2020

ASYMPTOTIC NORMALITY OF THE DECONVOLUTION KERNEL DENSITY ESTIMATORS BASED ON INDEPENDENT AS WELL AS STRONG MIXING RIGHT CENSORED DATA

Wenqing Zhu

ASYMPTOTIC NORMALITY OF THE DECONVOLUTION KERNEL DENSITY ESTIMATORS BASED ON INDEPENDENT AS WELL AS STRONG MIXING RIGHT CENSORED DATA

by

WENQING ZHU

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2020

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Dr. Shan Sun-Mitchell for her guidance, continuous support and motivation through each stage of my graduate study during these past four years. She has provided insightful discussions about my research and her valuable advice is essential to the completion of this work.

I would like to extend my sincere gratitude to Dr. Andrzej Korzeniowski, Dr. Suvra Pal, Dr. Li Wang and Dr. Jonghyun Yun. From Dr. Andrzej Korzeniowski's classes, I benefited a lot and improved my skills in doing proofs. I am very appreciative of Dr. Suvra Pal for his introduction to survival analysis, which led me to involve this part to my research. I'm grateful to Dr. Li Wang for her help with my programming part, she is always so nice and easygoing. I'm thankful to Dr. Jonghyun Yun, who taught me almost all statistics courses, ranging from the very basic one "Mathematical Statistics I" to the more advanced level class "Data Mining".

I wish to thank the Department of Mathematics, for the financial support in my doctoral program. I would also like to acknowledge all the professors, staff, and graduate students of our department for their help and inspirations. I thank the Chair, Dr. Jiangzhong Su, for his kindness and advice throughout my doctoral study at UTA. I thank Crystal Mackey, Anthony Thomas, Benie Justine N'Gozan, Omer Mogultay, and Ce Bian. It is your friendship and company that have made my days here so enjoyable and memorable. I would also like to thank my mentor Jie Tang, who is a role model of mine and has provided invaluable career suggestions and always inspires me.

To my father Songzhu Zhu, my mother Yangmei Luo, my uncle Zhiyong Yang, and my aunt Diane, thank you for your love, support, and encouragement which have enriched my soul and inspired me to pursue and complete my graduate study. A special thank you goes to my best friend Zihe Zeng for being there for me throughout the entire doctoral program. I would not be where I am today without them.

July 23, 2020

ABSTRACT

ASYMPTOTIC NORMALITY OF THE DECONVOLUTION KERNEL DENSITY ESTIMATORS BASED ON INDEPENDENT AS WELL AS STRONG MIXING RIGHT CENSORED DATA

Wenqing Zhu, Ph.D.

The University of Texas at Arlington, 2020

Supervising Professor: Shan Sun-Mitchell

We consider estimation of a density when observed lifetime from the convolution model contaminated by additive measurement errors. A kernel type deconvolution density estimator of the unknown distribution based on right censored data is proposed by using the Inverse-Probability-of-Censoring Weighted Average. Further, we discuss the asymptotic normality of the deconvolution kernel density estimators for independent and strong mixing vectors when the error distribution function is either ordinary smooth or supersmooth.

Our method is applied to the study conducted by UTSW medical center. The research team at UTSW collected the data of women who underwent cystoscopy fulguration for recurrent urinary tract infection (UTI) from 2004-2016. Using the estimators and the asymptotic distributions of the estimators, we estimate the survival probability of the time from infection to recurrent UTI.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In various fields of biological, engineering and social sciences, our event of interest is an undesirable one. For example, death of a patient (in biomedical science) or failure of a component in a machine (reliability). The time until the occurrence of the event is duration data or time to event data, which we know as the survival data. Survival data are typically observed in an incomplete way, due to the presence of a number of events which potentially censor the event of interest. We are going to focus on the right censored data-respondents left the survey or haven't achieved the milestone at the time of survey, which is the more prevalent situation.

There's a drawback when we collect the survival data, people may not accurately report when event occurs. If the timing of events happened in the distant past, people may tend up or down the year or time since the event occurred [4]. Under this consideration, variables of interest are not directly observable but only through some error contamination.

Suppose we have n observations $x_1, \ldots, x_n$ i.i.d from the convolution model

$$X = Z + E \tag{1.1}$$

We are interested in estimating the distribution of Z that cannot be directly observed. For example [7], in AIDs study, the survival data X we collect may be the time from some starting point to the time that symptoms appear, the error term E could be the time from some starting point to the time that infection occurs. Z is the incubation period (the time from the occurrence of infection to the time of symptoms).

1.1 Model Formulation

We consider the random censorship model from right, using $X_1, X_2, \ldots, X_n$ as observed lifetime from an unknown common distribution function $f_X(\cdot)$. X_i is censored from the right by the censoring time C_i, C_i is from a distribution $G(\cdot)$. Assume we have random censoring, X_i is independent of C_i. And both of them are non-negative random variables.

Note that we are only able to observe $W_i = \min(X_i, C_i)$ and $\delta_i = \mathrm{I}(X_i \leq C_i)$, δ_i is the indicator for the event that W_i is uncensored. Let H be the distribution function of W_i, then $H = 1 - (1 - f_X)(1 - G)$. We assume censoring indicators $\delta_1, \cdots, \delta_n$ follow a Bernoulli distribution: for $i = 1, \cdots, n$

$$P(\delta_i = 1) = p, \qquad P(\delta_i = 0) = 1 - p \tag{1.2}$$

where $0 < p < 1$. Therefore the joint distribution of (W_i, δ_i) is:

$$\begin{aligned} f_{W,\delta}(w, \delta) &= P_{W|\delta}(w \mid \delta = 1)\, P(\delta = 1) + P_{W|\delta}(w \mid \delta = 0)\, P(\delta = 0) \\ &= f_{X|\delta=1}(w)\, p + g_{C|\delta=0}(w)\, (1 - p) \end{aligned} \tag{1.3}$$

Z_i is contaminated in model $X_i = Z_i + E_i$, where X_i is an observable survival time with an unknown density function f_X in $W_i = \min(X_i, C_i)$, E_i is the error term. We assume E_i is a random variable that is independent of Z_i and δ_i, and E_i has known density f_E. We wish to estimate the unknown density f_Z based on observations of X_i.

To estimate f_Z, note that we have the following convolution $f_X = f_Z * f_E$, and

$$\varphi_{f_Z}(t) = \frac{\varphi_{f_X}(t)}{\varphi_{f_E}(t)} \tag{1.4}$$

where φ_g is the characteristic function of g. By Fourier inversion theorem, we write the estimator of f_Z, $\hat{f}_Z$ as

$$\hat{f}_Z(z) = \frac{1}{2\pi} \int e^{-\mathrm{i}tz} \frac{\varphi_{\hat{f}_X}(t)}{\varphi_{f_E}(t)} dt \tag{1.5}$$

where $\varphi_{\hat{f}_X}(t) = \int \exp(\mathrm{i}tx) \cdot \hat{f}_X(x)dx$, and $\hat{f}_X$ will be given by (1.7) and (1.10) as two methods.

A kernel density estimator of f_X can be motivated through the Kaplan-Meier estimator of F_X [12]

$$\hat{F}_{KM}(x) = 1 - \hat{S}_{KM}(x) = \begin{cases} 0, & 0 \le x \le W_{(1)} \\ 1 - \prod_{i=1}^{j-1} \left(\frac{n-i}{n-i+1}\right)^{\delta_{[i]}}, & W_{(j-1)} < x \le W_{(j)} \\ & j = 2, \ldots, n \\ 1, & x > W_{(n)} \end{cases} \tag{1.6}$$

where $(W_{(i)}, \delta_{[i]})$, $i = 1, \ldots, n$, denotes the (W_i, δ_i) ordered with respect to the W_i's.

The kernel estimator of f_X induced by Kaplan-Meier estimator $\hat{F}_{KM}$ of F_X is then

$$\begin{aligned} \hat{f}_X^{KM}(x) &= \frac{1}{h} \int K\left(\frac{x-y}{h}\right) d\hat{F}_{KM}(y) \\ &= \frac{1}{h} \sum_{j=1}^{n} K\left(\frac{x - W_{(j)}}{h}\right) s_j \end{aligned} \tag{1.7}$$

where s_j is the size of the jump of $\hat{F}_{KM}$ at $W_{(j)}$, h is a positive number, usually called as the bandwidth or window width, and kernel K is a symmetric function satisfying

$$\int K(t)dt = 1, \qquad \int tK(t)dt = 0, \qquad and \int t^2 K(t)dt < \infty$$

Thus $\hat{f}_Z$ is described as follows: from (1.5) and (1.7), Chakrabarty [4] showed that the deconvolution kernel density estimator of the target density $\hat{f}_Z$ can be written as

$$\hat{f}_Z^{KM}(z) = \frac{1}{h} \sum_{j=1}^{n} K^Z\left(\frac{z - W_{(j)}}{h}\right) s_j \tag{1.8}$$

where $K^Z\left[(z - W_{(j)})/h\right] = \int \exp(-\mathrm{i}y(z - W_{(j)})/h) \cdot \varphi_K(y)/\varphi_{f_E}(y/h)dy\,/(2\pi)$, and φ_g denotes the characteristic function of a density g.

In this research, we discuss an alternative way of estimating f_Z. Satten and Datta [19] showed that the Kaplan-Meier estimator $\hat{F}_{KM}$ in (1.6) is equivalent to

$$\hat{F}_{IP}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{I(W_i \leq x)\cdot\delta_i}{1-G(W_i)} \tag{1.9}$$

Therefore, application of the inverse-probability-of censoring weighted idea to estimate f_X leads to:

$$\begin{aligned}\hat{f}_X^{IP}(x) &= \frac{1}{h}\int K\left(\frac{x-y}{h}\right)d\hat{F}_{IP}(y)\\ &= \frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{W_j-x}{h}\right)\frac{\delta_j}{1-G(W_j)} \end{aligned}\tag{1.10}$$

By using the symmetry of K, the characteristic function of $\hat{f}_X^{IP}$ is

$$\begin{aligned}\varphi_{\hat{f}_X^{IP}}(t) &= \int e^{\mathrm{i}tx}\hat{f}_X^{IP}(x)dx\\ &= \int e^{\mathrm{i}tx}\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{W_j-x}{h}\right)\frac{\delta_j}{1-G(W_j)}dx\\ &= \frac{1}{nh}\sum_{j=1}^{n}\int e^{\mathrm{i}tx}K\left(\frac{W_j-x}{h}\right)\frac{\delta_j}{1-G(W_j)}dx\\ &= \frac{1}{nh}\sum_{j=1}^{n}\frac{\delta_j}{1-G(W_j)}\int e^{\mathrm{i}tx}K\left(\frac{x-W_j}{h}\right)dx\\ &= \frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j}{1-G(W_j)}\int e^{\mathrm{i}t(hu+W_j)}K(u)du\\ &= \frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j}{1-G(W_j)}e^{\mathrm{i}tW_j}\varphi_K(ht) \end{aligned}\tag{1.11}$$

Plug $\hat{f}_X^{IP}(x)$ in (1.5), we obtain the deconvolution kernel density estimator of f_Z as

$$\begin{aligned}\hat{f}_Z^{IP}(z) &= \frac{1}{2\pi}\int e^{-\mathrm{i}tz}\frac{\varphi_{\hat{f}_X^{IP}}(t)}{\varphi_{f_E}(t)}dt\\ &= \frac{1}{2\pi}\int e^{-\mathrm{i}tz}\frac{\frac{1}{n}\sum_{j=1}^{n}\frac{\delta_j}{1-G(W_j)}e^{\mathrm{i}tW_j}\varphi_K(ht)}{\varphi_{f_E}(t)}dt\\ &= \frac{1}{2\pi n}\sum_{j=1}^{n}\int\frac{e^{\mathrm{i}t(W_j-z)}\varphi_K(ht)}{\varphi_{f_E}(t)}dt\,\frac{\delta_j}{1-G(W_j)}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{nh}\sum_{j=1}^{n}\frac{1}{2\pi}\int \frac{e^{-\mathrm{i}y\frac{z-W_j}{h}}\varphi_K(y)}{\varphi_{f_E}\left(\frac{y}{h}\right)}dy\,\frac{\delta_j}{1-G(W_j)} \\
&:= \frac{1}{nh}\sum_{j=1}^{n} K^E\left(\frac{z-W_j}{h}\right)\frac{\delta_j}{1-G(W_j)} \qquad (1.12)
\end{aligned}
$$

where $K^E\left[(z-W_j)/h\right] = \int \exp(-\mathrm{i}y(z-W_j)/h)\cdot\varphi_K(y)/\varphi_{f_E}(y/h)dy\,/(2\pi)$.

CHAPTER 2

APPLICATION OF THE ESTIMATE

2.1 RUTI Data

A urinary tract infection (UTI) is an infection of any part of the urinary system, UTIs occur mainly in women. Out of 100 women, 50 will have a UTI at some point in their life according to the United States Department of Health and Human Services (HHS) [1].

UT Southwestern research team collected a group of 95 female patients information from 2004 to 2016, who underwent cystoscopy with fulguration (CF). Note when we cleaned the data, we didn't count the first 6 months following the procedure, as fulgurating the bladder can sometimes cause irritation which is confused for a UTI. Thus we set the starting point of recording recurrent UTI (RUTI) as 6 months after the CF, RUTI time we collected is the time from this starting point to the time RUTI first appear. Considering the rate of recurrence following an initial UTI is high, researchers defined RUTI as "when a patient experienced two UTIs within 6 months or three UTIs within 12 months".

