

A Comparative Study on the Performance of Gamma Kernels for  
Nonparametric Imputation

Mianbo Wang

A Thesis  
In the Department  
of  
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements  
For the Degree of  
Master of Science (Mathematics) at  
Concordia University  
Montreal, Quebec, Canada

August 2014

© Mianbo Wang, 2014

# CONCORDIA UNIVERSITY

## School of Graduate Studies

This is to certify that the thesis prepared

By: Mianbo Wang

Entitled: A Comparative Study on the Performance of Gamma Kernels for Nonparametric Imputation

and submitted in partial fulfillment of the requirements for the degree of

### Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

\_\_\_\_\_ Chair

Dr. A. Sen

\_\_\_\_\_ Examiner

Dr. A. Sen

\_\_\_\_\_ Examiner

Dr. D. Sen

\_\_\_\_\_ Thesis Supervisor

Dr. Y.P. Chaubey

Approved by

\_\_\_\_\_

Dr. J. Garrido

August 2014

\_\_\_\_\_

Dr. André Roy

# ABSTRACT

## A Comparative Study on the Performance of Gamma Kernels for Nonparametric Imputation

Mianbo Wang

The problems with using the symmetric kernels for nonparametric density and regression estimators for nonnegative data have been widely discussed. The use of asymmetric kernels for nonparametric regression, focusing on gamma kernels, have been recently proposed based on two different angles: one by [Chaubey et al. \(2010\)](#) and the other one by [Shi and Song \(2013\)](#). These estimators are based on the density estimators proposed by [Chaubey et al. \(2012\)](#) and [Chen \(2000\)](#). In the present thesis, we explore the performance of these estimators in the context of nonparametric imputation method under strongly missing at random assumption that has not been investigated yet in the literature. It is found that under certain assumption on the regression function, the estimator of [Chaubey et al. \(2010\)](#) may have a slight advantage over [Shi and Song \(2013\)](#) estimator whereas in other cases the comparison is not conclusive and further investigation may be needed.

## Acknowledgements

First, I would like to express my sincere gratitude to my supervisor Dr. Yogendra P. Chaubey, for his kind and continuous support on my thesis, also for his patience, motivation, enthusiasm and immense knowledge, all of which impressed me deeply.

Besides my advisor, I would like to thank all my instructors: Dr. Arusharka Sen, Dr. Debaraj Sen, Dr. Wei Sun and Dr. Abdellatif Bellahnid, for their excellent lectures and insightful comments. Also, I want to extend my thanks to Dr. José Garrido, Ms. Marie-France Leclere and Ms. Judy Thykootathil. They were always supportive during my study in Concordia.

I would like to express my special appreciation to Ms. Mengjiao Li for her constant encouragement and companion during my tough time. My sincere thanks also goes to all my dear friends in Montreal. They brought me a rich life experience during these two years and wish all of them have a wonderful future!

Last but not the least, I would like to express my deepest appreciation to my parents for their everlasting understanding and love, that I will never forget.

# Contents

<b>List of Figures</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Gamma kernel estimators</b>	<b>3</b>
2.1 A brief introduction about kernel function . . . . .	3
2.2 From symmetric kernels to asymmetric kernels . . . . .	4
2.3 Different modified gamma kernel estimators . . . . .	6
2.3.1 Gamma kernel proposed by Chaubey, Laïb and Sen (2010) . . . . .	6
2.3.2 Gamma kernel proposed by Shi and Song (2013) . . . . .	9
<b>3 Nonparametric imputation methods</b>	<b>12</b>
3.1 Missing data and imputation . . . . .	12
3.1.1 Missing data mechanism . . . . .	12
3.1.2 Strongly ignorable MAR assumption . . . . .	13
3.1.3 Different imputation approaches . . . . .	14
3.2 Nonparametric imputation method . . . . .	15
3.2.1 Kernel-weighted regression (KR) method . . . . .	15
3.2.2 Horvitz-Thompson (HT) inverse weighting method . . . . .	16
3.2.3 Double-robustness (DR) HT method . . . . .	17
<b>4 Numerical study</b>	<b>18</b>
4.1 Simulation study . . . . .	18

4.2	Empirical study . . . . .	19
4.3	Conclusions and future research . . . . .	27
4.3.1	Conclusions . . . . .	27
4.3.2	Future research . . . . .	27
	<b>References</b>	<b>31</b>
	<b>A R Code for Simulation Study</b>	<b>32</b>

# List of Figures

4.1	Regression of orthodontic growth data with KR method . . . . .	25
4.2	Regression of orthodontic growth data with DR method . . . . .	26

# Chapter 1

## Introduction

The boundary problem when a symmetric kernel is applied to estimate nonnegative data has been discussed in [Wand et al. \(1991\)](#), [Marron and Ruppert \(1994\)](#), [Bagai and Rao \(1996\)](#), and [Chaubey and Sen \(1996\)](#), especially in the context of density estimation. To alleviate these problems, some asymmetric kernel estimators have been proposed recently in the literature (see [Chen \(2000\)](#) and [Scaillet \(2004\)](#)). Asymmetric kernels are also suggested in the method of [Chaubey et al. \(2012\)](#) that is based on a stochastic approximation of the distribution function. These estimators may be adapted for nonparametric regression for nonnegative data. [Chaubey et al. \(2010\)](#) focus on the use of gamma kernels for nonparametric regression by adopting the density estimator proposed in [Chaubey et al. \(2012\)](#), whereas [Shi and Song \(2013\)](#) consider nonparametric regression based on gamma kernels by adopting the density estimator developed in [Chen \(2000\)](#). The method by [Chaubey et al. \(2012\)](#) is based on the kernel:

$$Q_{x+\epsilon_n, v_n}(t) = \frac{t^{1/v_n^2-1}}{v_n^2(x+\epsilon_n)^{1/v_n^2}\Gamma(1/v_n^2)} \exp\left(-\frac{t}{v_n^2(x+\epsilon_n)}\right) \quad (1.1.1)$$

where both  $v_n$  and  $\epsilon_n$  are smoothing parameters.

