right],
\end{aligned}$$

where $a(x) = f'_X(x) + \frac{1}{2}x^2 f''_X(x)$, and $b(x) = \left(x + \frac{1}{2}\right) f''_X(x) + x^2\left(\frac{x}{3} + \frac{1}{2}\right) f'''_X(x)$. By taking the natural logarithm and using its expansion, we have

$$\begin{aligned}
\log J_h(x) &= \log f_X(x) + \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}\left[\frac{a(x)}{f_X(x)}\sqrt{h} + \frac{b(x)}{f_X(x)}h + o(h)\right]^k \\
&= \log f_X(x) + \frac{a(x)}{f_X(x)}\sqrt{h} + \left[b(x) - \frac{a^2(x)}{2f_X(x)}\right]\frac{h}{f_X(x)} + o(h).
\end{aligned}$$

Next, if we define $J_{4h}(x) = E[A_{4h}(x)]$ (using quadrupled bandwidth), i.e.,

$$\ln J_{4h}(x) = \ln f_X(x) + \frac{2a(x)}{f_X(x)}\sqrt{h} + \frac{4}{f_X(x)}\left[b(x) - \frac{a^2(x)}{2f_X(x)}\right] h + o(h),$$

we can set up conditions to eliminate the term $\sqrt{h}$ while keeping the term $\ln f_X(x)$. Now, since $\ln[J_h(x)]^{t_1}[J_{4h}(x)]^{t_2}$ equals

$$(t_1 + t_2)\ln f_X(x) + (t_1 + 2t_2)\frac{a(x)}{f_X(x)}\sqrt{h} + (t_1 + 4t_2)\left[b(x) - \frac{a^2(x)}{2f_X(x)}\right]\frac{h}{f_X(x)} + o(h),$$

the conditions we need are $t_1 + t_2 = 1$ and $t_1 + 2t_2 = 0$. It is obvious that the solution is $t_1 = 2$ and $t_2 = -1$, and we get

$$\ln[J_h(x)]^2[J_{4h}(x)]^{-1} = \ln f_X(x) - \frac{2}{f_X(x)}\left[b(x) - \frac{a^2(x)}{2f_X(x)}\right] h + o(h).$$

If we take the exponential function and use its expansion, we have

$$\begin{aligned}
[J_h(x)]^2[J_{4h}(x)]^{-1} &= f_X(x)\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\left[\frac{2}{f_X(x)}\left\{b(x)-\frac{a^2(x)}{2f_X(x)}\right\}h+o(h)\right]^k \\
&= f_X(x)\left[1-\frac{2}{f_X(x)}\left\{b(x)-\frac{a^2(x)}{2f_X(x)}\right\}h+o(h)\right] \\
&= f_X(x)-2\left[b(x)-\frac{a^2(x)}{2f_X(x)}\right]h+o(h).
\end{aligned}$$

# Proof of theorem 2.1.6

Because of the definition of $J_h(x)$ and $J_{4h}(x)$, we can rewrite $A_h(x)=J_h(x)+Y$ and $A_{4h}(x)=J_{4h}(x)+Z$, where $Y$ and $Z$ are random variables with $E(Y)$ and $E(Z)$ are both 0, $Var(Y)=Var[A_h(x)]$, and $Var(Z)=Var[A_{4h}(x)]$. Then, by the expansion $(1+p)^q=1+pq+O(p^2)$, we get

$$\begin{aligned}
\widetilde{f}_X(x) &= [J_h(x)]^2[J_{4h}(x)]^{-1}\left[1+\frac{Y}{J_h(x)}\right]^2\left[1+\frac{Z}{J_{4h}(x)}\right]^{-1} \\
&= [J_h(x)]^2[J_{4h}(x)]^{-1}\left[1+\frac{2Y}{J_h(x)}+O\left\{\frac{Y^2}{J_h^2(x)}\right\}\right] \\
&\quad\times\left[1-\frac{Z}{J_{4h}(x)}+O\left\{\frac{Z^2}{J_{4h}^2(x)}\right\}\right] \\
&= [J_h(x)]^2[J_{4h}(x)]^{-1}+\frac{2J_h(x)}{J_{4h}(x)}Y-\left[\frac{J_h(x)}{J_{4h}(x)}\right]^2Z+O[(Y+Z)^2].
\end{aligned}$$

Hence,

$$\begin{aligned}
E[\widetilde{f}_X(x)] &= [J_h(x)]^2[J_{4h}(x)]^{-1}+\frac{2J_h(x)}{J_{4h}(x)}E(Y)-\left[\frac{J_h(x)}{J_{4h}(x)}\right]^2E(Z) \\
&\quad +O[E\{(Y+Z)^2\}] \\
&= f_X(x)-2\left[b(x)-\frac{a(x)}{2f_X(x)}\right]h+o(h)+O\left(\frac{1}{nh^{\frac{1}{4}}}\right),
\end{aligned}$$

and its bias is

$$Bias[\widetilde{f}_X(x)]=-2\left[b(x)-\frac{a(x)}{2f_X(x)}\right]h+o(h)+O\left(\frac{1}{nh^{\frac{1}{4}}}\right).$$

# Proof of theorem 2.1.7

By usual calculation of i.i.d. random variables, we have

$$Cov[A_h(x), A_{4h}(x)] = \frac{1}{n}E[K(X_1;x,h)K(X_1;x,4h)] + O\left(\frac{1}{n}\right).$$

Now, for the expectation,

$$\begin{aligned}
&E[K(X_1;x,h)K(X_1;x,4h)] \\
&= \int_0^\infty \frac{t^{\frac{1}{\sqrt{h}}-1}e^{-\frac{t}{x\sqrt{h}+h}}}{\Gamma\left(\frac{1}{\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{1}{\sqrt{h}}}} \frac{t^{\frac{1}{2\sqrt{h}}-1}e^{-\frac{t}{2x\sqrt{h}+4h}}}{\Gamma\left(\frac{1}{2\sqrt{h}}\right)(2x\sqrt{h}+4h)^{\frac{1}{2\sqrt{h}}}} f_X(t)\mathrm{d}t \\
&= \frac{\Gamma\left(\frac{3}{2\sqrt{h}}-1\right)\left[\frac{2\sqrt{h}(x+\sqrt{h})(x+2\sqrt{h})}{3x+5\sqrt{h}}\right]^{\frac{3}{2\sqrt{h}}-1}}{\Gamma\left(\frac{1}{\sqrt{h}}\right)\Gamma\left(\frac{1}{2\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{1}{\sqrt{h}}}(2x\sqrt{h}+4h)^{\frac{1}{2\sqrt{h}}}} \\
&\quad\times \int_0^\infty \frac{t^{\left(\frac{3}{2\sqrt{h}}-1\right)-1}e^{-t\left[\frac{3x+5\sqrt{h}}{2\sqrt{h}(x+\sqrt{h})(x+2\sqrt{h})}\right]}}{\Gamma\left(\frac{3}{2\sqrt{h}}-1\right)\left[\frac{2\sqrt{h}(x+\sqrt{h})(x+2\sqrt{h})}{3x+5\sqrt{h}}\right]^{\frac{3}{2\sqrt{h}}-1}} f_X(t)\mathrm{d}t \\
&= C(x,h)E[f_X(T)],
\end{aligned}$$

where $C(x,h)$ is the factor outside the integral, and $T$ is a random variable with mean

$$\mu_T = \frac{3(x+\sqrt{h})(x+2\sqrt{h})}{3x+5\sqrt{h}} + O(\sqrt{h})$$

and variance $Var(T) = O(\sqrt{h})$. Utilizing Taylor expansion results in

$$\begin{aligned}
E[f_X(T)] &= f_X(x) + \left[\frac{3(x+\sqrt{h})(x+2\sqrt{h})}{3x+5\sqrt{h}} - x + O(\sqrt{h})\right] f'_X(x) + o(\sqrt{h}) \\
&\quad + \frac{1}{2}f''_X\left[\frac{3(x+\sqrt{h})(x+2\sqrt{h})}{3x+5\sqrt{h}} + O(\sqrt{h})\right] O(\sqrt{h}) \\
&= f_X(x) + O(\sqrt{h}).
\end{aligned}$$