Le [14] found out Smoking (the patient who is a smoker), CoitalPre (using Coital prophylaxis before CF procedure), FQ (having Fluoroquinolone resistance), ESBL (Extended spectrum beta lactamase resistance) are the strongest risk factors which likely increase the RUTI rate of UTI patients by using Cox Proportional-hazards model. Considering Cox model doesn't take the cure probability of the patients into account, he extended the Cox model and applied it to the mixture cure model to identify the effect of significant factors on the cure probability of the patients. First year success (no UTI in the first year after the first follow up appointment), EC (E. Coli pathogen), MultBugs (at least two bacterial species grew in the urine culture), ESBL (Extended spectrum beta lactamase resistance) are the strongest risk factors that affect the cure rate of UTI patients by using the mixture cure model.

In this chapter, we are interested in investigating the time from infection to the occurrence of RUTI, that is the time from exposure to the causative agent to when RUTI first appear. Knowing the time from infection to the occurrence of RUTI can provide important information during outbreak, including when the patient will have the symptoms of RUTI and is most likely to spread it. Because the time to RUTI reflects the growth of pathogen, copying rate and toxin discharge, time from infection to the occurrence of RUTI will be helpful to provide some clues for identifying the causative agent [18]. Time from infection to the occurrence of RUTI also offers insights for the prognosis, like the expected duration of the disease and the severity of it, which will be beneficial for doctors to develop the treatments.

In the study, denote the time from six months after CF to the time first RUTI appear by the random variable X. The data we use contains the censoring time, the percent of censoring in data is 61%. Censoring time here means the time when patients left the survey before the survey ends, or if the patient doesn't have any RUTIs during the survey, then censoring time will be the end time of the survey. C is a random variable representing the time from the six months after CF to a censoring time.

2.1.1 Survival probability of the time from infection to the occurrence of RUTI

To estimate the survival probability of Z, recall that we are only able to observe the survival time $W_i = \min(X_i, C_i)$ and $\delta_i = \mathrm{I}(X_i \leq C_i)$. $\{X_i, i = 1, \cdots, n\}$ denotes the uncensored lifetime from an unknown common distribution function $f_X(\cdot)$. X_i

is censored from the right by the censoring time C_i, where C_i is from an unknown distribution $G(\cdot)$. Assume X_i is independent of C_i, and both of them are non-negative random variables.

We now apply the deconvolution density estimators $\hat{f}_Z^{IP}$ in (1.12), $\hat{f}_Z^{KM}$ in (1.8) to the RUTI data, where the goal is to estimate the survival probability $S(t) = \int_t^{\infty} f_Z(u)du$ of the time from infection to the occurrence of RUTI.

Recall our model (1.1), $Z = X - E$, Z is the time from infection to the occurrence of RUTI, E is independent of Z, and we use sample variance s_X^2 of X to estimate σ_X^2. The variance σ_E^2 of the error E is chosen so that the reliability ratio [10]:

$$r = \frac{\text{VarZ}}{\sigma_E^2 + \text{VarZ}} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2} \approx \frac{s_X^2 - \sigma_E^2}{s_X^2} = 0.7 \tag{2.1}$$

Fan [7] mentioned that the convergence rate of estimators in the presence of supersmooth errors is slower than the case in the presence of ordinary smooth errors, and considering E should have non-negative values, we assume E has a truncated double exponential distribution, which belongs to the ordinary smooth case:

$$f_E(u) = \frac{1}{\sigma_E} e^{\frac{-u}{\sigma_E}} \tag{2.2}$$

where $u \geq 0$. The characteristic function of E is

$$\begin{aligned}
\varphi_{f_E}(t) &= \int_0^{\infty} e^{\mathrm{i}tu} f_E(u)\, du \\
&= \frac{1}{\sigma_E} \int_0^{\infty} e^{\mathrm{i}tu} e^{\frac{-u}{\sigma_E}}\, du \\
&= \frac{1}{\sigma_E} \int_0^{\infty} \cos(tu)\, e^{\frac{-u}{\sigma_E}}\, du + \mathrm{i}\, \frac{1}{\sigma_E} \int_0^{\infty} \sin(tu)\, e^{\frac{-u}{\sigma_E}}\, du \\
&= \frac{1}{\sigma_E} \frac{\frac{1}{\sigma_E}}{\left(\frac{1}{\sigma_E}\right)^2 + t^2} + \mathrm{i}\, \frac{1}{\sigma_E} \frac{t}{\left(\frac{1}{\sigma_E}\right)^2 + t^2} \\
&= \frac{1}{1 - \mathrm{i}\, \sigma_E\, t}
\end{aligned} \tag{2.3}$$

From (1.12), we get

$$
\begin{aligned}
K^E(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}yx}\,\frac{\varphi_K(y)}{\varphi_{f_E}(\frac{y}{h})}dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}yx}\,\varphi_K(y)\left(1-\mathrm{i}\,\sigma_E\,\frac{y}{h}\right)dy \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}yx}\,\varphi_K(y)dy + \frac{\sigma_E}{h}\,\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\mathrm{i}yx}\,(-\mathrm{i}y)\,\varphi_K(y)dy \\
&= K(x) + \frac{\sigma_E}{h}\,K'(x) \qquad (2.4)
\end{aligned}
$$

Since the distribution G of C is unknown, one way to overcome this is to use the smooth kernel density estimator, therefore, G can be estimated as $\hat{G}(x) = \sum_{i=1}^{l} K_C[(x-C_i)/h]/(lh)$, where l is the count number of $\delta_i = 0$ $(i = 1, \cdots, n)$, we choose K_C as the normal kernel: $K_C(x) = \phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$. And we set the bandwidth $h = 0.4$. For kernel function K in (2.4), we use 3 different kernel functions as in Table 2.1.

Table 2.1: Kernel functions and their characteristic functions.

Name	**Kernel** $K(x)$	**Characteristic Function** $\varphi_K(t)$
de la Vallée-Poussin (K_1)	$\frac{1-\cos x}{\pi x^2}$	$(1-\lvert t\rvert)I_{[-1,1]}(t)$
Triweight Ft (K_2)	$\frac{48(1-15x^{-2})\cos x}{\pi x^4} - \frac{144(2-5x^{-2})\sin x}{\pi x^5}$	$(1-t^2)^3 I_{[-1,1]}(t)$
Tricube Ft (K_3)	$\frac{648(-3x^4+90x^2-560)\sin x}{x^9\pi} + \frac{162(x^6-80x^4+1120x^2-2240)\cos x}{x^{10}\pi} + \frac{18(20160-x^6)}{x^{10}\pi}$	$(1-\lvert t\rvert^3)^3 I_{[-1,1]}(t)$

When we graph the survival probability estimator, we set the grid of z to be equispaced, where the distance between two neighbour points is equal to 1. Furthermore, we need to consider rescaling the estimated probability density function $\hat{f}_Z$ by using $\hat{f}_Z(z_i)/\left(\sum_{i=1}^{N}\hat{f}_Z(z_i)\cdot dz\right)$ so that it integrates to 1 after the negative ones have been dropped. The estimated curves, shown in below (Figure 2.1 to 2.3), used 3 different kernels in Table 2.1. Dashed line KMfdec in each plot is the estimated survival probability $\hat{S}_Z^{KM}$ of Z by using the rescaled $\hat{f}_Z^{KM}(z)$, solid line IPfdec is $\hat{S}_Z^{IP}$ by using the rescaled $\hat{f}_Z^{IP}(z)$.

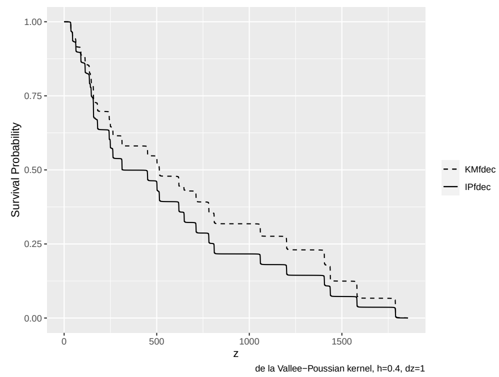

Figure 2.1: Estimated Survival Probability of Z by using K_1

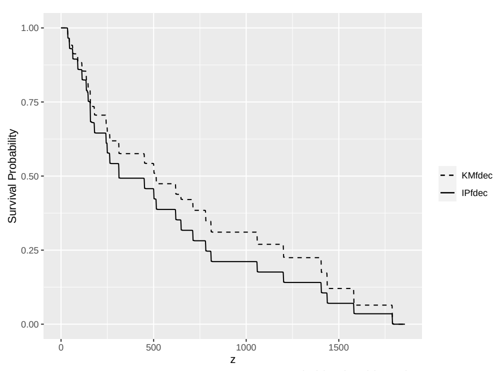

Figure 2.2: Estimated Survival Probability of Z by using K_2

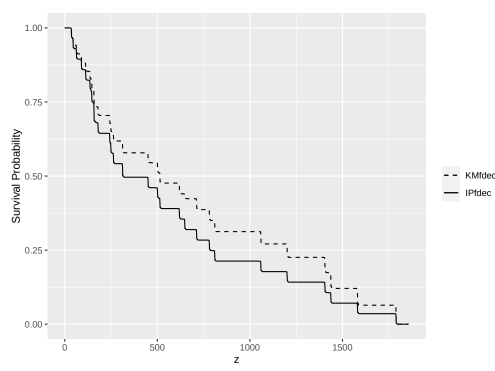

Figure 2.3: Estimated Survival Probability of Z by using K_3

In these plots, we see that there's no big difference between 3 figures based on 3 kernels we choose (K_1, K_2, K_3), which implies that the choice of kernel function doesn't play an important role in the estimator. The survival probability estimator of Z derived by $\hat{f}_Z^{IP}$ has lower survival probability than the one derived by $\hat{f}_Z^{KM}$.

2.2 Confidence Interval

By using $\hat{f}_Z^{IP}(z)$ in (1.12), we obtain a point estimate of EZ: $\hat{\theta}_Z = \int z\hat{f}_Z^{IP}(z)\,dz$. But we don't know the distribution of $\hat{\theta}_Z$, thus we use the percentile bootstrap confidence interval [13] of EZ, which is a nonparametric version. The procedures of obtaining confidence interval are as follows:

(1) Draw B samples $W_{i1}^*, \cdots, W_{in}^*$ from the observations $W_1, \cdots, W_n$ with replacement:

$$\begin{array}{c} W_{11}^*, \cdots, W_{1n}^* \\ \cdots\cdots\cdots\cdots \\ W_{B1}^*, \cdots, W_{Bn}^*. \end{array} \tag{2.5}$$

(2) Find the density estimator $\hat{f}_Z^{*i}(z) = \sum_{j=1}^n \delta_j\, K^E\left((z - W_{ij}^*)/h\right)/(nh(1 - G(W_{ij}^*)))$, $(i = 1, \cdots, B)$ based on the bootstrap sample, where K^E is given in (1.12). Thus the bootstrap estimated mean is $\hat{\theta}_i^* = \int z\hat{f}_Z^{*i}(z)\,dz$, from which we obtain $\hat{\theta}_1^*, \cdots, \hat{\theta}_B^*$.

(3) Order the estimates $\hat{\theta}_{(1)}^*, \cdots, \hat{\theta}_{(B)}^*$ to find the approximate $(1-\alpha)*100\%$ percentile bootstrap confidence interval $(\hat{\theta}_{(A)}^*, \hat{\theta}_{(B-A)}^*)$ of EZ, where $A = [\alpha/2 * B]$.

CHAPTER 3

INDEPENDENT CASE

We now are going to discuss the asymptotic normality of $\hat{f}_Z^{IP}(z)$ when $\{(W_i, \delta_i) =: \xi_i, i \geq 1\}$ is a sequence of independent random vectors.

Fan [7] pointed out the rates of convergence depend strongly on the smoothness of error distributions. Similar to Fan's results, we separate our results into two cases as well: when the smoothness of error distribution is ordinary smooth and when it is supersmooth. In model $X = Z + E$, the examples of ordinary smooth distributions are gamma, exponential, uniform with non-negative support, and their mixtures; examples of supersmooth error distributions include truncated normal, truncated Cauchy, and truncated mixture normal whose supports are all non-negative. Hence we assume that φ_{f_E} satisfies either:

Assumption 3.0.1.

(i) Ordinary smooth case:

$$\varphi_{f_E}(t)\, t^{\beta} \to c \quad \textit{as } t \to \infty \tag{3.1}$$

with some constant $c \neq 0$ and $\beta \geq 0$. Moreover we assume $\varphi_{f_E}(t) \neq 0$, for all t.

Or

(ii) Supersmooth case:

$$c_1\, |t|^{\beta_0} \exp\left(-|t|^{\beta}/\gamma\right) \leq |\varphi_{f_E}(t)| \leq c_2\, |t|^{\beta_0} \exp\left(-|t|^{\beta}/\gamma\right) \quad \textit{as } t \to \infty \tag{3.2}$$

with β, γ, c_1, $c_2 > 0$ and some real number β_0. $\varphi_{f_E}(t) \neq 0$ for all t.