Whereas the kernel proposed in [Shi and Song \(2013\)](#) is given by,

$$K_{x,h}(t) = t^{x/h} \exp(-t/h) \quad (1.1.2)$$

where  $h$  represents the smoothing parameter.



In their paper, [Shi and Song \(2013\)](#) compared the finite sample performance of the gamma kernel regression estimate between both aforementioned estimators and implied that their estimator is a bit better than that of [Chaubey et al. \(2010\)](#). However, [Shi and Song \(2013\)](#) set the parameter  $v_n$  and  $\epsilon_n = v_n^2$  and used two exponential functions, both of which were not equal to 0 at  $x = 0$ , ignoring other cases.

The methods to handle missing data depend on the missing data mechanisms. Missing completely at random(MCAR), missing at random(MAR) and not missing at random(NMAR) are common missing data patterns proposed by [Little and Rubin \(2002\)](#). Two main approaches could be used to handle missing data: parametric imputation and nonparametric imputation, the latter of which will be discussed further under the strongly ignorable missing at random (MAR) assumption introduced by [Rosenbaum and Rubin \(1983\)](#).

The goal of this thesis is to compare the performance of both gamma kernels by applying them into the procedure of nonparametric imputation under strongly ignorable MAR assumption. Three nonparametric imputation methods will be applied: kernel-weighted regression method (see [Cheng and Wei \(1986\)](#), [Cheng \(1994\)](#)), Horvitz-Thompson inverse weighting method (see [Horvitz and Thompson \(1952\)](#)) and double-robustness HT method (see [Scaillet et al. \(1999\)](#)). Furthermore, the CLS estimator will be discussed in two situations:  $\epsilon_n = 0$  and  $\epsilon_n \neq 0$  through both simulation study and empirical study, in which the regression functions are different from previous research.

In chapter 2, a brief introduction of kernel functions and different asymmetric kernels will be given. The origin of two gamma kernels will be demonstrated in detail. A series of nonparametric imputation methods will be shown in Chapter 3 based on the background of missing data mechanism. In chapter 4, both gamma kernels will be applied into the nonparametric imputation methods in a simulation study, together with the orthodontic growth data in a empirical study. Besides, some conclusions and related future topics are also summarized in this chapter.

# Chapter 2

## Gamma kernel estimators

### 2.1 A brief introduction about kernel function

Let  $X_1, \dots, X_n$  be a random sample drawn from a distribution with unknown probability density function  $f_X$ . A common kernel estimator (see Silverman (1986), Wald and Jones (1995)) is used to estimate  $f_X$  as follows:

$$\hat{f}_X(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \quad (2.1.1)$$

where  $K_h(u) = \frac{1}{h}K(u/h)$ . Here  $K(\cdot)$  is a symmetric function satisfying  $\int K(x)dx = 1$ , which is called the *kernel*.  $h$  is a positive number, usually called the *bandwidth* or *window width*, which is very important to determine the smoothing applied. Usually  $K(\cdot)$  is chosen to be symmetric about zero but not necessarily positive function, which ensures  $\hat{f}_X(x)$  is itself also a density. Several types of kernel functions are commonly used as follows:

Uniform

$$K(u) = \frac{1}{2}I(|u| \leq 1),$$

Triangular

$$K(u) = (1 - |u|)I(|u| \leq 1),$$

Quartic

$$K(u) = \frac{15}{16}(1 - u^2)^2 I(|u| \leq 1),$$

Tricube

$$K(u) = \frac{70}{81}(1 - u^3)^3 I(|u| \leq 1),$$

Epanechnikov

$$K(u) = \frac{3}{4}(1 - u^2) I(|u| \leq 1),$$

Gaussian

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}.$$

## 2.2 From symmetric kernels to asymmetric kernels

The major disadvantage of applying a symmetric kernel estimator to a nonnegative data is that it puts positive weights outside the density support  $[0, \infty)$  and may cause the boundary problem that the expected value of  $\hat{f}(x)$  could not consistently estimate  $f(0)$ , especially when  $f(0) > 0$  such as the Exponential density (see [Silverman \(1986\)](#), [Wand et al. \(1991\)](#) for detail discussion).

Different approaches are proposed to solve this problem such as the transformation method. For example, if  $\log t$  denotes the transformation function, then we take  $\hat{f}_n(x) = (1/x)\hat{g}_n(\log x)$ , where  $\hat{g}_n$  is the kernel density estimator on the transformed data.

[Wand et al. \(1991\)](#) used the "back-transform" approach to estimate the density functions of the nonnegative random variables, which changes the variables of the global window width while [Marron and Ruppert \(1994\)](#) proposed a three-step computation-intensive transformation method.

[Bagai and Rao \(1996\)](#) proposed a kernel type estimator for  $f(x)$ , replacing the symmetric kernel by a bounded density function with nonnegative support, which is given by

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_{(i)}}{h_n}\right) \quad (2.2.1)$$

where  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  are ordered statistics. However, only the first  $r$  order statistics are used for estimating the density function  $f(x)$ , where  $X_r < x \leq X_{(r+1)}$ , which has an obviously undesirable feature for estimation because it does not include the whole data.