Using the definition of $R(z)$ as before, we get

$$
\begin{aligned}
C(x,h) &= \frac{[2\sqrt{h}(x+\sqrt{h})(x+2\sqrt{h})]^{\frac{3}{2\sqrt{h}}-1}}{(x\sqrt{h}+h)^{\frac{1}{\sqrt{h}}}(2x\sqrt{h}+4h)^{\frac{1}{2\sqrt{h}}}(3x+5\sqrt{h})^{\frac{3}{2\sqrt{h}}-1}} \\
&\quad\times \frac{\sqrt{2\pi}\left(\frac{3}{2\sqrt{h}}-2\right)^{\frac{3}{2\sqrt{h}}-\frac{3}{2}}}{e^{\frac{3}{2\sqrt{h}}-2}R\left(\frac{3}{2\sqrt{h}}-2\right)} \frac{e^{\frac{1}{\sqrt{h}}-1}R\left(\frac{1}{\sqrt{h}}-1\right)}{\sqrt{2\pi}\left(\frac{1}{\sqrt{h}}-1\right)^{\frac{1}{\sqrt{h}}-\frac{1}{2}}} \frac{e^{\frac{1}{2\sqrt{h}}-1}R\left(\frac{1}{2\sqrt{h}}-1\right)}{\sqrt{2\pi}\left(\frac{1}{2\sqrt{h}}-1\right)^{\frac{1}{2\sqrt{h}}-\frac{1}{2}}} \\
&= \frac{R\left(\frac{1}{\sqrt{h}}-1\right)R\left(\frac{1}{2\sqrt{h}}-1\right)}{2h^{\frac{1}{4}}\sqrt{\pi}R\left(\frac{3}{2\sqrt{h}}-2\right)(3x+5\sqrt{h})} \frac{\left(\frac{3}{2}-2\sqrt{h}\right)^{\frac{3}{2\sqrt{h}}-\frac{3}{2}}}{(2-2\sqrt{h})^{\frac{1}{\sqrt{h}}-\frac{1}{2}}(1-2\sqrt{h})^{\frac{1}{2\sqrt{h}}-\frac{1}{2}}} \\
&\quad\times \left(\frac{x+\sqrt{h}}{3x+5\sqrt{h}}\right)^{\frac{1}{2\sqrt{h}}-1}\left(\frac{2x+4\sqrt{h}}{3x+5\sqrt{h}}\right)^{\frac{1}{\sqrt{h}}-1},
\end{aligned}
$$

when $x > h$ (for $x \leq h$, the calculation is similar). Hence, the covariance term is

$$
\begin{aligned}
Cov[A_h(x), A_{4h}(x)] = {} & \frac{R\left(\frac{1}{\sqrt{h}}-1\right)R\left(\frac{1}{2\sqrt{h}}-1\right)}{2\sqrt{\pi}R\left(\frac{3}{2\sqrt{h}}-2\right)(3x+5\sqrt{h})} \frac{\left(\frac{3}{2}-2\sqrt{h}\right)^{\frac{3}{2\sqrt{h}}-\frac{3}{2}}}{(2-2\sqrt{h})^{\frac{1}{\sqrt{h}}-\frac{1}{2}}(1-2\sqrt{h})^{\frac{1}{2\sqrt{h}}-\frac{1}{2}}} \\
& \times \left(\frac{x+\sqrt{h}}{3x+5\sqrt{h}}\right)^{\frac{1}{2\sqrt{h}}-1}\left(\frac{2x+4\sqrt{h}}{3x+5\sqrt{h}}\right)^{\frac{1}{\sqrt{h}}-1} \frac{f_X(x)}{nh^{\frac{1}{4}}} + O\left(\frac{h^{\frac{1}{4}}}{n}\right),
\end{aligned}
$$

when $xh^{-1} \to \infty$, and

$$
\begin{aligned}
Cov[A_h(x), A_{4h}(x)] = {} & \frac{R\left(\frac{1}{\sqrt{h}}-1\right)R\left(\frac{1}{2\sqrt{h}}-1\right)}{2\sqrt{\pi}R\left(\frac{3}{2\sqrt{h}}-2\right)(3c\sqrt{h}+5)} \frac{\left(\frac{3}{2}-2\sqrt{h}\right)^{\frac{3}{2\sqrt{h}}-\frac{3}{2}}}{(2-2\sqrt{h})^{\frac{1}{\sqrt{h}}-\frac{1}{2}}(1-2\sqrt{h})^{\frac{1}{2\sqrt{h}}-\frac{1}{2}}} \\
& \times \left(\frac{c\sqrt{h}+1}{3c\sqrt{h}+5}\right)^{\frac{1}{2\sqrt{h}}-1}\left(\frac{2c\sqrt{h}+4}{3c\sqrt{h}+5}\right)^{\frac{1}{\sqrt{h}}-1} \frac{f_X(x)}{nh^{\frac{3}{4}}} + O\left(\frac{1}{nh^{\frac{1}{4}}}\right),
\end{aligned}
$$

when $xh^{-1} \to c > 0$.

## Proof of theorem 2.1.8

It is easy to prove that $[J_h(x)][J_{4h}(x)]^{-1} = 1+O(\sqrt{h})$ by using the expansion of $(1+p)^q$. This fact brings us to

$$\begin{aligned} Var[\widetilde{f}_X(x)] &= Var[2\{1+O(\sqrt{h})\}Y - \{1+O(\sqrt{h})\}^2 Z] + Var[O\{(Y+Z)^2\}] \\ &= Var[2A_h(x) - A_{4h}(x)] + o\left(\frac{1}{nh^{\frac{1}{4}}}\right) \\ &= 4Var[A_h(x)] + Var[A_{4h}(x)] - 4Cov[A_h(x), A_{4h}(x)] + o\left(\frac{1}{nh^{\frac{1}{4}}}\right). \end{aligned}$$

Last, since the equation above is just a linear combination of two variance formulas, the orders of the variance do not change, which are $n^{-1}h^{-1/4}$ in the interior and $n^{-1}h^{-3/4}$ in the boundary region.

## Proof of theorem 3.1.1

Let $j_h(x) = E[\widehat{F}_h(x)]$, and extend the expansion until $h^4$ term. In detail,

$$\begin{aligned} j_h(x) &= \int_{-\infty}^{\infty} W\left(\frac{x-v}{h}\right) f_X(v)\mathrm{d}v \\ &= \int_{-\infty}^{\infty} F_X(x-hw)K(w)\mathrm{d}w \\ &= F_X(x) + h^2 a_2(x) + h^4 a_4(x) + o(h^4) \\ &= F_X(x)\left[1 + h^2\frac{a_2(x)}{F_X(x)} + h^4\frac{a_4(x)}{F_X(x)} + o(h^4)\right], \end{aligned}$$

where $a_2(x) = \frac{f'_X(x)}{2}\int_{-\infty}^{\infty} w^2 K(w)\mathrm{d}w$ and $a_4(x) = \frac{f'''_X(x)}{24}\int_{-\infty}^{\infty} w^4 K(w)\mathrm{d}w$. By taking a natural logarithm and using its expansion, we have

$$\begin{aligned} \log j_h(x) &= \log F_X(x)\left[1 + h^2\frac{a_2(x)}{F_X(x)} + h^4\frac{a_4(x)}{F_X(x)} + o(h^4)\right] \\ &= \log F_X(x) + \sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\left[h^2\frac{a_2(x)}{F_X(x)} + h^4\frac{a_4(x)}{F_X(x)} + o(h^4)\right]^k \\ &= \log F_X(x) + h^2\frac{a_2(x)}{F_X(x)} + h^4\frac{2a_4(x)F_X(x) - a_2^2(x)}{2F_X^2(x)} + o(h^4). \end{aligned}$$

Next, if we define $j_{ah}(x) = E[\widehat{F}_{ah}(x)]$, i.e.

$$\log j_{ah}(x) = \log F_X(x) + a^2h^2\frac{a_2(x)}{F_X(x)} + a^4h^4\frac{2a_4(x)F_X(x) - a_2^2(x)}{2F_X^2(x)} + o(h^4),$$

we can set up some conditions to eliminate the term $h^2$ but keep the term $\log F_X(x)$. Since

$$\begin{aligned}\log[j_h(x)]^{t_1}[j_{ah}(x)]^{t_2} &= (t_1+t_2)\log F_X(x) + (t_1+a^2t_2)h^2\frac{a_2(x)}{F_X(x)} \\ &\quad +(t_1+a^4t_2)h^4\frac{2a_4(x)F_X(x)-a_2^2(x)}{2F_X^2(x)} + o(h^4),\end{aligned}$$

the conditions we need are $t_1+t_2=1$ and $t_1+a^2t_2=0$. It is obvious that the solutions are $t_1=\frac{a^2}{a^2-1}$ and $t_2=-\frac{1}{a^2-1}$, and we get

$$\log[j_h(x)]^{\frac{a^2}{a^2-1}}[j_{ah}(x)]^{-\frac{1}{a^2-1}} = \log F_X(x) - h^4a^2\frac{2a_4(x)F_X(x)-a_2^2(x)}{2F_X^2(x)} + o(h^4).$$

To neutralize the natural logarithmic function, we can use exponential function and its expansion, then we have

$$\begin{aligned}[j_h(x)]^{\frac{a^2}{a^2-1}}[j_{ah}(x)]^{-\frac{1}{a^2-1}} &= \exp[\log\{j_h(x)\}^{\frac{a^2}{a^2-1}}\{j_{ah}(x)\}^{-\frac{1}{a^2-1}}] \\ &= F_X(x)\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}\left[h^4a^2\frac{2a_4(x)F_X(x)-a_2^2(x)}{2F_X^2(x)} + o(h^4)\right]^k \\ &= F_X(x) + h^4a^2\frac{a_2^2(x)-2a_4(x)F_X(x)}{2F_X(x)} + o(h^4).\end{aligned}$$