Write $\varphi_{f_E}(t) = R_E(t) + \mathrm{i}I_E(t)$, where $R_E(t)$ and $I_E(t)$ denote the real and the imaginary part of the characteristic function $\varphi_{f_E}(t)$. Assume furthermore that either $I_E(t) = o(R_E(t))$ or $R_E(t) = o(I_E(t))$ as $t \to \infty$.

In this chapter, we will consider two different cases: E follows an ordinary smooth distribution and a supersmooth distribution when $\{(W_i, \delta_i) =: \xi_i, i \geq 1\}$ is a sequence of independent random vectors.

We assume $\sup_x 1/[1 - G(x)] \leq B_0 < \infty$, B_0 is a positive constant.

3.1 Case I: Ordinary Smooth Distribution

First we consider $\hat{f}_Z^{IP}(z)$ under ordinary smooth error distribution, which satisfies Assumption 3.0.1 (i). Following assumptions on the characteristic function φ_K of kernel function K are also needed.

Assumption 3.1.1.

(i) $\varphi_K(t)$ *is a symmetric function, having* $s+2$ *bounded integrable derivatives,* $\varphi_K(0) = 1$.

(ii) $\int_{-\infty}^{\infty} [|\varphi_K(t)| + |\varphi_K'(t)|]\ |t|^{\beta} dt < \infty$, $\int_{-\infty}^{\infty} |t|^{2\beta} |\varphi_K(t)|^2 dt < \infty$.

(iii) $\int_{-\infty}^{\infty} f_{X|\delta=1}(x)/[1 - G(x)]^{2+\nu}\ dx < \infty$ *for some* $\nu > 0$.

The following theorem gives the asymptotic normality of the estimator $\hat{f}_Z^{IP}(z)$.

Theorem 3.1.1. *Use Assumption 3.1.1, if $h = h_n \to 0$, and $nh \to \infty$, then*

$$\frac{\hat{f}_Z^{IP}(z) - E\hat{f}_Z^{IP}(z)}{\sqrt{\mathrm{Var}(\hat{f}_Z^{IP}(z))}} \xrightarrow{d} N(0,1). \tag{3.3}$$

In order to check (3.3), we need the following lemma, which generalizes the Theorem 1A of Parzen [17].

Lemma 3.1.1. *Suppose that $K_n(\cdot)$ is a sequence of Borel functions satisfying*

$$K_n(x) \to \widetilde{K}(x) \quad \textit{and} \quad \sup_n |K_n(x)| \leq K^*(x)$$

where $K^*(x)$ *satisfies* $\int_{-\infty}^{\infty} K^*(x)\,dx < \infty$, $\lim_{x\to\infty} |xK^*(x)| = 0$. *Let* $m(x)$ *satisfy* $\int_{-\infty}^{\infty} |m(x)|\,dx < \infty$, *if* z *is a continuity point of* $m(\cdot)$, *then for any sequence* $h = h_n \to 0$,

$$\lim_{n\to\infty} \frac{1}{h} \int_{-\infty}^{\infty} K_n\left(\frac{z-x}{h}\right) m(x)\,dx = m(z) \int_{-\infty}^{\infty} \widetilde{K}(x)\,dx$$

Proof. Let $b > 0$, and split the region of integration into two parts: $|x| \le b$ and $|x| > b$, we get

$$\begin{aligned}
&\left|\frac{1}{h} \int_{-\infty}^{\infty} K_n\left(\frac{z-x}{h}\right) m(x)\,dx - m(z) \int_{-\infty}^{\infty} \widetilde{K}(x)\,dx\right| \\
&= \left|\frac{1}{h} \int_{-\infty}^{\infty} K_n\left(\frac{x}{h}\right) m(z-x)\,dx - m(z) \int_{-\infty}^{\infty} \widetilde{K}(x)\,dx\right| \\
&\le \left|\frac{1}{h} \int_{-\infty}^{\infty} K_n\left(\frac{x}{h}\right) [m(z-x) - m(z)]\,dx\right| + |m(z)| \left|\int_{-\infty}^{\infty} \left(\frac{1}{h} K_n\left(\frac{x}{h}\right) - \widetilde{K}(x)\right) dx\right| \\
&\le \max_{|x|\le b} |m(z-x) - m(z)| \int_{-\infty}^{\infty} K^*(x)\,dx + \int_{|x|\ge b} \frac{|m(z-x)|}{x} \frac{x}{h} K_n\left(\frac{x}{h}\right) dx \\
&\quad + |m(z)| \int_{|x|\ge b} \frac{1}{h} K_n\left(\frac{x}{h}\right) dx + |m(z)| \left|\int_{-\infty}^{\infty} \left(K_n(x) - \widetilde{K}(x)\right) dx\right| \\
&\le \max_{|x|\le b} |m(z-x) - m(z)| \int_{-\infty}^{\infty} K^*(x)\,dx + \frac{1}{b} \sup_{|y|\ge \frac{b}{h}} |yK^*(y)| \int_{-\infty}^{\infty} |m(x)|\,dx \\
&\quad + |m(z)| \int_{|y|\ge \frac{b}{h}} K^*(y)\,dy + |m(z)| \left|\int_{-\infty}^{\infty} \left(K_n(x) - \widetilde{K}(x)\right) dx\right| \qquad (3.4)
\end{aligned}$$

When $n \to \infty$, the last three terms tend to 0 by the assumptions and Lebesgue's dominated convergence theorem. Let $b \to 0$, the first term also tends to 0. □

We now are ready to prove Theorem 3.1.1.

Proof of Theorem 3.1.1:

Proof. To discuss the asymptotic normality of the estimator $\hat{f}_Z^{IP}(z)$, we first rewrite $\hat{f}_Z^{IP}(z)$ in (1.12) as:

$$\hat{f}_Z^{IP}(z) = \frac{1}{nh} \sum_{j=1}^{n} K^E\left(\frac{z-W_j}{h}\right) \frac{\delta_j}{1-G(W_j)} := \frac{1}{n} \sum_{j=1}^{n} U_j \qquad (3.5)$$

where $U_j = \delta_j\, K^E((z-W_j)/h)/(h(1-G(W_j))$. Note that $\hat{f}_Z^{IP}(z)$ is the sum of an independent sequence and

$$\frac{\hat{f}_Z^{IP}(z) - E\hat{f}_Z^{IP}(z)}{\sqrt{\mathrm{Var}(\hat{f}_Z^{IP}(z))}} = \frac{n\,\hat{f}_Z^{IP}(z) - E\left(\sum_{j=1}^n U_j\right)}{\sqrt{\mathrm{Var}\left(\sum_{j=1}^n U_j\right)}}$$

$$= \frac{\sum_{j=1}^n (U_j - EU_j)}{\sum_{j=1}^n \mathrm{Var}U_j} \tag{3.6}$$

Hence it's equivalent to show (3.6) $\xrightarrow{\mathrm{d}} N(0,1)$, for which a sufficient condition is that Lyapunov's condition holds, i.e. for some $\nu > 0$

$$\frac{\sum_{j=1}^n E|U_j - EU_j|^{2+\nu}}{\left[\sum_{j=1}^n \mathrm{Var}(U_j)\right]^{1+\nu/2}} \to 0 \tag{3.7}$$

Therefore we proceed the proof of (3.7).

First, we evaluate the limit of $E{U_j}^2$ by (1.3).

$$\begin{aligned} E{U_j}^2 &= E\left[\frac{1}{h}\,K^E\left(\frac{z-W_j}{h}\right)\frac{\delta_j}{1-G(W_j)}\right]^2 \\ &= \frac{1}{h^2}\left\{\int_{-\infty}^{\infty}\left[K^E\left(\frac{z-w}{h}\right)\right]^2 \frac{1}{[1-G(w)]^2}\, f_{X|\delta=1}(w)\,p\,dw \right.\\ &\quad \left. + \int_{-\infty}^{\infty}\left[K^E\left(\frac{z-w}{h}\right)\right]^2 \frac{0}{[1-G(w)]^2}\, g_{C|\delta=0}(w)\,(1-p)\,dw\right\} \\ &= \frac{p}{h^2}\int_{-\infty}^{\infty}\left[\frac{K^E\left(\frac{z-w}{h}\right)}{1-G(w)}\right]^2 f_{X|\delta=1}(w)\,dw \\ &= \frac{p}{h^{2+2\beta}}\int_{-\infty}^{\infty}\left[\frac{h^\beta K^E\left(\frac{z-w}{h}\right)}{1-G(w)}\right]^2 f_{X|\delta=1}(w)\,dw \end{aligned} \tag{3.8}$$

Now we need to check whether $[h^\beta K^E(w)]^2$ satisfies Lemma 3.1.1's conditions before we apply Lemma 3.1.1 to (3.8). By Assumption 3.0.1 (i) and Lebesgue's dominated convergence theorem,

$$h^\beta K^E(w) \to \frac{1}{2\pi c}\int_{-\infty}^{\infty} e^{-\mathrm{i}yw}\; y^\beta \varphi_K(y)dy$$

Applying Plancherel's theorem and by Assumption 3.1.1 (ii),

$$\int_{-\infty}^{\infty} [h^\beta K^E(w)]^2 dw \to \frac{1}{(2\pi c)^2} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-\mathrm{i}yw}\; y^\beta \varphi_K(y) dy \right]^2 dw$$
$$= \frac{1}{2\pi c^2} \int_{-\infty}^{\infty} |y^\beta \varphi_K(y)|^2 dy < \infty \tag{3.9}$$

Fan [7] has proved $[h^\beta K^E(w)]^2 \le C_1/w^2$ for some constant C_1, which leads to $\lim_{w\to\infty} |w \cdot (C_1/w^2)| = 0$. Using Assumption 3.1.1 (iii), we now are ready to apply Lemma 3.1.1 to (3.8). By (3.9),

$$\lim_{n\to\infty} h^{1+2\beta}\, E U_j{}^2 = \frac{p\, f_{X|\delta=1}(z)}{[1-G(z)]^2} \int_{-\infty}^{\infty} [h^\beta\, K^E(w)]^2 dw \tag{3.10}$$

$|EU_j|$ can be expressed as

$$|EU_j| = \left| E\left[\frac{1}{h}\; K^E\left(\frac{z-W_j}{h} \right) \frac{\delta_j}{1-G(W_j)} \right] \right|$$
$$= \frac{p}{h} \left| \int_{-\infty}^{\infty} K^E\left(\frac{z-w}{h} \right) \frac{f_{X|\delta=1}(w)}{1-G(w)} dw \right|$$
$$\le \frac{p}{h}\, B_0 \left| \int_{-\infty}^{\infty} K^E\left(\frac{z-w}{h} \right) f_{X|\delta=1}(w)\, dw \right|$$
$$= \frac{p}{h}\, B_0 \left| E\left[K^E\left(\frac{z-X_j}{h} \right) \middle| \, \delta_j = 1 \right] \right| \tag{3.11}$$

From [21], [7], and recall E_i is independent of Z_i and δ_i, apply Lemma 3.1.1 we know

$$\frac{1}{h}\, E\left[K^E\left(\frac{z-X_j}{h} \right) \middle| \, \delta_j = 1 \right] = \frac{1}{h}\, E\left[K\left(\frac{z-Z_j}{h} \right) \middle| \, \delta_j = 1 \right]$$
$$= \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{z-x}{h} \right) f_{Z|\delta=1}(x)\, dx$$
$$\to f_{Z|\delta=1}(z) \int_{-\infty}^{\infty} K(x) dx = f_{Z|\delta=1}(z) \tag{3.12}$$

Thus we conclude

$$|EU_j| = \mathcal{O}(1). \tag{3.13}$$

Similarly, by $|h^\beta \cdot K^E(w)| \le \min(C_2, C_1/w)$ in [7] and Lemma 3.1.1, we show that

$$E|U_j|^{2+\nu} = E \left| \frac{1}{h}\; K^E\left(\frac{z-W_j}{h} \right) \frac{\delta_j}{1-G(W_j)} \right|^{2+\nu}$$

$$= \frac{p}{h^{(2+\nu)(1+\beta)}} \int_{-\infty}^{\infty} \left[\frac{h^\beta K^E\left(\frac{z-w}{h}\right)}{1-G(w)} \right]^{2+\nu} f_{X|\delta=1}(w)\, dw$$
$$= \frac{p}{h^{(2+\nu)(1+\beta)}} \int_{-\infty}^{\infty} \left[h^\beta K^E\left(\frac{z-w}{h}\right) \right]^{2+\nu} \frac{f_{X|\delta=1}(w)}{[1-G(w)]^{2+\nu}}\, dw$$

note that $\int_{\mathbb{R}} |h^\beta K^E(u)|^{2+\nu} du \le C(1+o(1))$, C is some positive constant by [15] (Lemma 2), hence

$$\lim_{n\to\infty} h^{(2+\nu)(1+\beta)-1} E|U_j|^{2+\nu} = \frac{p\, f_{X|\delta=1}(z)}{[1-G(z)]^{2+\nu}} \int_{-\infty}^{\infty} [h^\beta K^E(w)]^{2+\nu} dw, \tag{3.14}$$

which implies

$$E|U_j|^{2+\nu} = \mathcal{O}(h^{-(2+\nu)(1+\beta)+1}) \tag{3.15}$$

Thus, by c_r-inequality [6], Lyapunov's condition (3.7) follows from

$$\frac{\sum_{j=1}^n E|U_j - EU_j|^{2+\nu}}{\left[\sum_{j=1}^n \mathrm{Var}(U_j)\right]^{1+\nu/2}} \le 2^{1+\nu} \frac{\sum_{j=1}^n (E|U_j|^{2+\nu} + |EU_j|^{2+\nu})}{\left[\sum_{j=1}^n \mathrm{Var}(U_j)\right]^{1+\nu/2}} \tag{3.16}$$
$$= \mathcal{O}(1/(nh)^{\nu/2}) \to 0$$

□

3.2 Case II: Supersmooth Distribution

Supersmooth error models will be considered in this subsection, and we assume that φ_{f_E} satisfies Assumption 3.0.1 (ii) and φ_K satisfies the following assumptions.