In contrast to the proposal by [Bagai and Rao \(1996\)](#), [Chaubey and Sen \(1996\)](#) proposed a new smooth density estimator

$$\tilde{S}_n = \sum_{k=0}^n w_{nk}(x, \lambda_n) S_n(k/\lambda_n), x \in \mathbb{R}^+ \quad (2.2.2)$$

where the nonnegative array  $w_{nk}(s, t)$  is

$$w_{nk}(s, t) = \{(st)^k/k!\} \left\{ \sum_{i=0}^n (st)^i/i! \right\}$$

and

$$\sum_{k=0}^n w_{nk}(s, t) = 1, \forall s, t \in \mathbb{R}^+$$

for nonnegative random variables based on Hille's lemma (see [Hille \(1948\)](#)) as follows.

**Lemma 2.2.1.** *If  $u(t)$  is a bounded, continuous function on  $\mathbb{R}^+$ , then as  $\lambda \rightarrow \infty$ ,*

$$e^{-\lambda t} \sum_{k \geq 0} u(k/\lambda) (t\lambda)^k/k! \rightarrow u(t)$$

*uniformly in any finite interval contained in  $\mathbb{R}^+$ ;*

A series of asymmetric kernels estimators were proposed to solve the problem caused by symmetric kernels. The first gamma kernel estimator which was considered by [Chen \(2000\)](#) has the following form:

$$\hat{f}(x) = n^{-1} \sum_{i=1}^n K_{\rho_h(x)}(X_i) \quad (2.2.3)$$

where

$$K_{\rho_h(x), h}(t) = \frac{t^{\rho_h(x)-1} e^{-t/h}}{h^{\rho_h(x)} \Gamma(\rho_h(x))} \quad (2.2.4)$$

For the function  $\rho_h(x)$ , two options were given and compared. One is

$$\rho_h(x) = x/h + 1 \quad (2.2.5)$$

which leads to the first estimator  $\hat{f}_1(x)$  given by [Chen \(2000\)](#). The other estimator  $\hat{f}_2(x)$  is led by

$$\rho_h(x) = \begin{cases} x/h, & \text{if } x \geq 2h; \\ \frac{1}{4}(x/h)^2 + 1, & \text{if } x \in [0, 2h). \end{cases} \quad (2.2.6)$$

which showed smaller MISE properties as that of  $\hat{f}_1(x)$ .

By utilising the inverse Gaussian density

$$K_{IG(m,\lambda)}(y) = \frac{\sqrt{\lambda}}{\sqrt{2\pi y^3}} \exp\left(-\frac{\lambda}{2m} \left(\frac{y}{m} - 2 + \frac{m}{y}\right)\right), \quad y > 0. \quad (2.2.7)$$

and reciprocal inverse Gaussian density

$$K_{RIG(m,\lambda)}(z) = \frac{\sqrt{\lambda}}{\sqrt{2\pi z}} \exp\left(-\frac{\lambda}{2m} \left(mz - 2 + \frac{1}{mz}\right)\right), \quad z > 0. \quad (2.2.8)$$

as kernels, [Scaillet \(2004\)](#) suggested two estimators of the probability density function

$$\hat{f}_{IG}(x) = n^{-1} \sum_{i=1}^n K_{IG(x,1/h)}(X_i), \quad (2.2.9)$$

and

$$\hat{f}_{RIG}(x) = n^{-1} \sum_{i=1}^n K_{RIG(1/(x-h),1/h)}(X_i), \quad (2.2.10)$$

However, the estimators of Scaillet are not consistent at  $x = 0$  (see [Chaubey et al. \(2010\)](#)).

## 2.3 Different modified gamma kernel estimators

### 2.3.1 Gamma kernel proposed by Chaubey, Laïb and Sen (2010)

In order to extend the results in [Chaubey and Sen \(1996\)](#), [Chaubey et al. \(2012\)](#) proposed a new density estimator generated from gamma function by using the following generalization of the Hille's lemma (see [Feller \(1965\)](#)).

**Lemma 2.3.1.** *Let  $u$  be any bounded and continuous function. Let  $G_{x,n}, n = 1, 2, \dots$  be a family of distributions with mean  $\mu_n(x)$  and variance  $h_n^2(x)$  then we have as  $\mu_n(x) \rightarrow x$*

and  $h_n(x) \rightarrow 0$

$$\tilde{u}(x) = \int_{-\infty}^{\infty} u(t) dG_{x,n}(t) \rightarrow u(x)$$

The convergence is uniform in every subinterval in which  $h_n(x) \rightarrow 0$  and  $u$  is uniformly continuous.

The density estimators suggested by [Chaubey et al. \(2012\)](#) are obtained from the derivatives of estimators of  $F(x)$  via smoothing the empirical distribution function. Adapted from this method, [Chaubey et al. \(2010\)](#) (CLS is short for the method by Chaubey, Laïb and Sen) proposed a generalized kernel smoothing technique to estimate the regression function in a class of nonnegative stationary ergodic processes.

To avoid the possible inconsistency at 0, their regression estimator used a perturbation term to a Gamma kernel as follows:

$$\tilde{m}_n(x) = \frac{\sum_{i=1}^n Q_{x+\epsilon_n, v_n}(X_i) Y_i}{\sum_{i=1}^n Q_{x+\epsilon_n, v_n}(X_i)}. \quad (2.3.1)$$

Here the modified gamma kernel is

$$Q_{x+\epsilon_n, v_n}(t) = \frac{t^{\alpha_n-1}}{\beta_{x+\epsilon_n}^{\alpha_n} \Gamma(\alpha_n)} \exp\left(-\frac{t}{\beta_{x+\epsilon_n}}\right) \quad (2.3.2)$$

where

$$\alpha_n = 1/v_n^2 \quad \text{and} \quad \beta_{x+\epsilon_n} = v_n^2(x + \epsilon_n)$$

The quantity  $v_n$  and  $\epsilon_n$  are two smoothing parameters, and  $\epsilon_n$  is a positive real number that goes to 0 at an appropriate rate as  $n \rightarrow \infty$ .