# Proof of theorem 3.1.4

In order to investigate the bias of our proposed estimator, we rewrite $\widehat{F}_h(x) = j_h(x)+Y$ and $\widehat{F}_{ah}(x) = j_{ah}(x)+Z$, where $Y$ and $Z$ are random variables with $E(Y)=E(Z)=0$, $Var(Y)=Var[\widehat{F}_h(x)]$, and $Var(Z)=Var[\widehat{F}_{ah}(x)]$. These forms are actually reasonable, because of the definition of $j_h(x)$ and $j_{ah}(x)$. Then, by the expansion $(1+p)^q = 1+pq+O(p^2)$, we have

$$\begin{aligned}\widetilde{F}_X(x) &= [j_h(x)+Y]^{\frac{a^2}{a^2-1}}[j_{ah}(x)+Z]^{-\frac{1}{a^2-1}} \\ &= [j_h(x)]^{\frac{a^2}{a^2-1}}[j_{ah}(x)]^{-\frac{1}{a^2-1}}\left[1+\frac{Y}{j_h(x)}\right]^{\frac{a^2}{a^2-1}}\left[1+\frac{Z}{j_{ah}(x)}\right]^{-\frac{1}{a^2-1}} \\ &= [j_h(x)]^{\frac{a^2}{a^2-1}}[j_{ah}(x)]^{-\frac{1}{a^2-1}} + \frac{a^2}{a^2-1}\left[\frac{j_h(x)}{j_{ah}(x)}\right]^{\frac{1}{a^2-1}}Y \\ &\quad -\frac{1}{a^2-1}\left[\frac{j_h(x)}{j_{ah}(x)}\right]^{\frac{a^2}{a^2-1}}Z + O[(Y+Z)^2].\end{aligned}$$

Hence,

$$\begin{aligned} E[\widetilde{F}_X(x)] &= [j_h(x)]^2[j_{ah}(x)]^{-1} + \frac{2j_h(x)}{j_{ah}(x)}E(Y) - \left[\frac{j_h(x)}{j_{ah}(x)}\right]^2 E(Z) \\ &\quad +O[E\{(Y+Z)^2\}] \\ &= F_X(x) + h^4a^2\frac{a_2^2(x) - 2a_4(x)F_X(x)}{2F_X(x)} + o(h^4) + O\left(\frac{1}{n}\right), \end{aligned}$$

and the bias is obtained.

## Proof of theorem 3.1.7

Before we derive the variance, we need to calculate $\frac{j_h(x)}{j_{ah}(x)}$ first. Once again by using $(1+p)^q = 1 + pq + O(p^2)$, we get

$$\begin{aligned} \frac{j_h(x)}{j_{ah}(x)} &= \frac{F_X(x)\left[1 + h^2\frac{a_2(x)}{F_X(x)} + h^4\frac{a_4(x)}{F_X(x)} + o(h^4)\right]}{F_X(x)\left[1 + a^2h^2\frac{a_2(x)}{F_X(x)} + a^4h^4\frac{a_4(x)}{F_X(x)} + o(h^4)\right]} \\ &= \frac{1 + h^2\frac{a_2(x)}{F_X(x)} + h^4\frac{a_4(x)}{F_X(x)} + o(h^4)}{1 + O(h^2)} \\ &= 1 + h^2\frac{a_2(x)}{F_X(x)} + h^4\frac{a_4(x)}{F_X(x)} + o(h^4) + O(h^2) = 1 + O(h^2). \end{aligned}$$

The calculation of the variance is

$$\begin{aligned} Var[\widetilde{F}_X(x)] &= Var\left[\frac{a^2}{a^2-1}\{1+O(h^2)\}^{\frac{1}{a^2-1}}Y - \frac{1}{a^2-1}\{1+O(h^2)\}^{\frac{a^2}{a^2-1}}Z\right] \\ &\quad +O[Var\{(Y+Z)^2\}] \\ &= Var\left(\frac{a^2}{a^2-1}Y - \frac{1}{a^2-1}Z\right) + O\left(\frac{h^2}{n} + \frac{1}{n^2}\right) \\ &= Var\left[\frac{a^2}{a^2-1}\widehat{F}_h(x) - \frac{1}{a^2-1}\widehat{F}_{ah}(x)\right] + o\left(\frac{h}{n}\right). \end{aligned}$$

Because this is just a variance of linear combination of two standard kernel distribution function estimators, the order of the variance does not change, that is $\frac{1}{n}$. For the explicit formula of the variance, first we calculate

$$\frac{a^4}{(a^2-1)^2}Var[\widehat{F}_h(x)] + \frac{1}{(a^2-1)^2}Var[\widehat{F}_{ah}(x)] - \frac{2a^2}{(a^2-1)^2}Cov[\widehat{F}_h(x), \widehat{F}_{ah}(x)].$$

Since we already knew about the formulas of $Var[\widehat{F}_h(x)]$ and $Var[\widehat{F}_{ah}(x)]$, we only need to take a look to the covariance part, that is

$$\frac{1}{n}\left[E\left\{W\left(\frac{x-X_1}{h}\right)W\left(\frac{x-X_1}{ah}\right)\right\}-E\left\{W\left(\frac{x-X_1}{h}\right)\right\}E\left\{W\left(\frac{x-X_1}{ah}\right)\right\}\right].$$

Because,

$$E\left[W\left(\frac{x-X_1}{h}\right)\right]=E\left[W\left(\frac{x-X_1}{ah}\right)\right]=F_X(x)+O(h^2),$$

we only need to calculate

$$\begin{aligned}
&E\left[W\left(\frac{x-X_1}{h}\right)W\left(\frac{x-X_1}{ah}\right)\right]\\
&\quad=\int_{-\infty}^{\infty}F_X(x-yh)\left[K(y)W\left(\frac{y}{a}\right)+\frac{1}{a}W(y)K\left(\frac{y}{a}\right)\right]\mathrm{d}y\\
&\quad=\int_{-\infty}^{\infty}[F_X(x)-hyf_X(x)+o(h)]\left[K(y)W\left(\frac{y}{a}\right)+\frac{1}{a}W(y)K\left(\frac{y}{a}\right)\right]\mathrm{d}y\\
&\quad=F_X(x)\left[\int_{-\infty}^{\infty}K(y)W\left(\frac{y}{a}\right)\mathrm{d}y+\frac{1}{a}\int_{-\infty}^{\infty}W(y)K\left(\frac{y}{a}\right)\mathrm{d}y\right]-hf_X(x)r_2+o(h),
\end{aligned}$$

where $r_2=\int_{-\infty}^{\infty}y\left[K(y)W\left(\frac{y}{a}\right)+\frac{1}{a}W(y)K\left(\frac{y}{a}\right)\right]\mathrm{d}y$. For the first term of the right-hand side, we have

$$\int_{-\infty}^{\infty}K(y)W\left(\frac{y}{a}\right)\mathrm{d}y=1-\frac{1}{a}\int_{-\infty}^{\infty}W(y)K\left(\frac{y}{a}\right)\mathrm{d}y.$$

Thus we get

$$E\left[W\left(\frac{x-X_1}{h}\right)W\left(\frac{x-X_1}{ah}\right)\right]=F_X(x)-hf_X(x)r_2+o(h).$$

As a result, we can show that

$$Cov[\widehat{F}_h(x),\widehat{F}_{ah}(x)]=\frac{1}{n}F_X(x)[1-F_X(x)]-\frac{h}{n}f_X(x)r_2+o\left(\frac{h}{n}\right),$$

and then

$$Var[\widetilde{F}_X(x)]=\frac{1}{n}F_X(x)[1-F_X(x)]-\frac{h}{n}\left[\frac{2(a^4+1)}{(a^2-1)^2}r_1+r_2\right]f_X(x)+o\left(\frac{h}{n}\right).$$

## Proof of theorem 4.1.1

Utilizing the usual reasoning of *i.i.d.* random variables and the transformation property of expectation, with $Y = g^{-1}(X_1)$, we have

$$\begin{aligned}
E[\widetilde{F}_X(x)] &= \int_{-\infty}^{\infty} W\left(\frac{g^{-1}(x)-y}{h}\right) f_Y(y)\mathrm{d}y \\
&= \frac{1}{h}\int_{-\infty}^{\infty} F_Y(y) K\left(\frac{g^{-1}(x)-y}{h}\right)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} F_Y(g^{-1}(x)-hv)K(v)\mathrm{d}v \\
&= \int_{-\infty}^{\infty}\left[F_Y(g^{-1}(x)) - hvf_Y(g^{-1}(x)) + \frac{h^2}{2}v^2 f'_Y(g^{-1}(x)) + o(h^2)\right]K(v)\mathrm{d}v \\
&= F_X(x) + \frac{h^2}{2}c_1(x)\int_{-\infty}^{\infty} v^2 K(v)\mathrm{d}v + o(h^2),
\end{aligned}$$

and we obtained the $Bias[\widetilde{F}_X(x)]$. For the variance of $\widetilde{F}_X(x)$, we first calculate

$$\begin{aligned}
&E\left[W^2\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right)\right] \\
&\quad= \frac{2}{h}\int_{-\infty}^{\infty} F_Y(y) W\left(\frac{g^{-1}(x)-y}{h}\right) K\left(\frac{g^{-1}(x)-y}{h}\right)\mathrm{d}y \\
&\quad= 2\int_{-\infty}^{\infty}[F_Y(g^{-1}(x)) - hvf_Y(g^{-1}(x)) + o(h)]W(v)K(v)\mathrm{d}v \\
&\quad= F_X(x) - 2hg'(g^{-1}(x))f_X(x)r_1 + o(h),
\end{aligned}$$

and we got the variance.