Assumption 3.2.1.

(i) $\varphi_K(t)$ *is a symmetric function, supported with* $[-1,1]$ *and having the first* $s+2$ *continuous derivatives,* $\varphi_K(t) > c_3(1-t)^{s+3}$, *for* $t \in [1-\psi, 1)$ *for some* $\psi > 0$, *and* $c_3 > 0$.

(ii) $\int_{-\infty}^{\infty} f_{X|\delta=1}(x)/[1-G(x)]^{2+\nu}\, dx < \infty$ *for some* $\nu > 0$.

Following theorem states the limiting distribution of $\hat{f}_Z^{IP}(z)$ when the error distribution is supersmooth.

Theorem 3.2.1. *Under Assumption 3.2.1, we have*

$$\frac{\hat{f}_Z^{IP}(z) - E\hat{f}_Z^{IP}(z)}{\sqrt{\mathrm{Var}(\hat{f}_Z^{IP}(z))}} \xrightarrow{d} N(0,1). \tag{3.17}$$

provided $h = h_n = a(\ln n)^{-1/\beta}$, *for some* $a > 0$.

Before we proceed with the proof of Theorem 3.2.1, we need the following lemma from Fan [7].

Lemma 3.2.1. *Under Assumption 3.2.1, as* $n \to \infty$,

$$|K^E(x)| \geq c_4\, q(x)\ \exp\left(\frac{(1-b_n)^\beta}{\gamma h^\beta}\right) h^{\beta_0}\, b_n^{s+4} \tag{3.18}$$

uniformly over $x \in [0, \pi/2]$, *where* $h = h_n$, $b_n = h_n^{\beta/(2(s+5))}$ *and* c_4 *is a positive constant, and*

$$q(x) = \begin{cases} |\cos x|, & if\ I_E(t) = o(R_E(t)) \\ |\sin x|, & if\ R_E(t) = o(I_E(t)) \end{cases}$$

where $R_E(t)$ *and* $I_E(t)$ *denote the real and the imaginary part of the characteristic function* $\varphi_{f_E}(t)$.

Now, by using Lemma 3.2.1, we are ready to prove Theorem 3.2.1.

Proof of Theorem 3.2.1:

Proof. We will check (3.7) holds, which is the sufficient condition of (3.17).

By Lemma 3.2.1, when n is sufficiently large, we first find the lower bound of $EU_j{}^2$.

$$\begin{aligned} EU_j{}^2 &= E\left[\frac{1}{h}\, K^E\left(\frac{z - W_j}{h}\right) \frac{\delta_j}{1 - G(W_j)}\right]^2 \\ &= \frac{p}{h^2} \int_{-\infty}^{\infty} \left[\frac{K^E\left(\frac{z-w}{h}\right)}{1 - G(w)}\right]^2 f_{X|\delta=1}(w)\, dw \end{aligned}$$

$$
\begin{aligned}
&= \frac{p}{h} \int_{-\infty}^{\infty} \left[K^E(t)\right]^2 \frac{f_{X|\delta=1}(z-ht)}{[1-G(z-ht)]^2}\, dt \\
&\geq \frac{p}{h} \int_{0}^{\frac{\pi}{2}} \left[c_4\, q(x)\, \exp\left(\frac{(1-b_n)^\beta}{\gamma h^\beta}\right) h^{\beta_0}\, b_n^{s+4}\right]^2 \frac{f_{X|\delta=1}(z-ht)}{[1-G(z-ht)]^2}\, dt \\
&= \frac{p}{h} \left[c_4\, \exp\left(\frac{(1-b_n)^\beta}{\gamma h^\beta}\right) h^{\beta_0}\, b_n^{s+4}\right]^2 \int_{0}^{\frac{\pi}{2}} q^2(t) \frac{f_{X|\delta=1}(z-ht)}{[1-G(z-ht)]^2} dt \\
&\geq c_5\, h^{c_6}\, \frac{f_{X|\delta=1}(z)}{[1-G(z)]^2}\, \exp\left(\frac{2(1-b_n)^\beta}{\gamma h^\beta}\right) \qquad (3.19)
\end{aligned}
$$

Recall $b_n = h_n^{\beta/(2(s+5))}$, $h = a(\ln n)^{-1/\beta}$, we obtain $b_n = a^{\beta/(2s+10)}(\ln n)^{-1/(2s+10)} \to 0$ (as $n \to \infty$). According to Taylor's theorem, $(1-b_n)^\beta = 1 - \beta b_n + R(b_n)$, where $R(b_n) = o(b_n)$. It follows that $(1-b_n)^\beta \to 1 - \beta b_n \geq 1 - 2\beta b_n$. Thus,

$$
EU_j^{\,2} \geq c_5\, \frac{f_{X|\delta=1}(z)}{[1-G(z)]^2}\, h^{c_6}\, \exp\left(\frac{2-4\beta b_n}{\gamma h^\beta}\right) \qquad (3.20)
$$

for some constant $c_5 > 0$ and $c_6 = 2\beta_0 - 1 + \beta(s+4)/(s+5)$.

Next we will find the upper bound of EU_j and $E|U_j|^{2+\delta}$ by using the upper bound for $|K^E|$ provided in [7]: $|K^E(t)| \leq \mathcal{O}(\exp(1/(\gamma h^\beta))c_n)$, where $h = h_n$,

$$
c_n = \begin{cases} 1, & \beta_0 \geq 0 \\ h_n^{\,\beta_0}, & \beta_0 < 0 \end{cases} \qquad (3.21)
$$

Note that $|EU_j|$ is bounded as shown in (3.13). Moreover,

$$
\begin{aligned}
E|U_j|^{2+\nu} &= E\left|\frac{1}{h}\, K^E\left(\frac{z-W_j}{h}\right) \frac{\delta_j}{1-G(W_j)}\right|^{2+\nu} \\
&= \frac{p}{h^{2+\nu}} \int_{-\infty}^{\infty} \left[\frac{K^E\left(\frac{z-w}{h}\right)}{1-G(w)}\right]^{2+\nu} f_{X|\delta=1}(w)\, dw \\
&= \frac{p}{h^{2+\nu}} \int_{-\infty}^{\infty} \left[K^E\left(\frac{z-w}{h}\right)\right]^{2+\nu} \frac{f_{X|\delta=1}(w)}{[1-G(w)]^{2+\nu}}\, dw \\
&\leq \frac{p}{h^{2+\nu}}\, \mathcal{O}\left(\exp\left(\frac{2+\nu}{\gamma h^\beta}\right) c_n^{\,2+\nu}\right) \int_{-\infty}^{\infty} \frac{f_{X|\delta=1}(w)}{[1-G(w)]^{2+\nu}}\, dw \\
&= \mathcal{O}(h^{-(2+\nu)}\, \exp\left(\frac{2+\nu}{\gamma h^\beta}\right) c_n^{\,2+\nu}) \qquad (3.22)
\end{aligned}
$$

In consequence, by (3.16),

$$\frac{\sum_{j=1}^{n} E|U_j - EU_j|^{2+\nu}}{\left[\sum_{j=1}^{n} \mathrm{Var}(U_j)\right]^{1+\nu/2}} \leq 2^{1+\nu} \frac{\sum_{j=1}^{n} \left(E|U_j|^{2+\nu} + |EU_j|^{2+\nu}\right)}{\left[\sum_{j=1}^{n} EU_j^2\right]^{1+\nu/2}} \to 0 \tag{3.23}$$

by choosing $h = a(\ln n)^{-1/\beta}$, for some a such that $a > (2/\gamma)^{1/\beta}$. □

CHAPTER 4

STRONG MIXING CASE

Based on the work of Masry [16], we are also interested in finding the asymptotic normality of estimator $\hat{f}_Z^{IP}(z)$ when $\{(W_i, \delta_i) =: \xi_i, i \geq 1\}$ is a sequence of strong mixing (α mixing) random vectors, i.e.

$$\alpha(p) := \sup_{i\geq 1}\{|P(A\cap B) - P(A)P(B)| : A \in \mathcal{F}_1^i, B \in \mathcal{F}_{p+i}^{\infty}\} \tag{4.1}$$

converges to 0 as $p \to \infty$, where $\mathcal{F}_a^b$ denotes the σ-algebra generated by $\{\xi_i, a \leq i \leq b\}$.

We assume $\sup_x 1/[1 - G(x)] \leq B_0 < \infty$, B_0 is a positive constant. Let $P(\delta_i = 1, \delta_j = 1) = p^*$ $(1 \leq i < j \leq 1)$.

In order to introduce the main results under two different cases of error distributions, we need the following Lemma 4.0.1 [11] and Lemma 4.0.2 [22].

Lemma 4.0.1. *For random variable variables U and V which are $\mathcal{F}_1^i$, $\mathcal{F}_{p+i}^{\infty}$ measurable, respectively, with $E|U|^{\nu} < \infty$, $E|V|^{\nu} < \infty$, for $\nu > 2$,*

$$|\mathrm{Cov}(U, V)| \leq 8\,[\alpha(p)]^{1-\frac{2}{\nu}}\,\{E|U|^{\nu}\; E|V|^{\nu}\}^{\frac{1}{\nu}} \tag{4.2}$$

Lemma 4.0.2. *Let $V_1, \ldots, V_L$ be random variables measurable with respect to the σ-algebra $\mathcal{F}_{i_1}^{j_1}, \cdots, \mathcal{F}_{i_L}^{j_L}$, respectively, with $1 \leq i_1 < j_1 < i_2 < \cdots < j_L \leq n$, $i_{l+1} - j_l \geq \chi \geq 1$ and $|V_j| \leq 1$ for $j = 1, \cdots, L$. Then*

$$\left| E\left(\prod_{j=1}^{L} V_j\right) - \prod_{j=1}^{L} E(V_j)\right| \leq 16\,(L-1)\,\alpha(\chi) \tag{4.3}$$

For the remainder of this section, we will follow the framework Masry used in [16].

4.1 Case I: Ordinary Smooth Distribution

We first present the case when the smoothness of error distribution is ordinary smooth of order β, we need the Assumption 3.0.1 (i) on the tail of characteristic function φ_{f_E} and following assumptions on φ_K and the process $\{Z_i, E_i\}_{i=1}^n$.

Assumption 4.1.1.

(i) $\int_{-\infty}^{\infty} |t|^{\beta-2}\, |\varphi_K(t)| dt < \infty$ *for* $\beta > 1$; $\int_{-\infty}^{\infty} |t|^{2\beta} |\varphi_K(t)|^2 dt < \infty$.

(ii) $\varphi_K(t)$ *is twice differentiable with bounded derivative such that* $\int_{-\infty}^{\infty} |t|^{\beta-1}\, |\varphi_K'(t)| dt < \infty$; $\int_{-\infty}^{\infty} |t|^{\beta}\, |\varphi_K''(t)| dt < \infty$.

(iii) $\sup_x f_{X|\delta=1}(x) \le M_1 < \infty$; $\sup_{x_i, x_j} f_{(X_i,X_j)|(\delta_i=1,\delta_j=1)}(x_i, x_j) \le M_2 < \infty$ *(*$f_{(X_i,X_j)|(\delta_i=1,\delta_j=1)}$ *is the conditional probability density function of joint distribution* (X_i, X_j)*, given that* $\delta_i = 1$, $\delta_j = 1$. $1 \le i < j \le n$*).* M_1 *and* M_2 *are positive constants.*

(iv) $\int_{-\infty}^{\infty} f_{X|\delta=1}(x)/[1 - G(x)]^2\, dx < \infty$.

(v) $\sum_{j=1}^{\infty} j^{\rho}\, [\alpha(j)]^{1-2/\nu} < \infty$ *for some* $\nu > 2$ *and* $\rho > 1 - 2/\nu$.

Following theorem states the asymptotic normality of $\hat{f}_Z^{IP}(z)$ when the ordinary smooth error is involved.