In this paper, a simpler form of the modified gamma kernel in [\(2.3.2\)](#) may be used, as given by

$$Q_{x+\epsilon_n, v_n}(X_i) = X_i^{1/v_n^2-1} \exp\left(-\frac{X_i}{v_n^2(x + \epsilon_n)}\right) \quad (2.3.3)$$

The proof of the uniform strong consistency of the regression estimator  $\tilde{m}_n$  is given in the following theorem (see [Chaubey et al. \(2010\)](#)), which holds true even for  $x = 0$ .

**Theorem 2.3.1.** *In addition to some necessary assumptions (see [Chaubey et al. \(2010\)](#) for detail), suppose that there exist sequences of real numbers  $M_n \rightarrow \infty$  and  $v_n \rightarrow \infty$  as  $n \rightarrow \infty$  that satisfy*

$$a_n^{-2} b_n^{\alpha_n - 1} \alpha_n^{7/2} M_n v_n^{-1} \rightarrow 0 \text{ as } n \rightarrow \infty$$

and

$$\sum_{n \geq 1} v_n \exp(-\pi \lambda^2 a_n^2 n M_n^{-2} v_n^2) < \infty$$

where,  $\lambda > 0$  and  $\alpha_n$  is defined in [\(2.3.2\)](#). Then we have

$$\sup_{x \in [a, b]} |\tilde{m}_n(x) - m(x)| = o_{a.s.}(1) \text{ as } n \rightarrow \infty.$$

The asymptotic normality of the regression estimator is given by the following theorem (see [Chaubey et al. \(2010\)](#)).

**Theorem 2.3.2.** *Assume that some necessary conditions hold (see [Chaubey et al. \(2010\)](#) for detail). Let  $Z_i = (X_i, Y_i)$  be a  $\mathbb{R}^+ \times \mathbb{R}^+$ -valued strictly stationary ergodic process defined on a probability space and let  $\phi$  be a Borel function of  $\mathbb{R}^+$  into  $\mathbb{R}$  such that  $E(|\phi(Y_1)|) < \infty$ .  $W_2(X)$  is the conditional expectation of  $\phi^2(Y_i)$  given  $X_{i-1}$ . Let  $\mathbb{F}$  be the  $\sigma$ -field generated by  $(X_k, Y_k)$ ,  $k = 1, \dots, i$  and the centralizing conditional parameter is defined as:*

$$\tilde{B}_n(x) = \frac{[\bar{h}_n(x + \epsilon_n) - h(x)] - m(x)[\bar{f}_n(x + \epsilon_n) - f(x)]}{\bar{f}_n(x + \epsilon_n)}$$

where

$$\bar{h}_n(x + \epsilon_n) = \frac{1}{n} \sum_{i=1}^n E[\phi(Y_i) Q_{x+\epsilon_n, v_n}(X_i) | \mathbb{F}_{i-1}]$$

$$\bar{f}_n(x + \epsilon_n) = \frac{1}{n} \sum_{i=1}^n E[Q_{x+\epsilon_n, v_n}(X_i) | \mathbb{F}_{i-1}].$$

(i) Let  $f(x) > 0$  at given  $x \geq 0$  and let

$$\sigma^2(x) = \frac{1}{2\sqrt{\pi}} \frac{W_2(x) - m^2(x)}{xf(x)}$$

Then, we have

$$\sqrt{nv_n}(\tilde{m}_n(x) - m(x) - \tilde{B}_n(x)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(x))$$

(ii) Suppose in addition the condition  $\sup_y \sum_{i=1}^{\infty} \|\mathcal{P}_1 f(y|\mathcal{F}_i)\|_2 < \infty$  is satisfied and that

$$nv_n^5 \rightarrow 0 \quad \text{and} \quad (nv_n)^{\frac{1}{2}} \epsilon_n \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

Then we have

$$\sqrt{nv_n}(\tilde{m}_n(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x))$$

(iii) Let  $x = 0$  and suppose moreover that

$$\epsilon_n v_n \rightarrow 0, \quad nv_n^5 \epsilon_n \rightarrow 0, \quad nv_n \epsilon_n^3 \rightarrow 0 \quad \text{and} \quad nv_n \epsilon_n \rightarrow \infty \quad \text{as } n \rightarrow \infty$$

Then

$$\sqrt{nv_n \epsilon_n}(\tilde{m}_n(0) - m(0)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_0^2(0))$$

where

$$\sigma_0^2(0) = \frac{1}{2\sqrt{\pi}} \frac{W_2(0) - m^2(0)}{f(0)}$$

whenever  $f(0) > 0$ .

### 2.3.2 Gamma kernel proposed by Shi and Song (2013)

Different from the gamma kernel density used in CLS,  $\hat{f}(x)$  defined in (2.2.3) has the definition at  $x = 0$  and this feature makes the estimators free from suffering from the boundary bias. Also [Chen \(2000\)](#) generated a different way to obtain the density estimator from CLS by directly using smooth underlying density. [Shi and Song \(2013\)](#) (SS stands for Shi and Song's method) extended [Chen \(2000\)](#)'s idea to nonparametric regression based on the kernel  $K_{x/h+1,h}$  by similar to the Nadaraya-Watson kernel regression (see [Nadaraya \(1964\)](#) and [Watson \(1964\)](#)).