## Proof of theorem 4.1.3

For some $\delta > 0$, using Hölder and Cramér $c_r$ inequalities, we have

$$\begin{aligned}
&E\left[\left|W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right) - E\left\{W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right)\right\}\right|^{2+\delta}\right] \\
&\qquad\le 2^{2+\delta}E\left[\left|W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right)\right|^{2+\delta}\right].
\end{aligned}$$

But, since $0 \le W(v) \le 1$ for any $v \in \mathbb{R}$, then

$$E\left[\left|W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right) - E\left\{W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right)\right\}\right|^{2+\delta}\right] \le 2^{2+\delta} < \infty.$$

Also, because $Var\left[W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right)\right] = O(1)$, we get

$$\frac{E\left[\left|W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right) - E\left\{W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right)\right\}\right|^{2+\delta}\right]}{n^{\delta/2}\left[Var\left\{W\left(\frac{g^{-1}(x)-g^{-1}(X_1)}{h}\right)\right\}\right]^{1+\delta/2}} \to 0$$

when $n \to \infty$. Hence, by Loeve (1963), and with the fact $\widetilde{F}_X(x) \to_p F_X(x)$, we can conclude its asymptotic normality.

# Proof of theorem 4.1.4

Let $F_Y$ and $\widehat{F}_Y$ be the distribution function and the naive kernel distribution function estimator, respectively, of $Y_1, Y_2, ..., Y_n$, where $Y_i = g^{-1}(X_i)$. Since $\widehat{F}_Y$ is a naive kernel distribution function, then Nadaraya (1964) guarantees that $\sup_{y\in\mathbb{R}} |\widehat{F}_Y(y) - F_Y(y)| \to_{a.s.} 0$, which implies that

$$\sup_{x\in\Omega} |\widehat{F}_Y(g^{-1}(x)) - F_Y(g^{-1}(x))| \to_{a.s.} 0.$$

However, because $F_Y(g^{-1}(x)) = F_X(x)$, and it is clear that $\widehat{F}_Y(g^{-1}(x)) = \widetilde{F}_X(x)$, then this theorem is proven.

# Proof of theorem 4.1.5

Using the similar reasoning as in the proof of theorem 4.1.1, we have

$$\begin{aligned}
E[\widehat{f}_X(x)] &= \frac{1}{hg'(g^{-1}(x))}\int_{-\infty}^{\infty} K\left(\frac{g^{-1}(x)-y}{h}\right) f_Y(y)\mathrm{d}y \\
&= \frac{1}{g'(g^{-1}(x))}\int_{-\infty}^{\infty} f_Y(g^{-1}(x)-hv)K(v)\mathrm{d}v \\
&= \frac{f_Y(g^{-1}(x))}{g'(g^{-1}(x))} + \frac{h^2 f_Y''(g^{-1}(x))}{2g'(g^{-1}(x))}\int_{-\infty}^{\infty} v^2K(v)\mathrm{d}v + o(h^2),
\end{aligned}$$

and we obtained the bias formula. For the variance, first we have to calculate

$$\begin{aligned}
\frac{1}{hg'(g^{-1}(x))}E\left[K^2\left(\frac{g^{-1}(x)-Y}{h}\right)\right] &= \frac{1}{g'(g^{-1}(x))}\int_{-\infty}^{\infty} f_Y(g^{-1}(x)-hv)K^2(v)\mathrm{d}v \\
&= f_X(x)\int_{-\infty}^{\infty} K^2(v)\mathrm{d}v + o(1),
\end{aligned}$$

and the rests are easily done.

## Proof of theorem 4.2.2

First, we need to consider the following inequality

$$\begin{aligned}
|KS_n - \widetilde{KS}| &= \left|\sup_{v\in\Omega}|F_n(v) - F(v)| - \sup_{z\in\Omega}|\widetilde{F}_X(z) - F(z)|\right| \\
&\leq \sup_{x\in\Omega}\Big||F_n(x) - F(x)| - |\widetilde{F}_X(x) - F(x)|\Big| \\
&\leq \sup_{x\in\Omega}|F_n(x) - F(x) - \widetilde{F}_X(x) + F(x)| \\
&= \sup_{x\in\Omega}|\widetilde{F}_X(x) - F_n(x)|.
\end{aligned}$$

Now, let $F_{n,Y}$ and $\widehat{F}_Y$ be the empirical distribution function and the naive kernel distribution function estimator, respectively, of $Y_1, Y_2, ..., Y_n$, where $Y_i = g^{-1}(X_i)$. Hence, Omelka *et al.* (2009) guarantees that $\sup_{y\in\mathbb{R}}|\widehat{F}_Y(y) - F_{n,Y}(y)| = o_p(n^{-1/2})$, which further implies that

$$\sup_{x\in\Omega}|\widehat{F}_Y(g^{-1}(x)) - F_{n,Y}(g^{-1}(x))| \to_p 0$$

with rate $n^{-1/2}$. But, $\widehat{F}_Y(g^{-1}(x)) = \widetilde{F}_X(x)$ and $F_{n,Y}(g^{-1}(x)) = F_n(x)$, which conclude that the equivalency is proven.

## Proof of theorem 4.2.3

In this proof, we assume the bandwidth $h = o(n^{-1/4})$. Let us define

$$\Delta_n = n\int_{-\infty}^{\infty}[\widetilde{F}_X(x) - F(x)]^2 \mathrm{d}F(x) - n\int_{-\infty}^{\infty}[F_n(x) - F(x)]^2 \mathrm{d}F(x).$$

Then, we have

$$\begin{aligned}
\Delta_n &= n\int_{-\infty}^{\infty}\left[\widetilde{F}_X(x) - F(x) - F_n(x) + F(x)\right]\left[\widetilde{F}_X(x) - F(x) + F_n(x) - F(x)\right]\mathrm{d}F(x) \\
&= n\int_{-\infty}^{\infty}\frac{1}{n}\sum_{i=1}^{n}[W_i^*(x) - I_i^*(x)]\,\frac{1}{n}\sum_{j=1}^{n}\left[W_j^*(x) + I_j^*(x)\right]\mathrm{d}F(x),
\end{aligned}$$

where

$$W_i^*(x) = W\left(\frac{g^{-1}(x) - g^{-1}(X_i)}{h}\right) - F(x) \quad \text{and} \quad I_i^*(x) = I(X_i \leq x) - F(x).$$

Note that if $i \neq j$, $W_i(\cdot)$ and $W_j(\cdot)$, also $I_i^*(\cdot)$ and $I_j^*(\cdot)$, are independent.

It follows from the Chauchy-Schwarz Inequality that

$$E(|\Delta_n|) \leq n \int_{-\infty}^{\infty} \sqrt{E\left[\left\{\frac{1}{n}\sum_{i=1}^{n}(W_i^*(x) - I_i^*(x))\right\}^2\right] E\left[\left\{\frac{1}{n}\sum_{j=1}^{n}(W_j^*(x) + I_j^*(x))\right\}^2\right]} \mathrm{d}F(x).$$

Let us define the bias

$$b_n(x) = E\left[W\left(\frac{g^{-1}(x) - g^{-1}(X_i)}{h}\right)\right] - F(x) = O(h^2).$$

Hence, it follows from the independence that

$$\begin{aligned} E\left[\left\{\frac{1}{n}\sum_{i=1}^{n}(W_i^*(x) - I_i^*(x))\right\}^2\right] &= E\left[\left\{\frac{1}{n}\sum_{i=1}^{n}(W_i^*(x) - b_n(x) - I_i^*(x))\right\}^2\right] + b_n^2(x) \\ &= \frac{1}{n}E[\{W_1^*(x) - b_n(x) + I_1^*(x)\}^2] + b_n^2(x). \end{aligned}$$

Furthermore, we have

$$\begin{aligned} & E[\{W_1^*(x) - b_n(x) - I_1^*(x)\}^2] \\ & \quad = E[\{W_1^*(x) - I_1^*(x)\}^2] - 2b_n(x)E[W_1^*(x) - I_1^*(x)] + b_n^2(x) \\ & \quad = E[\{W_1^*(x)\}^2 - 2W_1^*(x)I_1^*(x) - \{I_1^*(x)\}^2] - b_n^2(x). \end{aligned}$$

It follows from the mean squared error of $\widetilde{F}_X(x)$ that

$$E[\{W_1^*(x)\}^2] = F(x)[1 - F(x)] - 2hr_1 g'(g^{-1}(x))f_X(x) + O(h^2).$$

From the definition, we have

$$\begin{aligned} & E[W_1^*(x)I_1^*(x)] \\ & \quad = E\left[W\left(\frac{g^{-1}(x) - g^{-1}(X_1)}{h}\right) I(X_1 \leq x) - F(x)W\left(\frac{g^{-1}(x) - g^{-1}(X_1)}{h}\right)\right. \\ & \qquad \left. -F(x)I(X_1 \leq x) + F^2(x)\right] \\ & \quad = E\left[W\left(\frac{g^{-1}(x) - g^{-1}(X_1)}{h}\right) I(X_1 \leq x)\right] - F^2(x) - b_n(x)F(x). \end{aligned}$$

For the first term we have

$$\begin{aligned} & E\left[W\left(\frac{g^{-1}(x) - Y}{h}\right) I(Y \leq g^{-1}(x))\right] \\ & \quad = \int_{-\infty}^{g^{-1}(x)} W\left(\frac{g^{-1}(x) - y}{h}\right) f_Y(y)\mathrm{d}y \\ & \quad = W(0)F_Y(g^{-1}(x)) + \int_0^{\infty} F_Y(g^{-1}(x) - hv)K(v)dv \\ & \quad = W(0)F(x) + F(x)\int_0^{\infty} K(v)dv + O(h). \end{aligned}$$

Since $K(\cdot)$ is symmetric around the origin, we have $W(0) = 1/2$. Thus,

$$E[W_1^*(x)I_1^*(x)] = F(x)[1 - F(x)] + O(h).$$

Next we will evaluate $E\left[\{n^{-1}\sum_{i=1}^n (W_i^*(x) + I_i^*(x))\}^2\right]$. Using the bias term $b_n(x)$, we have

$$\begin{aligned}
E\left[\left\{\frac{1}{n}\sum_{i=1}^n (W_i^*(x) + I_i^*(x))\right\}^2\right] &= E\left[\left\{\frac{1}{n}\sum_{i=1}^n (W_i^*(x) - b_n(x) + I_i^*(x))\right\}^2\right] + b_n^2(x) \\
&= \frac{1}{n}E[\{W_1^*(x) - b_n(x) + I_1^*(x)\}^2] + b_n^2.
\end{aligned}$$

Based on previous calculations, we get

$$E\left[\left\{\frac{1}{n}\sum_{i=1}^n (W_i^*(x) + I_i^*(x))\right\}^2\right] = O\left(\frac{1}{n} + h^4\right).$$

Therefore, if $h = o(n^{-1/4})$, we have $E(|\Delta_n|) = o(1)$. Using the Markov Inequality, we can show that $\Delta_n \to_p 0$, and then two statistics are equivalent under $H_0$.