Theorem 4.1.1. *Suppose that Assumption 4.1.1 is satisfied and assume* $h = h_n \to 0$ *such that* $nh_n^{2\beta+1} \to \infty$, $\alpha(p)$ *satisfies* $(n/h)^{1/2}\, \alpha(r_n) \to 0$ *as* $n \to \infty$*. Let* $\{r_n\}$ *be a sequence of positive integers,* $r_n \to \infty$*, such that* $r_n = o(nh^{1/2})$ *as* $n \to \infty$*. Then*

$$\sqrt{nh^{2\beta+1}}\, \frac{[\hat{f}_Z^{IP}(z) - E\hat{f}_Z^{IP}(z)]}{\sigma_0(z)} \xrightarrow{d} N(0,1) \tag{4.4}$$

where

$$\sigma_0^2(z) = \frac{p}{2\pi c^2} \int_{-\infty}^{\infty} \left| y^{\beta}\, \varphi_K(y) \right|^2\, dy\, \frac{f_{X|\delta=1}(z)}{[1 - G(z)]^2} \tag{4.5}$$

In order to prove Theorem 4.1.1, we need the following lemma whose proof is similar to the work of Masry [15].

Lemma 4.1.1. *Under Assumptions 4.1.1, we have*

$$\lim_{n\to\infty} nh^{2\beta+1}\,\mathrm{Var}[\hat{f}_Z^{IP}(z)] = \sigma_0^2(z) \tag{4.6}$$

at points of continuity of $f_{X|\delta=1}/[1-G]^2$.

Proof. We first evaluate the variance of the estimator by separating it into two parts: $I_{n,0}$ and S_0. By (3.5),

$$\mathrm{Var}[\hat{f}_Z^{IP}(z)] = \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}(U_j) + \frac{2}{n^2}\sum_{1\le i<j\le n}\mathrm{Cov}(U_i,U_j) := I_{n,0} + S_0 \tag{4.7}$$

We obtain $|EU_j| = \mathcal{O}(1)$ from (3.13), hence $I_{n,0}$ can be written as:

$$\begin{aligned} I_{n,0} &= \frac{1}{n^2}\sum_{j=1}^{n}[E(U_j)^2 - (EU_j)^2] \\ &= \frac{1}{n^2h^2}\sum_{j=1}^{n} E\left[K^E\left(\frac{z-W_j}{h}\right)\frac{\delta_j}{1-G(W_j)}\right]^2 - \mathcal{O}(n^{-1}) \\ &= \frac{p}{nh^2}\int_{-\infty}^{\infty}\left[K^E\left(\frac{z-w}{h}\right)\right]^2 \frac{f_{X|\delta=1}(w)}{[1-G(w)]^2}dw + \mathcal{O}(n^{-1}) \end{aligned} \tag{4.8}$$

By Lemma 3.1.1 and Assumption 4.1.1 (iv), the proof is similar to (3.9) and (3.10),

$$\frac{h^{2\beta}}{h}\int_{-\infty}^{\infty}\left[K^E\left(\frac{z-w}{h}\right)\right]^2 \frac{f_{X|\delta=1}(w)}{[1-G(w)]^2}dw \to \frac{f_{X|\delta=1}(z)}{[1-G(z)]^2}\int_{-\infty}^{\infty}[h^\beta K^E(w)]^2\,dw \tag{4.9}$$

therefore

$$\lim_{n\to\infty} nh^{2\beta+1}\, I_{n,0} = \frac{p\, f_{X|\delta=1}(z)}{[1-G(z)]^2}\,\frac{1}{2\pi c^2}\int_{-\infty}^{\infty}\left|y^\beta\,\varphi_K(y)\right|^2\,dy + \mathcal{O}(h^{2\beta+1}) \tag{4.10}$$

Recall $\int_{-\infty}^{\infty}|y^\beta\,\varphi_K(y)|^2\,dy < \infty$ by Assumption 4.1.1 (i), we have as $h = h_n \to 0$,

$$\lim_{n\to\infty} nh^{2\beta+1}\, I_{n,0} = \frac{p}{2\pi c^2}\int_{-\infty}^{\infty}\left|y^\beta\,\varphi_K(y)\right|^2\,dy\,\frac{f_{X|\delta=1}(z)}{[1-G(z)]^2} = \sigma_0^2(z) \tag{4.11}$$

at points of continuity of $f_{X|\delta=1}/[1-G]^2$.

Next consider the covariance term S_0, we will show $S_0 = o(n^{-1} h^{-2\beta-1})$ as follows.

$$S_0 = \frac{2}{n^2} \sum_{1 \le i < j \le n} \text{Cov}(U_i, U_j) \le \frac{2}{n^2} \left| \sum_{1 \le i < j \le n} \text{Cov}(U_i, U_j) \right| \le \frac{2}{n^2} \sum_{1 \le i < j \le n} |\text{Cov}(U_i, U_j)| \tag{4.12}$$

We will then separate the right side of (4.12) into two parts:

$$\begin{aligned} & \frac{2}{n^2} \sum_{1 \le i < j \le n} |\text{Cov}(U_i, U_j)| \\ = & \frac{2}{n^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} |\text{Cov}(U_i, U_j)| + \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} |\text{Cov}(U_i, U_j)| := S_1 + S_2 \end{aligned} \tag{4.13}$$

where $l_n \to \infty$ and $l_n \, h \to 0$ as $n \to \infty$.

Now show $S_1 = o(n^{-1} h^{-1-2\beta})$.

$$\begin{aligned} S_1 = & \frac{2}{n^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} |\text{Cov}(U_i, U_j)| \\ \le & \frac{2}{n^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} |E(U_i \, U_j))| + \frac{2}{n^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} |E(U_i) \; E(U_j)| \\ \le & \frac{2}{n^2 h^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} \left| E\left[K^E \left(\frac{z - W_i}{h} \right) \frac{\delta_i}{1 - G(W_i)} \; K^E \left(\frac{z - W_j}{h} \right) \frac{\delta_j}{1 - G(W_j)} \right] \right| \\ & + \frac{2}{n^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} \mathcal{O}(1) := \frac{2}{n^2 h^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} |I_{i,j}| + \frac{2}{n^2} \sum_{\substack{1 \le i < j \le n \\ j-i \le l_n}} \mathcal{O}(1) \end{aligned} \tag{4.14}$$

where

$$\begin{aligned} I_{i,j} = & \sum_{\substack{\delta_i^* = 0,1 \\ \delta_j^* = 0,1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^E \left(\frac{z - w_i}{h} \right) \; K^E \left(\frac{z - w_j}{h} \right) \frac{f_{(W_i, W_j)|(\delta_i = \delta_i^*, \delta_j = \delta_j^*)}(w_i, w_j)}{[1 - G(w_i)][1 - G(w_j)]} \\ & \cdot \; \delta_i^* \, \delta_j^* \, P(\delta_i = \delta_i^*, \delta_j = \delta_j^*) \, dw_i \, dw_j \\ = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^E \left(\frac{z - w_i}{h} \right) \; K^E \left(\frac{z - w_j}{h} \right) \frac{f_{(X_i, X_j)|(\delta_i = 1, \delta_j = 1)}(w_i, w_j)}{[1 - G(w_i)][1 - G(w_j)]} \end{aligned}$$

$$\cdot\ P(\delta_i = 1, \delta_j = 1)\, dw_i\, dw_j \tag{4.15}$$

By Assumption 4.1.1 (iii),

$\sup_{w_i, w_j} f_{(X_i,X_j)|(\delta_i=1,\delta_j=1)}(w_i, w_j)/\{[1 - G(w_i)][1 - G(w_j)]\} \leq M_2\, B_0^2$ and note that $||K^E||_1 \leq C/h^\beta$, C is some positive constant, provided in [15] (Lemma 3) and $0 \leq P(\delta_i = 1, \delta_j = 1) = p^* \leq 1$, using (4.15) we get

$$|I_{i,j}| \leq C\ \left[h \int_{-\infty}^{\infty} |K^E(u)| du\right]^2 \leq \frac{Ch^2}{h^{2\beta}} \tag{4.16}$$

Hence by $l_n\, h \to 0$ as $n \to \infty$,

$$nh^{2\beta+1}\ S_1 \leq nh^{2\beta+1}\ \frac{(2n - l_n - 1)\, l_n}{2} \left(\frac{2}{n^2h^2}\ \frac{Ch^2}{h^{2\beta}} + \frac{2}{n^2}\mathcal{O}(1)\right) = \mathcal{O}(2l_n\, h) \to 0 \tag{4.17}$$

Next we consider S_2 and show $S_2 = o(n^{-1}\, h^{-1-2\beta})$. Apply Lemma 4.0.1, for $\nu > 2$,

$$\begin{aligned} S_2 &= \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} |\mathrm{Cov}(U_i, U_j)| \\ &\leq \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} 8\, [\alpha(j-i)]^{1-\frac{2}{\nu}}\, \{E|U_i|^\nu\ E|U_j|^\nu\}^{\frac{1}{\nu}} \\ &= \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} \frac{8}{h^2}\, [\alpha(j-i)]^{1-\frac{2}{\nu}} \left\{\int_{-\infty}^{\infty} \left|K^E\left(\frac{z-w}{h}\right)\right|^\nu \frac{f_{X|\delta=1}(w)\, p}{[1-G(w)]^\nu} dw\right\}^{\frac{2}{\nu}} \end{aligned} \tag{4.18}$$

From Assumption 4.1.1 (iii), we get $\sup_w f_{X|\delta=1}(w)/[1 - G(w)]^\nu \leq M_1\, B_0^\nu$ for $\nu > 2$. By [15] (Lemma 2), we have $\int_{\mathbb{R}} |K^E(u)|^\nu du \leq C(1 + o(1))/h^{\beta\nu}$, thus,

$$\begin{aligned} S_2 &\leq \frac{C}{n^2h^2} \sum_{j=l_n+1}^{n-1} (n-j)\, [\alpha(j)]^{1-\frac{2}{\nu}}\ \left[h \int_{-\infty}^{\infty} |K^E(u)|^\nu du\right]^{\frac{2}{\nu}} \\ &\leq \frac{C}{nh^2} \sum_{j=l_n+1}^{n-1} [\alpha(j)]^{1-\frac{2}{\nu}}\ \left(\frac{h}{h^{\beta\nu}}\right)^{\frac{2}{\nu}} \end{aligned}$$

$$\leq \frac{C}{nh^{2\beta+2-\frac{2}{\nu}}} \frac{1}{l_n^\rho} \sum_{j=l_n+1}^{n-1} j^\rho \left[\alpha(j)\right]^{1-\frac{2}{\nu}} \tag{4.19}$$

C is some positive constant, $\rho > 1 - 2/\nu$. By choosing $l_n = h^{(2/\nu-1)/\rho}$, $l_n\, h \to 0$ is also satisfied. Then using (4.19) and Assumption 4.1.1 (v), when $l_n \to \infty$ and $n \to \infty$ we have

$$nh^{2\beta+1} S_2 \leq \frac{C}{h^{1-\frac{2}{\nu}}\, l_n^\rho} \sum_{j=l_n+1}^{n-1} j^\rho [\alpha(j)]^{1-\frac{2}{\nu}} \to 0 \tag{4.20}$$

Combining (4.17) with (4.20), and using (4.11) we conclude that Lemma 4.1.1 follows:

$$\lim_{n\to\infty} nh^{2\beta+1}\mathrm{Var}[\hat{f}_Z^{IP}(z)] = \lim_{n\to\infty} nh^{2\beta+1}(I_{n,0} + S_0) = \lim_{n\to\infty} nh^{2\beta+1} I_{n,0} = \sigma_0^2(z) \tag{4.21}$$

□

We now are ready to prove Theorem 4.1.1.

Proof of Theorem 4.1.1:

Proof. In order to establish the asymptotic normality of $\hat{f}_Z^{IP}(z)$ for the dependent sequence, a standard method is using the classical big block-small block argument [3]. The procedure is as follows: define

$$\widetilde{U}_j = \frac{h^{\frac{2\beta+1}{2}}}{\sigma_0(z)} \left(U_j - EU_j\right) \tag{4.22}$$

recall $U_j = \delta_j\, K^E[(z - W_j)/h]/[h\,(1 - G(W_j))]$ and $\hat{f}_Z^{IP}(z) = \sum_{j=1}^n U_j/n$. It will be equivalent to show

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \widetilde{U}_j \xrightarrow{\mathrm{d}} N(0,1) \tag{4.23}$$

First, we need to partition the set $\{1, 2, \cdots, n\}$ into $2k+1$ subsets with large blocks of size q_n and small blocks of size r_n, such that $k = k_n = [n/(q_n + r_n)]$. And the remaining block has size $n - k(q_n + r_n)$. Thus,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \widetilde{U}_j = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} \eta_j + \frac{1}{\sqrt{n}} \sum_{j=1}^{k} \eta'_j + \frac{1}{\sqrt{n}} \eta''_k \tag{4.24}$$

where

$$\eta_j = \sum_{i=(j-1)(q_n+r_n)+1}^{(j-1)(q_n+r_n)+q_n} \widetilde{U}_i \quad (j = 1, 2, \cdots, k) \tag{4.25}$$

$$\eta'_j = \sum_{i=(j-1)(q_n+r_n)+q_n+1}^{j(q_n+r_n)} \widetilde{U}_i \tag{4.26}$$

$$\eta''_k = \sum_{i=k(q_n+r_n)+1}^{n} \widetilde{U}_i \tag{4.27}$$

To prove the main theorem, we need to check the following conditions of big block-small block procedure.

$$\frac{1}{n} E\left(\sum_{j=1}^{k} \eta'_j\right)^2 \to 0, \quad \frac{1}{n} E(\eta''_k)^2 \to 0 \tag{4.28}$$

$$\left| E\left(e^{\mathrm{i}t\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k} E\left(e^{\mathrm{i}t\eta_j}\right) \right| \to 0 \tag{4.29}$$

$$\frac{1}{n} \sum_{j=1}^{k} E(\eta_j^2) \to 1 \tag{4.30}$$

$$\frac{1}{n} \sum_{j=1}^{k} E\left[\eta_j^2 \, I\{|\eta_j| > \epsilon\sqrt{n}\}\right] \to 0 \tag{4.31}$$

for every $\epsilon > 0$.