The Nadaraya-Watson (N-W) estimator was proposed by [Nadaraya \(1964\)](#) and [Watson \(1964\)](#) to estimate  $m_n(x)$  as a locally average function, which is weighted by a kernel function:

$$\begin{aligned} \hat{m}(x) &= \frac{\sum_{i=1}^n K_h(x - x_i) y_i}{\sum_{i=1}^n K_h(x - x_i)} \\ &= \sum_{i=1}^n W_h(x, x_i) y_i \end{aligned} \tag{2.3.4}$$



where the weighted function  $W_h(x, x_i) = \frac{K_h(x-x_i)}{\sum_{i=1}^n K_h(x-x_i)}$ .

When the covariate  $X$  is nonnegative, the gamma kernel regression estimation of  $m(x)$  as defined by [Shi and Song \(2013\)](#) is give by

$$\hat{m}_n(x) = \frac{\sum_{i=1}^n K_{x/h+1,h}(X_i)Y_i}{\sum_{i=1}^n K_{x/h+1,h}(X_i)}. \quad (2.3.5)$$

A simpler expression for  $\hat{m}_n(x)$  in (2.3.4) could be derived from the definition of  $K_{x/h+1,h}$  by using the constants:

$$K_{x/h+1,h}(X_i) = (X_i)^{x/h} \exp(-X_i/h) \quad (2.3.6)$$

The asymptotic normality of  $\hat{m}_n(x)$  is shown in the following theorem (see [Shi and Song \(2013\)](#)).

**Theorem 2.3.3.** *Assume  $E(\varepsilon|X) = 0$ ; the second order derivative of  $f(x)$ ,  $f(x)m(x)$ ,  $f(x)m^2(x)$ ,  $\sigma^2(x)$ ,  $f(x)\sigma^2(x)$ ,  $E(|\varepsilon|^{2+\delta}|X = x)$  is continuous and bounded on  $[0, \infty)$  for some  $\delta > 0$ , where  $\sigma^2(x) = E(\varepsilon^2|X = x)$ ;  $h \rightarrow 0$ ,  $n\sqrt{h} \rightarrow \infty$  as  $n \rightarrow \infty$ . Then for any  $x \in (0, \infty)$  with  $f(x) > 0$ ,*

$$\left(\frac{v(x)}{n\sqrt{h}}\right)^{-1/2} [\hat{m}_n(x) - m(x) - hb(x) + o_p(h)] \xrightarrow{d} N(0, 1)$$

where  $b(x)$  and  $v(x)$  are defined as:

$$b(x) = m'(x) + \frac{1}{2}x''(x) + \frac{xm'(x)f''(x)}{f(x)}, \quad v(x) = \frac{\sigma^2(x)}{2f(x)\sqrt{\pi x}}$$

For  $x = 0$ , but  $f(0) > 0$ ,

$$\left(\frac{\sigma^2(0)}{2nhf(0)}\right)^{-1/2} [\hat{m}_n(0) - m(0) - hm'(0) + o^p(h)] \xrightarrow{d} N(0, 1)$$

If we further assume that  $\log n/(n\sqrt{h}) \rightarrow 0$ , then  $o_p(1)$  can be replaced by  $o(1)$  in the above results.

The following theorem (see [Shi and Song \(2013\)](#)) has given the uniform consistency of  $\hat{m}_n(x)$  to  $m(x)$  over the bounded sub-interval of  $(0, \infty)$ .

**Theorem 2.3.4.** *In addition to the the assumptions in theorem 2.3.3, assume that  $\log n/(n\sqrt{h}) \rightarrow 0$ . Then for any constants  $a$  and  $b$  such that  $0 < a < b < \infty$ ,*

$$\sup_{x \in [a, b]} |\hat{m}_n(x) - m(x)| = O(h) + o\left(\frac{\sqrt{\log n}}{\sqrt{n\sqrt{h}}}\right), \quad a.s.$$

# Chapter 3

## Nonparametric imputation methods

### 3.1 Missing data and imputation

#### 3.1.1 Missing data mechanism

Missing data mechanisms are very important because the property of missing data methods depends very strongly on the nature of the dependencies in the mechanisms. Three missing data mechanisms are defined by [Little and Rubin \(2002\)](#): Missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR).

Let  $Y = (y_{ij})$  denote an  $(n \times K)$  rectangular data set without missing values with  $i$ th row  $y_i = \{y_{ij}\}, j = 1, \dots, K$  where  $y_{ij}$  is the value of variable  $Y_j$  for subject  $i$ . With missing data, define the missing-data indicator matrix  $M = (m_{ij})$ , such that

$$m_{ij} = \begin{cases} 1, & \text{if } y_{ij} \text{ is missing,} \\ 0, & \text{if } y_{ij} \text{ is not missing,} \end{cases} \quad (3.1.1)$$

so the pattern of missing data is defined by the matrix  $M$ .

The missing-data mechanism is characterized by the conditional distribution of  $M$  given  $Y$ , say  $h(M|Y, \phi)$ , where  $\phi$  denotes unknown parameters. A missing-data mechanism is called missing completely at random (MCAR), if

$$h(M|Y, \phi) = h(M|\phi) \quad \forall Y \text{ \& } \phi \quad (3.1.2)$$

Here the missingness does not depend on the value of the data  $Y$ , no matter whether it is missing or observed.