# Some lemmas needed to prove the theorems

Though sometimes not stated explicitly in the proofs of our theorems, the following lemmas are needed for the calculations.

**Lemma .0.1.** *Under the condition D1, the following equations hold*

$$\int_{-\infty}^{\infty} V(x)K(x)\mathrm{d}x = \frac{1}{2} \tag{30}$$

$$\int_{-\infty}^{\infty} xV(x)K(x)\mathrm{d}x = -\frac{1}{2}\int_{-\infty}^{\infty} V(x)W(x)\mathrm{d}x \tag{31}$$

$$\int_{-\infty}^{\infty} \mathbb{V}(x)K(x)\mathrm{d}x = \int_{-\infty}^{\infty} V(x)W(x)\mathrm{d}x. \tag{32}$$

*Proof.* All of the above equations can be proven using the integration by parts and the definitions of $V(x)$, $\mathbb{V}(x)$, and $W(x)$. □

**Lemma .0.2.** *Let $f_Y(t)$ and $S_Y(t)$ be the probability density function and the survival function of $Y = g^{-1}(X)$, and let $a(t) = \int_t^\infty g'(y)S_Y(y)\mathrm{d}y$ and $A(t) = \int_t^\infty g'(y)a(y)\mathrm{d}y$. Then, under the condition D6, we have for $t \in \Omega$,*

$$f_Y(g^{-1}(t)) = g'(g^{-1}(t))f_X(t) \tag{33}$$

$$S_Y(g^{-1}(t)) = S_X(t) \tag{34}$$

$$a(g^{-1}(t)) = \mathbb{S}_X(t) \tag{35}$$

$$A(g^{-1}(t)) = \bar{\mathbb{S}}_X(t). \tag{36}$$

*Proof.* Using the change-of-variable technique, it is obvious that $f_Y(t) = g'(t)$ $f_X(g(t))$, and it validifies equation (33). For equation (34), by the definition,

$$S_Y(g^{-1}(t)) = \Pr[Y > g^{-1}(t)] = \Pr[g(Y) > t] = \Pr(X > t) = S_X(t).$$

Equation (35) is easily done using integration by $x = g(y)$ substitution, which is

$$a(g^{-1}(t)) = \int_{g^{-1}(t)}^{\infty} g'(y)S_Y(y)\mathrm{d}y = \int_t^{\omega''} S_Y(g^{-1}(x))\mathrm{d}x = \int_t^{\omega''} S_X(x)\mathrm{d}x = \mathbb{S}_X(t).$$

The same fashion goes for equation (36). $\square$

**Remark .0.3.** *The ideas to construct our proposed estimators actually came from lemma .0.2. We intentionally constructed the estimators of $S_X(t)$ and $\mathbb{S}_X(t)$ using the relationships stated at equation (34) and equation (35), respectively. We refer to this lemma as the change-of-variable properties.*

**Lemma .0.4.** *Let $a(t) = \int_t^\infty g'(y)S_Y(y)\mathrm{d}y$ and*

$$\widehat{a}(t) = \int_t^\infty g'(y)\widehat{S}_Y(y)\mathrm{d}y = \frac{1}{n}\sum_{i=1}^{n}\int_t^\infty g'(y)V\left(\frac{y - Y_i}{h}\right)\mathrm{d}y \qquad (37)$$

*be the naive kernel estimator of $a(t)$. If $B \subset \mathbb{R}$ is an interval where both $\widehat{a}(t)$ and $a(t)$ are bounded, then $\sup_{t\in B}|\widehat{a}(t) - a(t)| \to_{a.s.} 0$.*

*Proof.* Since $\widehat{a}(t)$ and $a(t)$ are both bounded, non-increasing, and continuous on $B$, then for any $\varepsilon > 0$, we can find $k$ number of points on $A$ such that

$$-\infty \le \inf A = t_1 < t_2 < ... < t_k = \sup A \le \infty,$$

and $a(t_j) - a(t_{j+1}) \le \varepsilon/2$, $j = 1, 2, ..., k-1$. For any $t \in B$, it is clear that there exists $j$ such that $t_j \le t < t_{j+1}$. For that particular $j$, we have

$$\widehat{a}(t_j) \ge \widehat{a}(t) \ge \widehat{a}(t_{j+1}) \quad \text{and} \quad a(t_j) \ge a(t) \ge a(t_{j+1}),$$

which result in

$$\widehat{a}(t_{j+1}) - a(t_{j+1}) - \frac{\varepsilon}{2} \le \widehat{a}(t) - a(t) \le \widehat{a}(t_j) - a(t_j) + \frac{\varepsilon}{2}.$$

Therefore,

$$\sup_{t\in B}|\widehat{a}(t) - a(t)| \le \sup_j |\widehat{a}(t_j) - a(t_j)| + \varepsilon.$$

Now, because $\widehat{a}(t)$ is a naive kernel estimator, it is clear that for fix $t_0$, $\widehat{a}(t_0)$ converges almost surely to $a(t_0)$. Thus, we get $|\widehat{a}(t_0) - a(t_0)| \to_{a.s.} 0$. Hence, for any $\varepsilon > 0$, almost surely $\sup_{t\in B}|\widehat{a}(t) - a(t)| \le \varepsilon$ when $n \to \infty$, which concludes the proof. $\square$

# Proof of theorem 5.1.1

Utilizing the usual reasoning of *i.i.d.* random variables and the transformation property of expectation, and with the fact

$$V_{1,h}(x,y) = \frac{1}{h}\int_{-\infty}^{y} K\left(\frac{x-z}{h}\right)\mathrm{d}z,$$

we have

$$\begin{aligned}
E[\widetilde{S}_{X,1}(t)] &= E[V_{1,h}(g^{-1}(t), g^{-1}(X_1))] \\
&= \int_{-\infty}^{\infty} V_{1,h}(g^{-1}(t), y) f_Y(y)\mathrm{d}y \\
&= \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{g^{-1}(t)-y}{h}\right) S_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} S_Y(g^{-1}(t)-hu)K(u)\mathrm{d}u \\
&= \int_{-\infty}^{\infty}\left[S_Y(g^{-1}(t)) + huf_Y(g^{-1}(t)) - \frac{h^2}{2}u^2 f_Y'(g^{-1}(t)) + o(h^2)\right]K(u)\mathrm{d}u \\
&= S_X(t) - \frac{h^2}{2}b_1(t)\int_{-\infty}^{\infty} u^2 K(u)\mathrm{d}u + o(h^2),
\end{aligned}$$

and we have $Bias[\widetilde{S}_{X,1}(t)]$. For the variance of $\widetilde{S}_{X,1}(t)$, we first calculate

$$\begin{aligned}
E[V_{1,h}^2(g^{-1}(t), g^{-1}(X_1))] &= \int_{-\infty}^{\infty} V_{1,h}^2(g^{-1}(t), y) f_Y(y)\mathrm{d}y \\
&= \frac{2}{h}\int_{-\infty}^{\infty} V_{1,h}(g^{-1}(t), y) K\left(\frac{g^{-1}(t)-y}{h}\right) S_Y(y)\mathrm{d}y \\
&= 2\int_{-\infty}^{\infty}[S_Y(g^{-1}(t)) + huf_Y(g^{-1}(t)) + o(h)]V(u)K(u)\mathrm{d}u \\
&= S_X(t) - hg'(g^{-1}(t))f_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h).
\end{aligned}$$

Hence, the variance is

$$\begin{aligned}
Var[\widetilde{S}_{X,1}(t)] &= \frac{1}{n}[E\{V_{1,h}^2(g^{-1}(t), g^{-1}(X_1))\} - E^2\{V_{1,h}(g^{-1}(t), g^{-1}(t))\}] \\
&= \frac{1}{n}S_X(t)F_X(t) - \frac{h}{n}g'(g^{-1}(t))f_X(t)\int_{-\infty}^{\infty} V(y)W(y)\mathrm{d}y + o\left(\frac{h}{n}\right).
\end{aligned}$$

For the calculation of $Bias[\mathbb{S}_{X,1}(t)]$, recall that

$$\mathbb{V}_{1,h}(g^{-1}(t), y) = \int_t^{\omega''} V_{1,h}(g^{-1}(z), y)\mathrm{d}z,$$

and by assuming we can change the order of the integral signs, we get

$$
\begin{aligned}
E[\widetilde{\mathbb{S}}_{X,1}(t)] &= \int_{-\infty}^{\infty}\int_{t}^{\omega''} V_{1,h}(g^{-1}(z),y)\mathrm{d}z f_Y(y)\mathrm{d}y \\
&= \int_{t}^{\omega''} E[V_{1,h}(g^{-1}(z),Y)]\mathrm{d}z \\
&= \mathbb{S}_X(t) - \frac{h^2}{2}\int_{t}^{\omega''} b_1(z)\mathrm{d}z \int_{-\infty}^{\infty} y^2 K(y)\mathrm{d}y + o(h^2).
\end{aligned}
$$

It is easy to see $b_2(t) = -\int_t^{\omega''} b_1(z)\mathrm{d}z$, and then the formula of $Bias[\widetilde{\mathbb{S}}_{X,1}(t)]$ is done.