In the big block-small block procedure, (4.28) implies $\sum_{j=1}^{k} \eta'_j/\sqrt{n}$, $\eta''_k/\sqrt{n}$ are asymptotic negligible; (4.29) states η_j is asymptotic independent in the sense of characteristic function; (4.30) and (4.31) are the standard Lindeberg-Feller conditions for asymptotic normality of $\sum_{j=1}^{k} \eta_j/\sqrt{n}$ under independence.

Now follow the big block-small block procedure, we first need to find the block size. By assumption $r_n = o(nh^{1/2})$, there exist integers $s_n \to \infty$ such that $s_n\, r_n = o(nh^{1/2})$, $s_n(n/h)^{1/2}\alpha(r_n) \to 0$ as $n \to \infty$. Now let $q_n = [nh^{1/2}/s_n]$, we obtain the following relationships as $n \to \infty$:

$$\frac{r_n}{q_n} \to 0 \tag{4.32}$$

$$\frac{q_n}{n} \to 0 \tag{4.33}$$

$$\frac{q_n}{(nh)^{\frac{1}{2}}} \to 0 \tag{4.34}$$

$$\frac{n}{q_n}\,\alpha(r_n) \to 0 \tag{4.35}$$

Next we will check the big block-small block procedure (4.28)-(4.31).

Step 1 (1). Check for the first part of (4.28): $E\left(\sum_{j=1}^{k}\eta'_j\right)^2/n \to 0$.

Notice $E\widetilde{U}_j = 0$, thus $E\eta_j = E\eta'_j = E\eta''_k = 0$. Observe that

$$E\left(\sum_{j=1}^{k}\eta'_j\right)^2 = \text{Var}\left(\sum_{j=1}^{k}\eta'_j\right) = \sum_{j=1}^{k}\text{Var}\eta'_j + 2\sum_{1\le i<j\le k}\text{Cov}(\eta'_i,\eta'_j) := J_1 + J_2 \tag{4.36}$$

To simplify the following proof, we let $\zeta_j = (j-1)(q_n + r_n) + q_n$, therefore each term of J_1 is

$$\text{Var}\eta'_j = \sum_{i=\zeta_j+1}^{\zeta_j+r_n}\text{Var}\widetilde{U}_i + 2\sum_{1\le i_1<i_2\le r_n}\text{Cov}(\widetilde{U}_{\zeta_j+i_1},\widetilde{U}_{\zeta_j+i_2}) \tag{4.37}$$

Similar to the proof of $I_{n,0}$ and S_0 in (4.7) in Lemma 4.1.1, the variance and the covariance part are:

$$\sum_{i=\zeta_j+1}^{\zeta_j+r_n}\text{Var}\widetilde{U}_i = \sum_{i=\zeta_j+1}^{\zeta_j+r_n}E\widetilde{U}_i^{\,2} = \frac{h^{2\beta+1}}{\sigma_0^2(z)}\sum_{i=\zeta_j+1}^{\zeta_j+r_n}\text{Var}U_i = r_n(1+o(1)) \tag{4.38}$$

$$\begin{aligned} 2\sum_{1\le i_1<i_2\le r_n}\text{Cov}(\widetilde{U}_{\zeta_j+i_1},\widetilde{U}_{\zeta_j+i_2}) &= \frac{2h^{2\beta+1}}{\sigma_0^2(z)}\sum_{1\le i_1<i_2\le r_n}\text{Cov}(U_{\zeta_j+i_1},U_{\zeta_j+i_2}) \\ &= \frac{1}{\sigma_0^2(z)}o(r_n^2\,h^{2\beta+1}/(r_n\,h^{2\beta+1})) = \frac{r_n}{\sigma_0^2(z)}\,o(1) \end{aligned} \tag{4.39}$$

Hence $\text{Var}\eta'_j = r_n(1 + o(1) + o(1)/\sigma_0^2(z))$, and

$$J_1 = k\,r_n(1+o(1)) \tag{4.40}$$

follows from (4.36).

Now check J_2,

$$J_2 = 2 \sum_{1 \le i < j \le k} \text{Cov}(\eta_i', \eta_j') = 2 \sum_{1 \le i < j \le k} \sum_{l_1=1}^{r_n} \sum_{l_2=1}^{r_n} \text{Cov}(\widetilde{U}_{\zeta_i + l_1}, \widetilde{U}_{\zeta_j + l_2}) \tag{4.41}$$

Notice $i < j$, $|\zeta_j + l_2 - \zeta_i - l_1| \ge q_n$, then by (4.22) and similar to the proof of S_0 in Lemma 4.1.1 we have

$$\begin{aligned} |J_2| &\le \frac{2h^{2\beta+1}}{\sigma_0^2(z)} \sum_{1 \le i < j \le k} \sum_{l_1=1}^{r_n} \sum_{l_2=1}^{r_n} \left| \text{Cov}(U_{\zeta_i + l_1}, U_{\zeta_j + l_2}) \right| \\ &= \frac{2n\, h^{2\beta+1}}{\sigma_0^2(z)} o(h^{-1-2\beta}) = \frac{o(n)}{\sigma_0^2(z)} \end{aligned} \tag{4.42}$$

As a result, by (4.36), (4.40), (4.42) and (4.32),

$$\frac{1}{n} E \left(\sum_{j=1}^{k} \eta_j' \right)^2 = \frac{1}{n}(J_1 + J_2) \le \frac{1}{n} \left(k\, r_n (1 + o(1)) + \frac{o(n)}{\sigma_0^2(z)} \right) \to 0 \tag{4.43}$$

Step 1 (2). After that, we consider the second part of (4.28): $E(\eta_k'')^2/n \to 0$.

Similar to the proof of J_1, when $n \to \infty$,

$$\begin{aligned} \frac{1}{n} E(\eta_k'')^2 &= \frac{1}{n} \text{Var}(\eta_k'') = \frac{1}{n} \text{Var} \left(\sum_{i=k(q_n+r_n)+1}^{n} \widetilde{U}_i \right) \\ &= \frac{1}{n} \sum_{i=k(q_n+r_n)+1}^{n} \text{Var} \widetilde{U}_i + \frac{2}{n} \sum_{k(q_n+r_n)+1 \le i_1 < i_2 \le n} \text{Cov}(\widetilde{U}_{i_1}, \widetilde{U}_{i_2}) \\ &= \frac{n - k(q_n + r_n)}{n} (1 + o(1)) \to 0 \end{aligned} \tag{4.44}$$

Step 2. Check for (4.29): $\left| E\left(e^{\text{it} \sum_{j=1}^{k} \eta_j}\right) - \prod_{j=1}^{k} E\left(e^{\text{it}\eta_j}\right) \right| \to 0$.

Using Lemma 4.0.2, we have η_j is $\mathcal{F}_{i_l}^{j_l}$-measurable with $i_l = (l-1)(q_n + r_n) + 1$, $j_l = (l-1)(q_n + r_n) + q_n$, and note that $i_{l+1} - j_l = r_n + 1$. Hence by (4.35),

$$\begin{aligned} \left| E\left(e^{\text{it} \sum_{j=1}^{k} \eta_j}\right) - \prod_{j=1}^{k} E\left(e^{\text{it}\eta_j}\right) \right| &\le 16\,(k-1)\,\alpha(r_n + 1) \le 16\, k\, \alpha(r_n) \\ &< 16\, \frac{n}{q_n}\, \alpha(r_n) \to 0 \end{aligned} \tag{4.45}$$

Step 3. Check for (4.30): $\sum_{j=1}^{k} E(\eta_j^2)/n \to 1$.

Similar to the proof of J_1 in (4.36), we have by (4.32),

$$\frac{1}{n}\sum_{j=1}^{k} E(\eta_j^2) = \frac{1}{n}\sum_{j=1}^{k} \text{Var}\eta_j \to \frac{k}{n}\, q_n \ (1+o(1)) \to 1 \tag{4.46}$$

Step 4. It remains to check (4.31): $\sum_{j=1}^{k} E\left[\eta_j^2\, I\{|\eta_j| > \epsilon\sqrt{n}\}\right]/n \to 0$.

Recall $U_j = \delta_j\, K^E[(z-W_j)/h]/[h\,(1-G(W_j))]$ and $\widetilde{U}_j = h^{(2\beta+1)/2}\,(U_j - EU_j)/\sigma_0(z)$. The upper bound of $|\widetilde{U}_j|$ can be found as

$$|\widetilde{U}_j| \le \frac{h^{\frac{2\beta+1}{2}}}{\sigma_0(z)} \left\{ \frac{1}{h} \sup_z \left| K^E\left(\frac{z-W_j}{h}\right) \frac{\delta_j}{1-G(W_j)} \right| + |EU_j| \right\} \tag{4.47}$$

From [7] (Proof of Theorem 2.1), we know $|K^E(u)| \le C/h^{\beta}$, C is some positive constant. And by the assumption $\sup_x 1/[1-G(x)] \le B_0$ and $|EU_j| = \mathcal{O}(1)$ from (3.13),

$$|\widetilde{U}_j| \le \frac{h^{\frac{2\beta+1}{2}}}{\sigma_0(z)} \left(\frac{B_0\, C}{h\, h^{\beta}} + \mathcal{O}(1) \right) \le \frac{C}{h^{\frac{1}{2}}} \tag{4.48}$$

uniformly in j. By the definition of η_j in (4.25) and using (4.34),

$$\max_{1\le j\le k} \frac{|\eta_j|}{\sqrt{n}} \le \frac{C q_n}{\sqrt{nh}} \to 0 \tag{4.49}$$

which implies that $P(|\eta_j| > \epsilon\sqrt{n}) = 0$. Hence (4.31) is verified by

$$\frac{1}{n}\sum_{j=1}^{k} E[\eta_j^2 I\{|\eta_j| > \epsilon\sqrt{n}\}] \le \frac{1}{n}\sum_{j=1}^{k} \left(\frac{C q_n}{h^{\frac{1}{2}}}\right)^2 P(|\eta_j| > \epsilon\sqrt{n}) \to 0 \tag{4.50}$$

We conclude the asymptotic normality of $\hat{f}_Z^{IP}(z)$ holds since all the conditions (4.28)-(4.31) of big block-small block are proved. □

4.2 Case II: Supersmooth Distribution

Under this part we will establish the asymptotic normality of $\hat{f}_Z^{IP}(z)$ when the smoothness of error distribution is supersmooth of order β, for which φ_{f_E} satisfies Assumption 3.0.1 (ii), and we also assume that:

Assumption 4.2.1.

(i) $\varphi_K(t)$ *has a finite support* $(-d, d)$. *Moreover,* $|\varphi_K(t)| \leq B_1(d-t)^s$ *and* $\varphi_K(t) \geq B_2(d-t)^s$ *for* $t \in (d-\psi, d)$*, where* s *and* ψ *are some positive constants.*

(ii) $\int_{-\infty}^{\infty} f_{X|\delta=1}(x)/[1-G(x)]^{\nu}\, dx < \infty$ *for some* $\nu > 2$.

(iii) $\sum_{j=1}^{\infty} j^{\lambda}\, [\alpha(j)]^{1-2/\nu} < \infty$ *for some* $\nu > 2$ *and* $\lambda > 0$.

Let $\sigma^2(z) = \mathrm{Var}\hat{f}_Z^{IP}(z)$, $\bar{\sigma}_n^2(z) = \mathrm{Var}(U_j)$. Main theorem is listed below:

Theorem 4.2.1. *Suppose Assumption 4.2.1 is satisfied. Assume* $h = h_n \to 0$ *such that* $nh^{\omega} \to \infty$ *as* $n \to \infty$ *for some* $\omega > 1$. *If* $r_n = [(nh^{\omega})]^{1/2}$ *and* $\alpha(p)$ *satisfies* $(n/h^{\omega})^{1/2}\, \alpha(r_n) \to 0$ *as* $n \to \infty$*, then*

$$\frac{\sqrt{n}\left[\hat{f}_Z^{IP}(z) - E\hat{f}_Z^{IP}(z)\right]}{\bar{\sigma}_n(z)} \xrightarrow{d} N(0,1) \tag{4.51}$$

as $n \to \infty$.

Before we proceed the proof of Theorem 4.2.1, we need the following lemmas: Lemma 4.2.1 is to find the bounds of $K^E(x)$ [8], Lemma 4.2.2 finds the lower bound of $\bar{\sigma}_n^2(z)$, and Lemma 4.2.3 provides the asymptotic relationship between $\sigma^2(z)$ and $\bar{\sigma}_n^2(z)$.