Let  $Y_{mis}$  denote the missing components or entries of  $Y$ , and  $Y_{obs}$  the observed components. A missing-data mechanism is called missing at random (MAR) if

$$h(M|Y, \phi) = h(M|Y_{obs}, \phi) \quad \forall Y_{mis} \ \& \ \phi \quad (3.1.3)$$

This is a less restrictive assumption than that of MCAR, where the missingness depends only on the components  $Y_{obs}$  of  $Y$  that are observed, and not on the components that are missing.

The mechanism is called not missing at random (NMAR) if the distribution of  $M$  depends on the missing values in the data matrix  $Y$ , that is

$$h(M|Y, \phi) = h(M|Y_{mis}, \phi) \quad \forall Y_{obs} \ \& \ \phi \quad (3.1.4)$$

### 3.1.2 Strongly ignorable MAR assumption

A simple missing data pattern is created based on the well-known double (or two-stage) sampling design, which was first proposed by [Neyman \(1938\)](#). Let  $X$  be a  $p$ -dimensional vector of factors and  $Y$  be a response variable influenced by  $X$ . A random sample of incomplete data is often denoted by

$$(X_i, Y_i, \delta_i), \quad i = 1, 2, \dots, n \quad (3.1.5)$$

where all the  $X_i$ 's are observed and  $\delta_i=1$  if  $Y_i$  is observed, otherwise  $\delta_i=0$ . Sometimes, the two-stage sampling plan is used to take more observations on the covariate  $X$  if there is lack of enough  $Y$  observation due to some special constraints.

A convenient nonparametric inference for the missing data was proposed by [Rosenbaum and Rubin \(1983\)](#) based on the assumption that  $\delta$  and  $Y$  are conditionally independent given  $X$ , which was called "strongly ignorable MAR" assumption. Without any parametric inference on the joint distribution of  $(\delta, X, y)$ , it assumes that

$$P(\delta = 1|Y, X) = P(\delta = 1|X) = p(X) \quad (3.1.6)$$

where  $p(X)$  is the response propensity score function given  $X$ , that is the missing pattern function defined under MAR.

### 3.1.3 Different imputation approaches

There are two main parametric approaches employed in the missing data analysis: maximum likelihood (ML) and multiple imputation (MI), both of which are based upon the ignorable MAR assumption defined in (3.1.6). Facilitated by the EM algorithm, the ML procedure could be used for inference when a parametric model could be defined for all the variables (see [Dempster et al. \(1977\)](#)). Multiple imputation is another popular parametric approach to missing data problems proposed by [Rubin \(1987\)](#). In MI, by making random draws from the predictive distribution, each missing value is replaced by a list of  $M > 1$  values to produce  $M$  complete data sets. Then, each of the data sets is analyzed based on a complete-data based inference method, all of which are combined to form a final inference that reflects the uncertainty because of the nonresponse (see [Little and Rubin \(2002\)](#)). In addition to the parametric model for the complete data, a prior distribution for the parameters is used to generate the imputation that needs to be specified first. And then the Bayesian arguments are applied to simulate independent draws from the distribution of  $Y_{mis}$  given  $Y_{obs}$ , which is usually carried out by markov chain monte carlo (MCMC) computational techniques.

Under more relaxed assumptions, [Robin et al. \(1994\)](#) used the inverse probability weighting to estimate a semiparametric regression function with the parametrically estimated propensity scores, of which the efficiency bound for parameter estimation was established by [Robin and Rotnitzky \(1995\)](#). A remarkable advantage is that this approach is more robust against model misspecification, although a correct model is often needed to reach the semiparametric efficiency bound for the conditional distribution of the missing variable given the observed variable.

## 3.2 Nonparametric imputation method

There are two main nonparametric approaches with missing data. One is to extend complete data rank testing procedures, which is only valid under rather restrictive MCAR assumption (see [Brunner et al. \(1999\)](#) for detail). The other approach handles missing data by relaxing the rigorous parametric assumption in the parametric approach through different smoothing techniques. [Titterington and Mill \(1983\)](#) considered a nonparametric estimation of the joint density  $(X, Y)$  utilizing kernel method to generate the empirical versions of the joint distribution. The smoothing process is done by using the observed values of the incomplete case and the corresponding values of these variables in the complete cases. Next, the following three nonparametric imputation methods in the vein will be introduced and applied.

### 3.2.1 Kernel-weighted regression (KR) method

The kernel-weighted regression (KR) method is a basic nonparametric regression imputation scheme, which was first introduced by [Cheng and Wei \(1986\)](#) and refined in the literature by [Cheng \(1994\)](#). Let  $X$  and  $Y$  be a pair of real valued data defined in [\(3.1.5\)](#) and  $m(x) = E(Y|X = x)$  be the regression function of  $Y$  given  $x$ . The parameter  $\theta = EY$  could be estimated using N-W estimator defined in [\(2.3.4\)](#). Let  $K(\cdot)$  be a kernel function, and  $h = h(n)$  be a bandwidth sequence and  $h \rightarrow 0$  as  $n \rightarrow \infty$ . Using a local weighted least squares, the following quantity should be minimized for each real  $x$

$$\sum_{i=1}^n K_h(X_i, x) \delta_i (Y_i - m(x))^2$$

where  $K_h(u, x) = h^{-1}K((u - x)/h)$ . The minimizer of  $m(x)$  is

$$\hat{m}(x) = \frac{\sum_{i=1}^n K_h(X_i, x) \delta_i Y_i}{\sum_{i=1}^n K_h(X_i, x) \delta_i}. \quad (3.2.1)$$

where  $\delta_i$  is given in (3.1.5). An estimator of  $\hat{\theta}$  considered by Cheng and Wei (1986) is the sample average of all the regression estimates:

$$\hat{\theta} = n^{-1} \sum_{i=1}^n \hat{m}(X_i) \quad (3.2.2)$$

Alternatively, a natural approach to estimate  $\theta$  is

$$\tilde{\theta} = n^{-1} \sum_{i=1}^n [\delta_i Y_i + (1 - \delta_i) \hat{m}(X_i)] \quad (3.2.3)$$

Estimators (3.2.2) and (3.2.3) were proved to be asymptotically equivalent by the following theorem given by Cheng (1994).