Before calculating $Var[\widetilde{\mathbb{S}}_{X,1}(t)]$, we must first note that

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{V}_{1,h}(x,y) &= \frac{1}{h}\int_{x}^{\infty} g'(z)K\left(\frac{z-y}{h}\right)\mathrm{d}z \\
&= \int_{\frac{x-y}{h}}^{\infty} g'(y+hz)K(z)\mathrm{d}z \\
&= g'(y)V\left(\frac{x-y}{h}\right) + hg''(y)\int_{\frac{x-y}{h}}^{\infty} zK(z)\mathrm{d}z + ... \\
&= g'(y)V\left(\frac{x-y}{h}\right) + o(h)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{V}_{1,h}(x,y) &= h\int_{\frac{x-y}{h}}^{\infty} g'(y+hz)V(z)\mathrm{d}z \\
&= hg'(y)\mathbb{V}\left(\frac{x-y}{h}\right) + h^2 g''(y)\int_{\frac{x-y}{h}}^{\infty} zV(z)\mathrm{d}z + ... \\
&= hg'(y)\mathbb{V}\left(\frac{x-y}{h}\right) + o(h).
\end{aligned}
$$

Now, we get

$$
\begin{aligned}
&E[\mathbb{V}_{1,h}^2(g^{-1}(t),g^{-1}(X_1))] \\
&\quad = 2\int_{-\infty}^{\infty} \mathbb{V}_{1,h}(g^{-1}(t),y)V\left(\frac{g^{-1}(t)-y}{h}\right)g'(y)S_Y(y)\mathrm{d}y + o(h) \\
&\quad = 2\int_{-\infty}^{\infty}\left[g'(y)V^2\left(\frac{g^{-1}(t)-y}{h}\right)\right. \\
&\qquad\quad \left. + \frac{1}{h}\mathbb{V}_{1,h}(g^{-1}(t),y)K\left(\frac{g^{-1}(t)-y}{h}\right)\right]a(y)\mathrm{d}y + o(h).
\end{aligned}
$$

Conducting integration by parts once again for the first term, we have

$$
\begin{aligned}
&2\int_{-\infty}^{\infty} g'(y)V^2\left(\frac{g^{-1}(t)-y}{h}\right)a(y)\mathrm{d}y\\
&\qquad= \frac{4}{h}\int_{-\infty}^{\infty} V\left(\frac{g^{-1}(t)-y}{h}\right)K\left(\frac{g^{-1}(t)-y}{h}\right)A(y)\mathrm{d}y\\
&\qquad= 4\int_{-\infty}^{\infty} A(g^{-1}(t)-hu)V(u)K(u)\mathrm{d}u\\
&\qquad= 2\bar{\mathbb{S}}_X(t) - 2hg'(g^{-1}(t))\mathbb{S}_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h).
\end{aligned}
$$

And the second term can be calculated with

$$
\begin{aligned}
&\frac{2}{h}\int_{-\infty}^{\infty} \mathbb{V}_{1,h}(g^{-1}(t),y)K\left(\frac{g^{-1}(t)-y}{h}\right)a(y)\mathrm{d}y\\
&\qquad= 2\int_{-\infty}^{\infty}\left[g'(y)\mathbb{V}\left(\frac{g^{-1}(t)-y}{h}\right)+o(1)\right]K\left(\frac{g^{-1}(t)-y}{h}\right)a(y)\mathrm{d}y\\
&\qquad= 2h\int_{-\infty}^{\infty} g'(g^{-1}(t)-hu)a(g^{-1}(t)-hu)\mathbb{V}(y)K(y)\mathrm{d}y + o(h)\\
&\qquad= 2h\int_{-\infty}^{\infty} [g'(g^{-1}(t))+o(1)][a(g^{-1}(t))+o(1)]\mathbb{V}(y)K(y)\mathrm{d}y + o(h)\\
&\qquad= 2hg'(g^{-1}(t))\mathbb{S}_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h).
\end{aligned}
$$

Thus, we get

$$
E[\mathbb{V}_{1,h}^2(g^{-1}(t),g^{-1}(X_1))] = 2\bar{\mathbb{S}}_X(t) + o(h),
$$

and this easily proves the $Var[\widetilde{\mathbb{S}}_{X,1}(t)]$ formula.

Before going into the calculation of the covariance, we have to take a look at

$$
\begin{aligned}
&E[\mathbb{V}_{1,h}(g^{-1}(t),g^{-1}(X_1))V_{1,h}(g^{-1}(t),g^{-1}(X_1))]\\
&\qquad= \int_{-\infty}^{\infty}\left[g'(y)V\left(\frac{g^{-1}(t)-y}{h}\right)V_{1,h}(g^{-1}(t),y)\right.\\
&\qquad\qquad\left.+\frac{1}{h}\mathbb{V}_{1,h}(g^{-1}(t),y)K\left(\frac{g^{-1}(t)-y}{h}\right)\right]S_Y(y)\mathrm{d}y.
\end{aligned}
$$

Once again we need to calculate them separately. The first term is

$$
\begin{aligned}
&\int_{-\infty}^{\infty} g'(y)V\left(\frac{g^{-1}(t)-y}{h}\right)V_{1,h}(g^{-1}(t),y)S_Y(y)\mathrm{d}y \\
&\qquad = \int_{-\infty}^{\infty}\left[\frac{1}{h^2}K\left(\frac{g^{-1}(t)-y}{h}\right)\int_{-\infty}^{y}K\left(\frac{g^{-1}(t)-z}{h}\right)\mathrm{d}z\right. \\
&\qquad\quad \left.+\frac{1}{h}V\left(\frac{g^{-1}(t)-y}{h}\right)K\left(\frac{g^{-1}(t)-y}{h}\right)\right]a(y)\mathrm{d}y \\
&\qquad = \int_{-\infty}^{\infty} a(g^{-1}(t)-hu)\left[K(u)\int_{u}^{\infty}K(v)\mathrm{d}v+V(u)K(u)\right]\mathrm{d}u \\
&\qquad = \mathbb{S}_X(t)-hg'g^{-1}(t))S_X(t)\int_{-\infty}^{\infty}V(u)W(u)\mathrm{d}u+o(h),
\end{aligned}
$$

while the second term is

$$
\begin{aligned}
&\frac{1}{h}\int_{-\infty}^{\infty}\mathbb{V}_{1,h}(g^{-1}(t),y)K\left(\frac{g^{-1}(t)-y}{h}\right)S_Y(y)\mathrm{d}y \\
&\qquad = h\int_{-\infty}^{\infty} g'(g^{-1}(t)-hu)S_Y(g^{-1}(t)-hu)\mathbb{V}(u)K(u)\mathrm{d}u+o(h) \\
&\qquad = h\int_{-\infty}^{\infty}[g'(g^{-1}(t))+o(1)][S_Y(g^{-1}(t))+o(1)]\mathbb{V}(u)K(u)\mathrm{d}u+o(h) \\
&\qquad = hg'(g^{-1}(t))S_X(t)\int_{-\infty}^{\infty}V(u)W(u)\mathrm{d}u+o(h),
\end{aligned}
$$

then we have

$$
E[\mathbb{V}_{1,h}(g^{-1}(t),g^{-1}(X_1))V_{1,h}(g^{-1}(t),g^{-1}(X_1))]=\mathbb{S}_X(t)+o(h).
$$

Hence, the covariance is

$$
\begin{aligned}
&Cov[\widetilde{\mathbb{S}}_{X,1}(t),\widetilde{S}_{X,1}(t)] \\
&\qquad = \frac{1}{n}[E\{\mathbb{V}_{1,h}(g^{-1}(t),g^{-1}(X_1))V_{1,h}(g^{-1}(t),g^{-1}(X_1))\} \\
&\qquad\quad -E\{\mathbb{V}_{1,h}(g^{-1}(t),g^{-1}(X_1))\}\{V_{1,h}(g^{-1}(t),g^{-1}(X_1))\}] \\
&\qquad = \frac{1}{n}[\mathbb{S}_X(t)-\mathbb{S}_X(t)S_X(t)+o(h)] \\
&\qquad \frac{1}{n}\mathbb{S}_X(t)F_X(t)+o\left(\frac{h}{n}\right).
\end{aligned}
$$