Lemma 4.2.1. *Under Assumption 4.2.1, as* $n \to \infty$,

$$\|K^E\|_{\infty} = \mathcal{O}(h^{(s+1)\beta+\beta_0}\ [\ln(1/h)]^s\ e^{(d/h)^{\beta}/\gamma}) \tag{4.52}$$

$$|K^E(x)| \geq B_3\, \widetilde{H}(x)\, h^{(s+1)\beta+\beta_0}\, e^{\left(\frac{d}{h}\right)^{\beta}/\gamma} \tag{4.53}$$

uniformly in x *on a bounded interval, where* B_3 *is a positive constant, and*

$$\widetilde{H}(x) = \begin{cases} |\cos\ dx|, & if\ I_E(t) = o(R_E(t)) \\ |\sin\ dx|, & if\ R_E(t) = o(I_E(t)) \end{cases}$$

$R_E(t)$ *and* $I_E(t)$ *denote the real and the imaginary part of the characteristic function* $\varphi_{f_E}(t)$.

Lemma 4.2.2. *Under Assumption 4.2.1, we have as* $n \to \infty$

$$\bar{\sigma}_n^2(z) \geq B_4\, h^{2[(s+1)\beta+\beta_0-\frac{1}{2}]}\, e^{2\left(\frac{d}{h}\right)^{\beta}/\gamma} \tag{4.54}$$

for some positive constant B_4*. Let* $h = h_n = d\,(2/(\gamma\theta \ln n))^{1/\beta}$*, where* $\theta \in (0,1)$*.*

Proof. We have shown $|EU_j|$ is bounded in (3.13), therefore

$$\begin{aligned}
\bar{\sigma}_n^2(z) = \mathrm{Var} U_j &= \frac{1}{h^2}\, E\left[K^E\left(\frac{z-W_j}{h}\right)\frac{\delta_j}{1-G(W_j)}\right]^2 + \mathcal{O}(1) \\
&= \frac{p}{h^2}\int_{-\infty}^{\infty}\left[K^E\left(\frac{z-w}{h}\right)\right]^2 \frac{f_{X|\delta=1}(w)}{[1-G(w)]^2} dw + \mathcal{O}(1) \\
&\geq \frac{p}{h}\int_{-1}^{1}\left[K^E(u)\right]^2 \frac{f_{X|\delta=1}(z-h\,u)}{[1-G(z-h\,u)]^2}\, du + \mathcal{O}(1) \qquad (4.55)
\end{aligned}$$

By Lemma 4.2.1 and $f_{X|\delta=1}/[1-G]^2$ is continuous, as $h \to 0$,

$$\begin{aligned}
\bar{\sigma}_n^2(z) &\geq B_3^2\, h^{2[(s+1)\beta+\beta_0-\frac{1}{2}]}\, e^{2\left(\frac{d}{h}\right)^{\beta}/\gamma}\, \frac{f_{X|\delta=1}(z)}{[1-G(z)]^2}\int_{-1}^{1}\left[\widetilde{H}(u)\right]^2 du\,(1+o(1)) \\
&\geq B_4\, h^{2[(s+1)\beta+\beta_0-\frac{1}{2}]}\, e^{2\left(\frac{d}{h}\right)^{\beta}/\gamma} \qquad (4.56)
\end{aligned}$$

for some positive constant B_4. □

Lemma 4.2.3. *Under Assumption 4.2.1, let* h *take the same value in Lemma 4.2.2, we have*

$$\sum_{1\leq i<j\leq n} |\mathrm{Cov}(U_i, U_j)| = o(n\,\bar{\sigma}_n^2(z)) \tag{4.57}$$

$$\sigma^2(z) = \frac{1}{n}\bar{\sigma}_n^2(z)(1+o(1)) \tag{4.58}$$

Proof. First consider $\sigma^2(z)$,

$$\begin{aligned}
\sigma^2(z) &= \mathrm{Var}\,\hat{f}_Z^{IP}(z) \\
&= \frac{\sum_{j=1}^{n} \mathrm{Var} U_j}{n^2} + \frac{2}{n^2}\sum_{1\leq i<j\leq n} \mathrm{Cov}(U_i, U_j)
\end{aligned}$$

$$\leq \frac{\sum_{j=1}^{n} \text{Var}U_j}{n^2} + \frac{2}{n^2} \left| \sum_{1 \leq i < j \leq n} \text{Cov}(U_i, U_j) \right|$$

$$\leq \frac{\sum_{j=1}^{n} \text{Var}U_j}{n^2} + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} |\text{Cov}(U_i, U_j)| \tag{4.59}$$

Variance part can be expressed as

$$\frac{\sum_{j=1}^{n} \text{Var}U_j}{n^2} = \frac{\text{Var}U_j}{n} = \frac{\bar{\sigma}_n^2(z)}{n} \tag{4.60}$$

Before continuing, we separate the summation of covariance terms into two parts:

$$\begin{aligned} &\frac{2}{n^2} \sum_{1 \leq i < j \leq n} |\text{Cov}(U_i, U_j)| \\ &= \frac{2}{n^2} \sum_{\substack{1 \leq i < j \leq n \\ j - i \leq l_n}} |\text{Cov}(U_i, U_j)| + \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} |\text{Cov}(U_i, U_j)| \\ &:= J_3 + J_4 \end{aligned} \tag{4.61}$$

where $l_n = e^{(d/h)^\beta / \gamma}$.

We now show $J_3 \to o(\bar{\sigma}_n^2(z)/n)$.

$$\begin{aligned} J_3 &= \frac{2}{n^2} \sum_{\substack{1 \leq i < j \leq n \\ j - i \leq l_n}} |\text{Cov}(U_i, U_j)| \\ &\leq \frac{2}{n^2} \sum_{\substack{1 \leq i < j \leq n \\ j - i \leq l_n}} |E(U_i\, U_j)| + \frac{2}{n^2} \sum_{\substack{1 \leq i < j \leq n \\ j - i \leq l_n}} |E(U_i)|\, |E(U_j)| \end{aligned} \tag{4.62}$$

By (4.15) and similar to the proof Zu used in [24], using the assumption $\sup_x 1/[1 - G(x)] \leq B_0 < \infty$, $P(\delta_i = 1, \delta_j = 1) = p^*$, we have

$$\begin{aligned} &|E(U_i\, U_j)| \\ &= \frac{1}{h^2} \left| E \left[K^E \left(\frac{z - W_i}{h} \right) \frac{\delta_i}{1 - G(W_i)}\, K^E \left(\frac{z - W_j}{h} \right) \frac{\delta_j}{1 - G(W_j)} \right] \right| \\ &= \frac{1}{h^2} \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^E \left(\frac{z - w_i}{h} \right)\, K^E \left(\frac{z - w_j}{h} \right) \frac{f_{(X_i, X_j)|(\delta_i = 1, \delta_j = 1)}(w_i, w_j)}{[1 - G(w_i)][1 - G(w_j)]} \right. \end{aligned}$$

$$
\begin{aligned}
&\cdot P(\delta_i = 1, \delta_j = 1)\, dw_i\, dw_j| \\
&= \frac{p^*}{h^2} \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K^E\left(\frac{z-w_i}{h}\right)\ K^E\left(\frac{z-w_j}{h}\right) \frac{f_{(X_i,X_j)|(\delta_i=1,\delta_j=1)}(w_i,w_j)}{[1-G(w_i)][1-G(w_j)]}\ dw_i\, dw_j \right| \\
&\le \frac{B_0^2\, p^*}{h^2} \left| E\left[K^E\left(\frac{z-X_i}{h}\right) K^E\left(\frac{z-X_j}{h}\right) \middle| \delta_i = 1, \delta_j = 1 \right] \right| \\
&= \frac{B_0^2\, p^*}{(2\pi)^2}\, \frac{1}{h^2} \left| E\left[\int_{-d}^{d}\int_{-d}^{d} \frac{\varphi_K(u)\ \varphi_K(v)}{\varphi_{f_E}(\frac{u}{h})\ \varphi_{f_E}(\frac{v}{h})}\ e^{\mathrm{i}u\frac{X_i-z}{h}}\ e^{\mathrm{i}v\frac{X_j-z}{h}} du\ dv \middle| \delta_i = 1, \delta_j = 1 \right] \right|
\end{aligned} \tag{4.63}
$$

Note that for random variables R, S, T, $E\{E[R|S,T]|S\} = E[R|S]$ and recall $X_i = Z_i + E_i$, E_i is independent of Z_i and δ_i, using $\varphi_{f_E}(u/h) = E(\exp(\mathrm{i}uE_i/h))$, (4.63) becomes

$$
\begin{aligned}
&|E(U_i\, U_j)| \\
&\le \frac{B_0^2\, p^*}{(2\pi)^2}\, \frac{1}{h^2} \left| \int_{-d}^{d}\int_{-d}^{d} \frac{\varphi_K(u)\ \varphi_K(v)}{\varphi_{f_E}(\frac{u}{h})\ \varphi_{f_E}(\frac{v}{h})} \right. \\
&\quad \cdot\ E\left\{ E\left[e^{\mathrm{i}u\frac{Z_i+E_i-z}{h}}\ e^{\mathrm{i}v\frac{Z_j+E_j-z}{h}} \middle| (Z_i, Z_j, \delta_i = 1, \delta_j = 1) \right] \middle| \delta_i = 1, \delta_j = 1 \right\} du\ dv \Bigg| \\
&\le \frac{B_0^2\, p^*}{(2\pi)^2}\, \frac{1}{h^2} \left| \int_{-d}^{d}\int_{-d}^{d} \frac{\varphi_K(u)\ \varphi_K(v)}{\varphi_{f_E}(\frac{u}{h})\ \varphi_{f_E}(\frac{v}{h})} \varphi_{f_E}\left(\frac{u}{h}\right)\ \varphi_{f_E}\left(\frac{v}{h}\right) \right. \\
&\quad \cdot E\left(e^{\mathrm{i}u\frac{Z_i-z}{h}}\ e^{\mathrm{i}v\frac{Z_j-z}{h}} |\delta_i = 1, \delta_j = 1 \right) du\ dv \Bigg| \\
&\le \frac{B_0^2\, p^*}{(2\pi)^2}\, \frac{1}{h^2} \left| \int_{-d}^{d}\int_{-d}^{d} |\varphi_K(u)\ \varphi_K(v)| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left| e^{\mathrm{i}u\frac{z_i-z}{h}}\ e^{\mathrm{i}v\frac{z_j-z}{h}} \right| \right. \\
&\quad \cdot\ f_{(Z_i,Z_j)|(\delta_i=1,\delta_j=1)}(z_i, z_j)\, dz_i\, dz_j\ du\ dv \Bigg| \\
&\le \frac{C}{h^2} \left| \int_{-d}^{d}\int_{-d}^{d} |\varphi_K(u)\ \varphi_K(v)|\ du\ dv \right| \le \frac{C}{h^2}
\end{aligned} \tag{4.64}
$$

where C is some positive constant, (4.64) follows from Assumption 4.2.1 (i), $|\varphi_K(t)| \le B_1(d-t)^s$. Now (4.62) becomes $J_3 \le 2\, l_n(C/h^2 + \mathcal{O}(1))/n = \mathcal{O}(l_n/(nh^2))$ by $|EU_j| = \mathcal{O}(1)$ in (3.13). Using the lower bound of $\bar{\sigma}_n^2(z)$ in Lemma 4.2.2, we have

$$
\frac{n\, J_3}{\bar{\sigma}_n^2(z)} \le \frac{C\, l_n}{B_4\, h^2\, h^{2[(s+1)\beta+\beta_0-\frac{1}{2}]}\, e^{2\left(\frac{d}{h}\right)^\beta/\gamma}} = \mathcal{O}(h^{-2[(s+1)\beta+\beta_0+1/2]}\, e^{-(d/h)^\beta/\gamma}) \to 0 \tag{4.65}
$$

Next, it remains to show $J_4 \to o(\bar{\sigma}_n^2(z)/n)$. By Lemma 4.0.1 and Assumption 4.2.1 (ii),

$$\begin{aligned} J_4 &= \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} |\text{Cov}(U_i, U_j)| \\ &\le \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} 8\,[\alpha(j-i)]^{1-\frac{2}{\nu}} \left\{E|U_i|^\nu \; E|U_j|^\nu\right\}^{\frac{1}{\nu}} \\ &= \frac{2}{n^2} \sum_{i=1}^{n-l_n-1} \sum_{j=i+l_n+1}^{n} \frac{8}{h^2}\,[\alpha(j-i)]^{1-\frac{2}{\nu}} \left\{\int_{-\infty}^{\infty} \left|K^E\left(\frac{z-w}{h}\right)\right|^\nu \frac{f_{X|\delta=1}(w)\,p}{[1-G(w)]^\nu} dw\right\}^{\frac{2}{\nu}} \\ &\le \frac{C}{n^2h^2} \sum_{j=l_n+1}^{n-1} (n-j)[\alpha(j)]^{1-\frac{2}{\nu}} \|K^E\|_\infty^2 \\ &\le \frac{C}{nh^2}\,\|K^E\|_\infty^2\,\frac{1}{l_n^\lambda} \sum_{j=l_n+1}^{n-1} j^\lambda\,[\alpha(j)]^{1-\frac{2}{\nu}} \end{aligned} \tag{4.66}$$

Using the upper bound of $\|K^E\|_\infty$ in Lemma 4.2.1 and the lower bound of $\bar{\sigma}_n^2(z)$ in Lemma 4.2.2, we obtain

$$\begin{aligned} \frac{nJ_4}{\bar{\sigma}_n^2(z)} &\le \frac{C\;h^{2[(s+1)\beta+\beta_0]}\;\left[\ln\left(\frac{1}{h}\right)\right]^{2s}\;e^{2\left(\frac{d}{h}\right)^\beta/\gamma}}{h^2\,l_n^\lambda\,h^{2[(s+1)\beta+\beta_0-\frac{1}{2}]}\;e^{2\left(\frac{d}{h}\right)^\beta/\gamma}} \sum_{j=l_n+1}^{n-1} j^\lambda\,[\alpha(j)]^{1-\frac{2}{\nu}} \\ &= \frac{C}{h\,l_n^\lambda}\left[\ln\left(\frac{1}{h}\right)\right]^{2s} \sum_{j=l_n+1}^{n-1} j^\lambda\,[\alpha(j)]^{1-\frac{2}{\nu}} \to 0 \end{aligned} \tag{4.67}$$

which follows from Assumption 4.2.1 (iii) as $n \to \infty$.