**Theorem 3.2.1.** *Assume equation (3.1.6) for the missing data in equation (3.1.5). Both  $n^{1/2}(\hat{\theta} - \theta)$  and  $n^{1/2}(\tilde{\theta} - \theta)$  have the same asymptotic normal distribution with mean 0 and variance*

$$\sigma^2 = E(\sigma^2(X)/p(X)) + \text{var}(m(X))$$

where  $\sigma^2(X) = \text{var}(Y|X)$  is the conditional variance of  $Y$  given  $X$ .

### 3.2.2 Horvitz-Thompson (HT) inverse weighting method

The Horvitz-Thompson (HT) inverse weighting estimator proposed by Horvitz and Thompson (1952) is frequently applied in the analysis of stratified sampling to estimate a population parameter. By inverting the sampling weights to reflect the effective sample size, the classical HT weighting scheme can be used to recover the missing data information. Under MAR, the basic HT imputation estimator of  $\hat{\theta}$  is defined by inverting the estimated propensity score:

$$\hat{\theta} = n^{-1} \sum_{i=1}^n \frac{\delta_i Y_i}{w_i} \quad (3.2.4)$$

where

$$w_i = \frac{\sum_{j=1}^n \delta_j K_h(X_j, X_i)}{\sum_{j=1}^n K_h(X_j, X_i)}$$

Here  $w_i$  is used to estimate the propensity score of  $p(X_i)$  in (3.1.6) using the same kernel smoothed estimate as defined by (3.2.1).

### 3.2.3 Double-robustness (DR) HT method

Double-robustness (DR) property refers to the advantage that a method is asymptotically efficient when either the parametric regression model or the propensity score model is correctly specified, which is defined by [Scaillet et al. \(1999\)](#) and has been extensively used within the semiparametric model. The requirement of using the DR property into the nonparametric regression is that both the regression function and the propensity score should be ideally smooth functions (see [Ning and Cheng \(2012\)](#)). The double-robust HT estimator, which is the modified form of the basic HT estimator in [\(3.2.4\)](#) is defined as:

$$\hat{\theta} = n^{-1} \sum_{i=1}^n \left[ \hat{m}(X_i) + \frac{\delta_i(Y_i - \hat{m}(X_i))}{w_i} \right] \quad (3.2.5)$$

where  $\hat{m}(x)$  and  $w_i$  are defined in [\(3.2.1\)](#) and [\(3.2.4\)](#) respectively.



# Chapter 4

## Numerical study

### 4.1 Simulation study

A simulation study is used to evaluate the performance of the aforementioned gamma kernels in (2.3.1) by [Chaubey et al. \(2010\)](#) and (2.3.5) by [Shi and Song \(2013\)](#) in the context of three different imputation methods under strongly ignorable MAR assumption. A common regression model is considered for each kernel in each imputation method

$$Y = m(X) + \varepsilon \tag{4.1.1}$$

where the random error  $\varepsilon$  is a normal with mean 0 and variance  $0.5^2$ , which is independent of  $X$ . For  $m(X)$ , the following 4 models were considered

$$\left\{ \begin{array}{l} m_1(x) = 2x \\ m_2(x) = 3 - 12(x - 0.5)^2 \\ m_3(x) = x + 2 \exp(-x^2) \\ m_4(x) = \sin(2x) + 2 \exp(-x^2) \end{array} \right. \tag{4.1.2}$$

where  $X \sim U(0, 1)$ . The model 3 and model 4 are not equal to 0 at  $x = 0$ , used by [Shi and Song \(2013\)](#). To compare with the aforementioned regression models, model 1 and model 2 are chosen to be linear and quadratic, both of which are equal to 0 at  $x = 0$ . The

response propensity score function  $p(x)$  is defined with a logistic function used by [Ning and Cheng \(2012\)](#)

$$p(x) = \frac{e^{2.5x}}{1 + e^{2.5x}}. \quad (4.1.3)$$

For each simulation, a random number  $u_i \in [0, 1]$  is created and  $\delta_i$  in [\(3.1.5\)](#) is defined as:

$$\delta_i = \begin{cases} 1, & \text{if } u_i < \int_0^1 p(x)dx, \\ 0, & \text{otherwise.} \end{cases} \quad (4.1.4)$$

For CLS estimator defined in [\(2.3.1\)](#), the parameter  $\epsilon_n$  is chosen in two different ways:  $\epsilon_n = v_n^2$  and  $\epsilon_n = 0$ . The optimal  $h$  and  $v_n$  values are searched from 200 equally spaced grid points from  $[0.001, 1]$ . The sample sizes are selected to be 100, 150 and 200. For each sample size, the simulation procedure is replicated for 200 times. The average values of the minimum MSE are reported for each kernel within the different imputation methods in [Table 4.1](#).

From the simulation result, it is interesting to notice that for all the regression models, SS estimator performs almost the same as the CLS estimator with parameter  $\epsilon_n = v_n^2$  in all of three imputation methods since they have the similar MSE values for any sample size. In model 1 and model 2, for which  $m(0) = 0$ , the CLS estimator  $\epsilon_n = 0$  shows a bit advantage over the other two estimators with both KR and DR method, although this advantage is very limited. However, there is no preference among all three estimators for model 3 and model 4, both of which are not equal to 0 when  $x = 0$ .