## Proof of theorem 5.1.3

The usual reasoning of *i.i.d.* random variables and the transformation property of expectation result in

$$
\begin{aligned}
E[\widetilde{S}_{X,2}(t)] &= \int_{-\infty}^{\infty} V\left(\frac{g^{-1}(t)-y}{h}\right) f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} K\left(\frac{g^{-1}(t)-y}{h}\right) S_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} S_Y(g^{-1}(t)-hu)K(u)\mathrm{d}u \\
&= S_X(t) - \frac{h^2}{2}b_1(t)\int_{-\infty}^{\infty} u^2K(u)\mathrm{d}u + o(h^2),
\end{aligned}
$$

and this gives us the $Bias[\widetilde{S}_{X,2}(t)]$. For the variance of $\widetilde{S}_{X,2}(t)$, first we calculate

$$
\begin{aligned}
E[V_{2,h}^2(g^{-1}(t), g^{-1}(X_1))] &= \frac{2}{h}\int_{-\infty}^{\infty} V\left(\frac{g^{-1}(t)-y}{h}\right) K\left(\frac{g^{-1}(t)-y}{h}\right) S_Y(y)\mathrm{d}y \\
&= 2\int_{-\infty}^{\infty} S_Y(g^{-1}(t)-hu)V(u)K(u)\mathrm{d}u \\
&= S_X(t) - hg'(g^{-1}(t))f_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h).
\end{aligned}
$$

The resulting variance is

$$
Var[\widetilde{S}_{X,2}(t)] = \frac{1}{n}S_X(t)F_X(t) - \frac{h}{n}g'(g^{-1}(t))f_X(t)\int_{-\infty}^{\infty} V(y)W(y)\mathrm{d}y + o\left(\frac{h}{n}\right).
$$

Next for $Bias[\widetilde{\mathbb{S}}_{X,2}(t)]$, utilizing similar reasoning as before, we get

$$
\begin{aligned}
E[\widetilde{\mathbb{S}}_{X,2}(t)] &= \int_{-\infty}^{\infty} \mathbb{V}_{2,h}(g^{-1}(t), y)f_Y(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} g'(y)V\left(\frac{g^{-1}(t)-y}{h}\right) S_Y(y)\mathrm{d}y \\
&= \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{g^{-1}(t)-y}{h}\right) a(y)\mathrm{d}y \\
&= \int_{-\infty}^{\infty} a(g^{-1}(t)-hu)K(u)\mathrm{d}u \\
&= \mathbb{S}_X(t) + \frac{h^2}{2}b_3(t)\int_{-\infty}^{\infty} u^2K(u)\mathrm{d}u + o(h^2),
\end{aligned}
$$

and this proves the bias part. For the variance, we need

$$\begin{aligned} E[\mathbb{V}_{2,h}^2(g^{-1}(t), g^{-1}(X_1))] &= 2\int_{-\infty}^{\infty} g'(y)\mathbb{V}_{2,h}(g^{-1}(t), y)V\left(\frac{g^{-1}(t)-y}{h}\right)S_Y(y)\mathrm{d}y \\ &= 2\int_{-\infty}^{\infty}\left[g'(y)V^2\left(\frac{g^{-1}(t)-y}{h}\right)\right. \\ &\qquad \left. +\frac{1}{h}\mathbb{V}_{2,h}(g^{-1}(t), y)K\left(\frac{g^{-1}(t)-y}{h}\right)\right]a(y)\mathrm{d}y. \end{aligned}$$

Once again using the integration by parts for the first term, we have

$$\begin{aligned} &2\int_{-\infty}^{\infty} g'(y)V^2\left(\frac{g^{-1}(t)-y}{h}\right)a(y)\mathrm{d}y \\ &\quad = \frac{4}{h}\int_{-\infty}^{\infty} V\left(\frac{g^{-1}(t)-y}{h}\right)K\left(\frac{g^{-1}(t)-y}{h}\right)A(y)\mathrm{d}y \\ &\quad = 4\int_{-\infty}^{\infty} A(g^{-1}(t)-hu)V(u)K(u)\mathrm{d}u \\ &\quad = 2\bar{\mathbb{S}}_X(t) - 2hg'(g^{-1}(t))\mathbb{S}_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h). \end{aligned}$$

The second term can be calculated in a similar way, which is

$$\begin{aligned} &\frac{2}{h}\int_{-\infty}^{\infty}\mathbb{V}_{2,h}(g^{-1}(t), y)K\left(\frac{g^{-1}(t)-y}{h}\right)a(y)\mathrm{d}y \\ &\quad = 2h\int_{-\infty}^{\infty} a(g^{-1}(t)-hu)\int_{u}^{\infty} g'(g^{-1}(t)-hv)V(v)\mathrm{d}vK(u)\mathrm{d}u \\ &\quad = 2h\int_{-\infty}^{\infty}[a(g^{-1}(t))+o(1)]\int_{u}^{\infty}[g'(g^{-1}(t))+o(1)]V(v)\mathrm{d}vK(u)\mathrm{d}u \\ &\quad = 2hg'(g^{-1}(t))\mathbb{S}_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h). \end{aligned}$$

Hence, we get

$$E[\mathbb{V}_{2,h}^2(g^{-1}(t), g^{-1}(X_1))] = 2\bar{\mathbb{S}}_X(t) + o(h),$$

proving the formula of $Var[\widetilde{\mathbb{S}}_{X,2}(t)]$.

Before moving onto the calculation of the covariance, we have to take a look at

$$\begin{aligned} &E[\mathbb{V}_{2,h}(g^{-1}(t), g^{-1}(X_1))V_{2,h}(g^{-1}(t), g^{-1}(X_1))] \\ &\quad = \int_{-\infty}^{\infty}\left[g'(y)V^2\left(\frac{g^{-1}(t)-y}{h}\right) + \frac{1}{h}\mathbb{V}_{2,h}(g^{-1}(t), y)K\left(\frac{g^{-1}(t)-y}{h}\right)\right]S_Y(y)\mathrm{d}y. \end{aligned}$$

Once again we need to calculate them separately. The first term is

$$\int_{-\infty}^{\infty} g'(y)V^2\left(\frac{g^{-1}(t)-y}{h}\right)S_Y(y)\mathrm{d}y$$
$$= \frac{2}{h}\int_{-\infty}^{\infty} V\left(\frac{g^{-1}(t)-y}{h}\right)K\left(\frac{g^{-1}(t)-y}{h}\right)a(y)\mathrm{d}y$$
$$= 2\int_{-\infty}^{\infty} a(g^{-1}(t)-hu)V(y)K(y)\mathrm{d}u$$
$$= \mathbb{S}_X(t) - hg'(g^{-1}(t))S_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h),$$

while the second term is

$$\frac{1}{h}\int_{-\infty}^{\infty} \mathbb{V}_{2,h}(g^{-1}(t),y)K\left(\frac{g^{-1}(t)-y}{h}\right)S_Y(y)\mathrm{d}y$$
$$= h\int_{-\infty}^{\infty} S_Y(g^{-1}(t)-hu)\int_{u}^{\infty} g'(g^{-1}(t)-hv)V(v)\mathrm{d}vK(u)\mathrm{d}u$$
$$= h\int_{-\infty}^{\infty} [S_Y(g^{-1}(t))+o(1)][g'(g^{-1}(t))\mathbb{V}(u)+o(1)]K(u)\mathrm{d}u$$
$$= hg'(g^{-1}(t))S_X(t)\int_{-\infty}^{\infty} V(u)W(u)\mathrm{d}u + o(h),$$

and the result is

$$E[\mathbb{V}_{2,h}(g^{-1}(t),g^{-1}(X_1))V_{2,h}(g^{-1}(t),g^{-1}(X_1))] = \mathbb{S}_X(t) + o(h).$$

Hence, the covariance is

$$Cov[\widetilde{\mathbb{S}}_{X,2}(t),\widetilde{S}_{X,2}(t)] = \frac{1}{n}\mathbb{S}_X(t)F_X(t) + o\left(\frac{h}{n}\right).$$

# Proof of theorem 5.2.1

As for a fixed $t$ we have that $\widetilde{S}_{X,1}(t)$ and $\widetilde{\mathbb{S}}_{X,1}(t)$ are consistent estimators for $S_X(t)$ and $\mathbb{S}_X(t)$, respectively, then

$$\begin{aligned}
\widetilde{m}_{X,1}(t) - m_X(t) &= \frac{\widetilde{\mathbb{S}}_{X,1}(t) - \widetilde{S}_{X,1}(t)m_X(t)}{S_X(t)}\left[1 + \frac{S_X(t)-\widetilde{S}_{X,1}(t)}{\widetilde{S}_{X,1}(t)}\right] \\
&= \frac{\widetilde{\mathbb{S}}_{X,1}(t) - \widetilde{S}_{X,1}(t)m_X(t)}{S_X(t)}[1+o_p(1)].
\end{aligned}$$

Thus, using theorem 5.1.1, we get

$$\begin{aligned}
Bias[\widetilde{m}_{X,1}(t)] &= \frac{1}{S_X(t)}[E\{\widetilde{\mathbb{S}}_{X,1}(t)\} - m_X(t)E\{\widetilde{S}_{X,1}(t)\}] \\
&= \frac{h^2}{2S_X(t)}[b_2(t) + m_X(t)b_1(t)] \int_{-\infty}^{\infty} y^2 K(y)\mathrm{d}y + o(h^2).
\end{aligned}$$

The same argument easily proves the formula of $Bias[\widetilde{m}_{X,2}(t)]$.