Consequently, we conclude that (4.57) holds and combine with (4.60), (4.58) is proved. □

We now are ready to prove the asymptotic normality of $\hat{f}_Z^{IP}(z)$ for supersmooth case.

Proof of Theorem 4.2.1:

Proof. Recall $U_j = \delta_j\,K^E((z-W_j)/h)/(h(1-G(W_j)))$, $\hat{f}_Z^{IP}(z) = \sum_{j=1}^n U_j/n$, let $\widetilde{U}_j = (U_j - EU_j)/\bar{\sigma}_n(z)$. By Lemma 4.2.3,

$$\text{Var}\widetilde{U}_j = 1 \tag{4.68}$$

$$\sum_{1\le i<j\le n} \left|\mathrm{Cov}(\widetilde{U}_i, \widetilde{U}_j)\right| = o(n) \tag{4.69}$$

we obtain

$$\mathrm{Var}\left(\sum_{j=1}^{n} \widetilde{U}_j\right) = \sum_{j=1}^{n} \mathrm{Var}\widetilde{U}_j + 2 \sum_{1\le<i<j\le n} \mathrm{Cov}(\widetilde{U}_i, \widetilde{U}_j) \ = n\,(1+o(1)) \tag{4.70}$$

Note that

$$\begin{aligned} \sqrt{n}\ \frac{\left[\hat{f}_Z^{IP}(z) - E\hat{f}_Z^{IP}(z)\right]}{\bar{\sigma}_n(z)} &= \frac{\sqrt{n}\ \left[\frac{1}{n}\sum_{j=1}^{n} U_j - E\left(\frac{1}{n}\sum_{j=1}^{n} U_j\right)\right]}{\bar{\sigma}_n(z)} \\ &= \frac{\sqrt{n}\ \frac{1}{n}\sum_{j=1}^{n}(U_j - EU_j)}{\bar{\sigma}_n(z)} \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \frac{U_j - EU_j}{\bar{\sigma}_n(z)} = \frac{\sum_{j=1}^{n} \widetilde{U}_j}{\sqrt{n}} \end{aligned} \tag{4.71}$$

Thus it's equivalent to show $\sum_{j=1}^{n} \widetilde{U}_j/\sqrt{n} \xrightarrow{\mathrm{d}} N(0,1)$, we will still apply the big block-small block procedure as discussed before in ordinary smooth case in Theorem 4.1.1 to prove it. Let η_j, η'_j and η''_K defined same as in (4.25)-(4.27), but with $\widetilde{U}_j = (U_j - EU_j)/\bar{\sigma}_n(z)$.

Before we check the big block-small block procedure (4.28)-(4.31), let the large block size $q_n = [nh^{\omega_1}]^{1/2}$, where $1 < \omega_1 < \omega$, as $n \to \infty$

$$\frac{r_n}{q_n} \to 0, \qquad \frac{q_n}{n} \to 0 \tag{4.72}$$

$$\frac{q_n}{(nh)^{\frac{1}{2}}}\left[\ln\left(\frac{1}{h}\right)\right]^s \to 0 \tag{4.73}$$

$$\frac{n}{q_n}\,\alpha(r_n) \to 0 \tag{4.74}$$

Step 1 (1). Check for the first part of (4.28): $E\left(\sum_{j=1}^{k} \eta'_j\right)^2/n \to 0$.

As in (4.37), using (4.69) we know

$$\mathrm{Var}(\eta'_j) \le \sum_{i=\zeta_j+1}^{\zeta_j+r_n} \mathrm{Var}\widetilde{U}_i + r_n\,o(1) = r_n\,(1+o(1)) \tag{4.75}$$

By (4.36) we get $J_1 = \sum_{j=1}^{k} \mathrm{Var}\eta'_j \le k\, r_n\,(1+o(1))$. In (4.41), note that we now have (4.69), thus $|J_2| \le o(n)$, which leads to

$$\frac{1}{n}\,E\left(\sum_{j=1}^{k}\eta'_j\right)^2 = \frac{1}{n}(J_1+J_2) \le \frac{1}{n}\,[k\, r_n\,(1+o(1)+o(n)] \to 0 \tag{4.76}$$

as $n \to \infty$ by (4.72).

Step 1 (2). Next check the second part of (4.28): $E(\eta''_k)^2/n \to 0$.

Similarly, by (4.44), using (4.68) and (4.69), it can be established as

$$\begin{aligned}\frac{1}{n}E(\eta''_k)^2 &\le \frac{1}{n}[n-k(q_n+r_n)] + \frac{1}{n}[n-k(q_n+r_n)]\,o(1)\\ &= \frac{1}{n}[n-k(q_n+r_n)](1+o(1)) \to 0\end{aligned} \tag{4.77}$$

as $n \to \infty$.

Step 2. Check for (4.29): $\left|E\left(e^{\mathrm{it}\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k} E\left(e^{\mathrm{it}\eta_j}\right)\right| \to 0$.

As in (4.45), using (4.74),

$$\left|E\left(e^{\mathrm{it}\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k} E\left(e^{\mathrm{it}\eta_j}\right)\right| \le 16\,k\,\alpha(r_n) < 16\,\frac{n}{q_n}\,\alpha(r_n) \to 0 \tag{4.78}$$

Step 3. Check for (4.30): $\sum_{j=1}^{k} E\eta_j^2/n \to 1$.

By (4.75), with r_n replaced by q_n and using (4.72),

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{k} E\eta_j^2 = \frac{1}{n}\sum_{j=1}^{k}\mathrm{Var}\eta_j &\le \frac{1}{n}\,k\,q_n\,(1+o(1))\\ &= \frac{1}{n}\left[\frac{n}{q_n+r_n}\right]\,q_n(1+o(1)) \to 1\end{aligned} \tag{4.79}$$

Step 4. Finally, we verify the Lindeberg-Feller condition (4.31): $\sum_{j=1}^{k} E\left[\eta_j^2\, I\{|\eta_j| > \epsilon\sqrt{n}\}\right]/n \to 0$.

From (3.13) we know $|EU_j| = \mathcal{O}(1)$ and recall the assumption $\sup_x 1/[1-G(x)] \le B_0$, we derive

$$|\widetilde{U}_j| \le B_0\,\frac{\|K^E\|_\infty}{h\,\bar{\sigma}_n(z)} + \frac{\mathcal{O}(1)}{\bar{\sigma}_n(z)} \tag{4.80}$$

uniformly in j, so that

$$|\eta_j| \leq C\, q_n \, \frac{\|K^E\|_\infty}{h\, \bar{\sigma}_n(z)} \tag{4.81}$$

where C is some positive constant. Moreover, using the upper bound of $\|K^E\|_\infty$ in Lemma 4.2.1 and the lower bound of $\bar{\sigma}_n(z)$ in Lemma 4.2.2, we have from (4.73)

$$\frac{|\eta_j|}{\sqrt{n}} \leq \frac{C\, q_n\, h^{(s+1)\beta+\beta_0}\, \left[\ln\left(\frac{1}{h}\right)\right]^s\, e^{\left(\frac{d}{h}\right)^\beta/\gamma}}{\sqrt{n}\, h\, h^{(s+1)\beta+\beta_0-\frac{1}{2}}\, e^{\left(\frac{d}{h}\right)^\beta/\gamma}} = \frac{C\, q_n\, \left[\ln\left(\frac{1}{h}\right)\right]^s}{(nh)^{\frac{1}{2}}} \to 0 \tag{4.82}$$

Thus $P(|\eta_j|/\sqrt{n} > \epsilon) = 0$ for sufficient large n, (4.31) follows from (4.73)

$$\frac{1}{n} \sum_{j=1}^{k} E\left[\eta_j^2\, I\{|\eta_j| > \epsilon\, \sqrt{n}\}\right] \leq C\, \sum_{j=1}^{k} \frac{q_n^2\, \left[\ln\left(\frac{1}{h}\right)\right]^{2s}}{nh}\, P\left(\frac{|\eta_j|}{\sqrt{n}} > \epsilon\right) \to 0 \tag{4.83}$$

By summary, conditions (4.28)-(4.31) are satisfied, we conclude (4.51) holds, which completes the proof. □

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

Overall, knowing distribution is crucial for statistical inference, we find sufficient conditions for the asymptotic normality of $\hat{f}_Z^{IP}(z)$ under independence and strong mixing when the error distribution is either ordinary smooth or supersmooth.

The work in this dissertation could possibly be used to extend the necessary conditions for the Central Limit Theorem to hold for $\hat{f}_Z^{IP}(z)$ and the asymptotic normality of the estimator $\hat{f}_Z^{KM}(z)$.

REFERENCES

[1] Urinary tract infection (uti). https : / / www . hhs . gov / opa / reproductive-health / fact-sheets / urinary-tract-infections / index . html, May 2019.

[2] Michael Bedward. The truncated laplace distribution. https://rdrr.io/github/mbedward/walkies/man/trunc_laplace_distribution.html.

[3] Serge Bernstein. Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, 97(1):1–59, 1927.

[4] TK Chakrabarty. Deconvolving kernel density estimation of right censored duration data with recall errors. *American Journal of Applied Mathematics and Statistics*, 2(6):416–422, 2014.

[5] Aurore Delaigle. Code for computing the deconvolution kernel density estimator. https://researchers.ms.unimelb.edu.au/~aurored@unimelb//links.html#Code.

[6] Jean-Marie Dufour. Properties of moments of random variables. *Cours de l'Université de Montréal*, 2003.

[7] Jianqing Fan. Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 97–110, 1991.

[8] Jianqing Fan and Elias Masry. Multivariate regression estimation with errors-in-variables: asymptotic normality for mixing processes. *Journal of multivariate analysis*, 43(2):237–271, 1992.

[9] Jianqing Fan and Young K Truong. Nonparametric regression with errors in variables. *The Annals of Statistics*, pages 1900–1925, 1993.

[10] Wayne A Fuller. *Measurement error models*, volume 305. John Wiley & Sons, 2009.

[11] Peter Hall and Christopher C Heyde. *Martingale limit theory and its application.* Academic press, 2014.

[12] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[13] John Kloke and Joseph W McKean. *Nonparametric statistical methods using R.* CRC Press, 2014.

[14] Daniel Le. The impact of some characteristics of female patients on the recurrent urinary tract infections rate infections rate after cystoscopy with fulguration (cf) surgery. Master's thesis, The University of Texas at Arlington, United States, 2019.

[15] Elias Masry. Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory*, 37(4):1105–1115, 1991.

[16] Elias Masry. Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of multivariate analysis*, 44(1):47–68, 1993.

[17] Emanuel Parzen. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076, 1962.

[18] Kara Rogers. Why is it important to know the incubation period of a disease? `https://www.britannica.com/story/why-is-it-important-to-know-the-incubation-period-of-a-disease`, 2020.

[19] Glen A Satten and Somnath Datta. The kaplan–meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician*, 55(3):207–210, 2001.

[20] Geoffrey H Schuette. *Asymptotic Proporties of the Deconvolution Kernel Density Estimate based on 2-Dependent Error Scturcture with Applications to Remaining Useful Life Problems in Reliability Theorey*. PhD thesis, The University of Texas at Arlington, United States, 2018.

[21] Leonard A Stefanski and Raymond J Carroll. Deconvolving kernel density estimators. *Statistics*, 21(2):169–184, 1990.

[22] VA Volkonskii and Yu A Rozanov. Some limit theorems for random functions. i. *Theory of Probability & Its Applications*, 4(2):178–197, 1959.

[23] Matt P Wand and M Chris Jones. *Kernel smoothing*. Chapman and Hall/CRC, 1994.

[24] Yang Zu. A note on the asymptotic normality of the kernel deconvolution density estimator with logarithmic chi-square noise. *Econometrics*, 3(3):561–576, 2015.

BIOGRAPHICAL STATEMENT

Wenqing Zhu was born in Huai'an, Jiangsu, China in 1994. She received her B.S. degree in Statistics from Chongqing University in 2016, her Ph.D. degree from The University of Texas at Arlington in Mathematics with Concentration in Statistics in 2020. During her four years doctoral studies, she also served as a Graduate Teaching Assistant in the Department of Mathematics. She received the Stephen R. Bernfeld Scholarship, Bobbitt Family Endowed Scholarship for the College of Science in 2019 and Summer Dissertation Fellowship in 2020.