Using a similar method, for $i = 1, 2$, we have

$$\begin{aligned}
&Var[\widetilde{m}_{X,i}(t)] \\
&\quad = Var[\widetilde{m}_{X,i}(t) - m_X(t)] \\
&\quad = Var\left[\frac{\widetilde{\mathbb{S}}_{X,1}(t) - \widetilde{S}_{X,1}(t)m_X(t)}{S_X(t)}\right] \\
&\quad = \frac{1}{S_X^2(t)}[Var\{\widetilde{\mathbb{S}}_{X,i}(t)\} + m_X^2(t)Var\{\widetilde{S}_{X,i}(t)\} \\
&\qquad -2m_X(t)Cov\{\widetilde{\mathbb{S}}_{X,i}, \widetilde{S}_{X,i}(t)\}] \\
&\quad = \frac{1}{n}\frac{b_4(t)}{S_X^2(t)} - \frac{h}{n}\frac{b_5(t)}{S_X^2(t)} \int_{-\infty}^{\infty} V(y)W(y)\mathrm{d}y + o\left(\frac{h}{n}\right).
\end{aligned}$$

## Proof of theorem 5.2.2

Because the proof of the case $i = 1$ is similar, we will only explain the case of $i = 2$ in detail. First, for some $\delta > 0$, using Hölder and $c_r$ inequalities

$$\begin{aligned}
&E[|V_{2,h}(g^{-1}(t), g^{-1}(X_1)) - E\{V_{2,h}(g^{-1}(t), g^{-1}(X_1))\}|^{2+\delta}] \\
&\qquad \leq 2^{2+\delta}E[|V_{2,h}(g^{-1}(t), g^{-1}(X_1))|^{2+\delta}].
\end{aligned}$$

But, since $0 \leq V_{2,h}(x, y) \leq 1$ for any $x, y \in \mathbb{R}$, then

$$E[|V_{2,h}(g^{-1}(t), g^{-1}(X_1)) - E\{V_{2,h}(g^{-1}(t), g^{-1}(X_1))\}|^{2+\delta}] \leq 2^{2+\delta} < \infty,$$

and because $Var[V_{2,h}(g^{-1}(t), g^{-1}(X_1))] = O(1)$, we get

$$\frac{E[|V_{2,h}(g^{-1}(t), g^{-1}(X_1)) - E\{V_{2,h}(g^{-1}(t), g^{-1}(X_1))\}|^{2+\delta}]}{n^{\delta/2}[Var\{V_{2,h}(g^{-1}(t), g^{-1}(X_1))\}]^{1+\delta/2}} \to 0$$

when $n \to \infty$. Hence, by Loeve (1963), and with the fact $\widetilde{S}_{X,2}(t) \to_p S_X(t)$, we can conclude that

$$\frac{\widetilde{S}_{X,2}(t) - S_X(t)}{\sqrt{Var[\widetilde{S}_{X,2}(t)]}} \to_D N(0, 1).$$

Next, with a similar reasoning as before, we have

$$\begin{aligned}E[|\mathbb{V}_{2,h}(g^{-1}(t), g^{-1}(X_1)) - E\{\mathbb{V}_{2,h}(g^{-1}(t), g^{-1}(X_1))\}|^{2+\delta}]\\ \leq 2^{2+\delta} E[|\mathbb{V}_{2,h}(g^{-1}(t), g^{-1}(X_1))|^{2+\delta}],\end{aligned}$$

which, by the same inequalities, results in

$$\begin{aligned}&E[|\mathbb{V}_{2,h}(g^{-1}(t), g^{-1}(X_1)) - E\{\mathbb{V}_{2,h}(g^{-1}(t), g^{-1}(X_1))\}|^{2+\delta}]\\ &\leq 2^{2+\delta} E\left[\left|\int_{-\infty}^{g^{-1}(X_1)} g'(z) V\left(\frac{g^{-1}(t)-z}{h}\right) \mathrm{d}z\right|^{2+\delta}\right]\\ &\leq 2^{2+\delta} E\left[\left|\int_{-\infty}^{g^{-1}(X_1)} g'(z) \mathrm{d}z\right|^{2+\delta}\right]\\ &\leq 2^{2+\delta} E(X_1^{2+\delta})\\ &< \infty.\end{aligned}$$

Therefore, with the same argument, we get

$$\frac{\widetilde{\mathbb{S}}_{X,2}(t) - \mathbb{S}_X(t)}{\sqrt{Var[\widetilde{\mathbb{S}}_{X,2}(t)]}} \to_D N(0,1).$$

At last, by Slutsky's Theorem for rational function, the theorem is proven.

**Remark .0.5.** *Since we only assume the existence of $E(X^3)$, then we should choose $\delta \leq 1$ in this proof.*

# Proof of theorem 5.2.3

Nadaraya (1964) guarantees that $\sup_{t\in\mathbb{R}} |\widehat{S}_Y(t) - S_Y(t)| \to_{a.s.} 0$, which implies

$$\sup_{t\in\Omega} |\widehat{S}_Y(g^{-1}(t)) - S_Y(g^{-1}(t))| \to_{a.s.} 0.$$

However, because $S_Y(g^{-1}(t)) = S_X(t)$, and it is clear that $\widehat{S}_Y(g^{-1}(t)) = \widetilde{S}_{X,1}(t)$, then $\sup_{t\in\Omega} |\widetilde{S}_{X,1}(t) - S_X(t)| \to_{a.s.} 0$ holds.

Next, since $\mathbb{S}_X(t) \geq 0$ is bounded above with

$$\sup_{t\in\Omega} \mathbb{S}_X(t) = \lim_{t\to\omega'+} \mathbb{S}_X(t) = E(X) - \omega',$$

then $a(g^{-1}(t)) = \mathbb{S}_X(t)$ is bounded on $\Omega$. Furthermore,

$$
\begin{aligned}
\widehat{a}(g^{-1}(t)) &= \frac{1}{n}\sum_{i=1}^{n}\int_{g^{-1}(t)}^{\infty} g'(z)V\left(\frac{z-g^{-1}(X_i)}{h}\right)\mathrm{d}z \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{t}^{\omega''} V\left(\frac{g^{-1}(z)-g^{-1}(X_i)}{h}\right)\mathrm{d}z \\
&= \widetilde{\mathbb{S}}_{X,1}(t) \\
&> 0
\end{aligned}
$$

is also bounded above almost surely with

$$
\sup_{t\in\Omega}\widetilde{\mathbb{S}}_{X,1}(t) = \lim_{t\to\omega'+}\widetilde{\mathbb{S}}_{X,1}(t) = \bar{X} - \omega' + O_p(h^2).
$$

Thus, lemma .0.4 implies $\sup_{t\in\Omega}|\widehat{a}(g^{-1}(t)) - a(g^{-1}(t))| \to_{a.s.} 0$, which is equivalent to $\sup_{t\in\Omega}|\widetilde{\mathbb{S}}_{X,1}(t) - \mathbb{S}_X(t)| \to_{a.s.} 0$. As a conclusion, $\sup_{t\in\Omega}|\widetilde{m}_{X,1}(t) - m_X(t)| \to_{a.s.} 0$ holds. The proof for the case of $i = 2$ is similar.

# Proof of theorem 5.2.4

Because, for $i = 1, 2$,

$$
\lim_{t\to\omega'+}\widetilde{m}_{X,i}(t) = \frac{\lim_{t\to\omega'+}\widetilde{\mathbb{S}}_{X,i}(t)}{\lim_{t\to\omega'+}\widetilde{S}_{X,i}(t)},
$$

we only need to see the limit behaviour of each estimators of the survival function and the cumulative survival function. First, we have

$$
\begin{aligned}
\lim_{t\to\omega'+}\widetilde{S}_{X,1}(t) &= \frac{1}{nh}\sum_{i=1}^{n}\lim_{t\to\omega'+}\int_{g^{-1}(t)}^{\infty} K\left(\frac{z-g^{-1}(X_i)}{h}\right)\mathrm{d}z \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} K(u)\mathrm{d}u \\
&= 1.
\end{aligned}
$$

For $\lim_{t\to\omega'+}\widetilde{\mathbb{S}}_{X,1}(t)$, the use of the integration by subsitution and by parts means

$$
\begin{aligned}
\lim_{t\to\omega'+}\widetilde{\mathbb{S}}_{X,1}(t) &= \frac{1}{n}\sum_{i=1}^{n}\lim_{t\to\omega'+}\int_{g^{-1}(t)}^{\infty} g'(z)V\left(\frac{z-g^{-1}(X_i)}{h}\right)\mathrm{d}z \\
&= -\omega' + \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty} g(g^{-1}(X_i)+hu)K(u)\mathrm{d}u \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}[g(g^{-1}(X_i)) + hg'(g^{-1}(X_i))u + O_p(h^2)]K(u)\mathrm{d}u - \omega' \\
&= \bar{X} - \omega' + O_p(h^2).
\end{aligned}
$$

On the other hand, the fact $\lim_{x\to-\infty} V(x) = 1$ results in

$$\lim_{t\to\omega'^+} \widetilde{S}_{X,2}(t) = \frac{1}{n}\sum_{i=1}^{n} \lim_{t\to\omega'^+} V\left(\frac{g^{-1}(t) - g^{-1}(X_i)}{h}\right) = 1,$$

and

$$\begin{aligned}
\lim_{t\to\omega'^+} \widetilde{\mathbb{S}}_{X,2}(t) &= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{g^{-1}(X_i)} g'(z) \lim_{t\to\omega'^+} V\left(\frac{g^{-1}(t) - z}{h}\right) \mathrm{d}z \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{g^{-1}(X_i)} g'(z)\mathrm{d}z \\
&= \bar{X} - \omega'.
\end{aligned}$$

Then, the theorem is proven.