## 4.2 Empirical study

Both gamma kernels will be applied into the nonparametric imputation procedure for the data set from [Potthoff and Roy \(1964\)](#) on the orthodontic growth measurements for 11 girls and 16 boys. The distance from the center of the pituitary to the maxillary fissure is recorded at the age of 12 and 14 years for each child. We assume that the finite population



mean of the 27 measures of all the boys and girls at age 14 is to be estimated under a MAR design. The measures at age 14 are the response  $Y$  values, and those at age 12 are the covariate  $X$  values. Our goal is to examine and compare the performance of applying both gamma kernels into different imputation methods, such as bias and MSE values. Some  $Y$  values will be deleted according to formulas (3.1.5) and (3.1.6) by defining a propensity score (see Ning and Cheng (2012))  $p(x)$  as

$$p(x) = \begin{cases} 0.9, & \text{if } x < 25, \\ 0.4, & \text{if } x \geq 25. \end{cases} \quad (4.2.1)$$

According to the propensity score, the expectation of the number of deletion is 7.7 for each simulation. Because of the small data size  $n = 27$ , the deletion process was simulated only 20 times. A typical simulated missing data set is presented in Table 4.2, where the missing  $Y$  value are quoted in parentheses.

For the 20 simulated data sets, both SS and CLS kernel estimators are applied into 3 nonparametric imputation methods. The parameter  $\epsilon_n$  is considered to be either equal to  $v_n^2$  or 0 respectively as that in the simulation study. The results are summarized in Table 4.3.

The result of empirical study seems to be consistent as well as the simulation study. When simulating the incomplete data, the HT method shows much larger bias and MSE values than KR and DR methods. Moreover, CLS kernel estimator with  $\epsilon_n = 0$  has a smaller bias and MSE values than SS estimator and CLS estimator with  $\epsilon_n = v_n^2$  in KR and DR imputation methods. We find the optimal bandwidth values for KR and DR method to compare the performances of SS estimator and CLS estimator with  $\epsilon_n = 0$  by splitting the interval  $[0.001, 1]$  into 500 grids. The optimal  $h$  values are 0.149 and 0.067 for corresponding estimator.

Thus, the regression procedure is applied different imputed data sets to compare the effects of using different kernels with the optimal bandwidth  $h$ . For convenience, the orthodontic growth data at age 12 years  $X$  will be transformed to  $X_1 = (X - 19)/2.8$ ,

Table 4.2: Orthodontic growth data for 11 girls and 16 boys

Girl	Age (in years)		Boy	Age (in years)	
	12( $X$ )	14( $Y$ )		12( $X$ )	14( $Y$ )
1	21.5	23.0	1	29.0	(31.5)
2	24.0	25.5	2	23.0	26.5
3	24.5	26.5	3	24.0	27.5
4	25.0	(26.5)	4	26.5	(27.0)
5	22.5	23.5	5	22.5	26.5
6	21.0	22.5	6	27.0	(28.5)
7	23.0	(25.0)	7	24.5	26.5
8	23.5	24.0	8	24.5	25.5
9	22.0	21.5	9	31.0	26.0
10	19.0	19.5	10	31.0	(31.5)
11	28.0	(28.0)	11	23.5	(25.0)
			12	24.0	28.0
			13	26.0	29.5
			14	25.5	26.0
			15	26.0	30.0
			16	23.5	25.0

<sup>1</sup> Data source: [Potthoff and Roy \(1964\)](#)

Table 4.3: A result of simulating the incomplete growth data with different kernels

		SS		CLS1		CLS2	
		Bias	MSE	Bias	MSE	Bias	MSE
KR	h						
	0.05	-0.229	3.946	-0.225	3.855	-0.223	3.395
	0.10	-0.226	3.738	-0.219	3.692	-0.195	3.074
	0.15	-0.207	3.486	-0.210	3.413	-0.201	3.320
	0.20	-0.247	3.813	-0.245	3.808	-0.231	3.792
	0.25	-0.285	4.097	-0.281	3.998	-0.269	3.810
HT	1.05	-2.373	597.14	-2.465	600.11	-2.451	599.85
	1.10	-2.342	597.03	-2.333	599.88	-2.321	599.77
	1.15	-2.311	596.98	-2.302	599.70	-2.292	599.74
	1.20	-2.282	596.96	-2.524	599.34	-2.646	599.79
	1.25	-2.255	597.03	-2.249	599.65	-2.238	599.86
DR	0.05	-0.191	2.109	-0.195	2.112	-0.182	2.012
	0.10	-0.178	1.953	-0.171	1.949	-0.139	1.848
	0.15	-0.156	1.878	-0.157	1.859	-0.124	1.787
	0.20	-0.210	2.302	-0.203	2.294	-0.176	2.108
	0.25	-0.234	2.460	-0.229	2.402	-0.217	2.215

<sup>1</sup> CLS1:  $\epsilon_n = v_n^2$ ; CLS2:  $\epsilon_n = 0$

where 19 is minimum of  $X$  and 2.8 is the standard deviation of  $X$ .

In Figure 4.1 and Figure 4.2, the solid line is the regression curve with normal kernel and the dotted line is the regression curve with CLS estimator when  $\epsilon_n = 0$ . we observe that the dotted line seems to be less likely to be off from the solid line so that the estimated curve for CLS kernel with  $\epsilon_n = 0$  captures a little more characteristic of the data structure than both SS estimator and CLS estimator with  $\epsilon_n = v_n^2$  in KR and DR imputation methods.

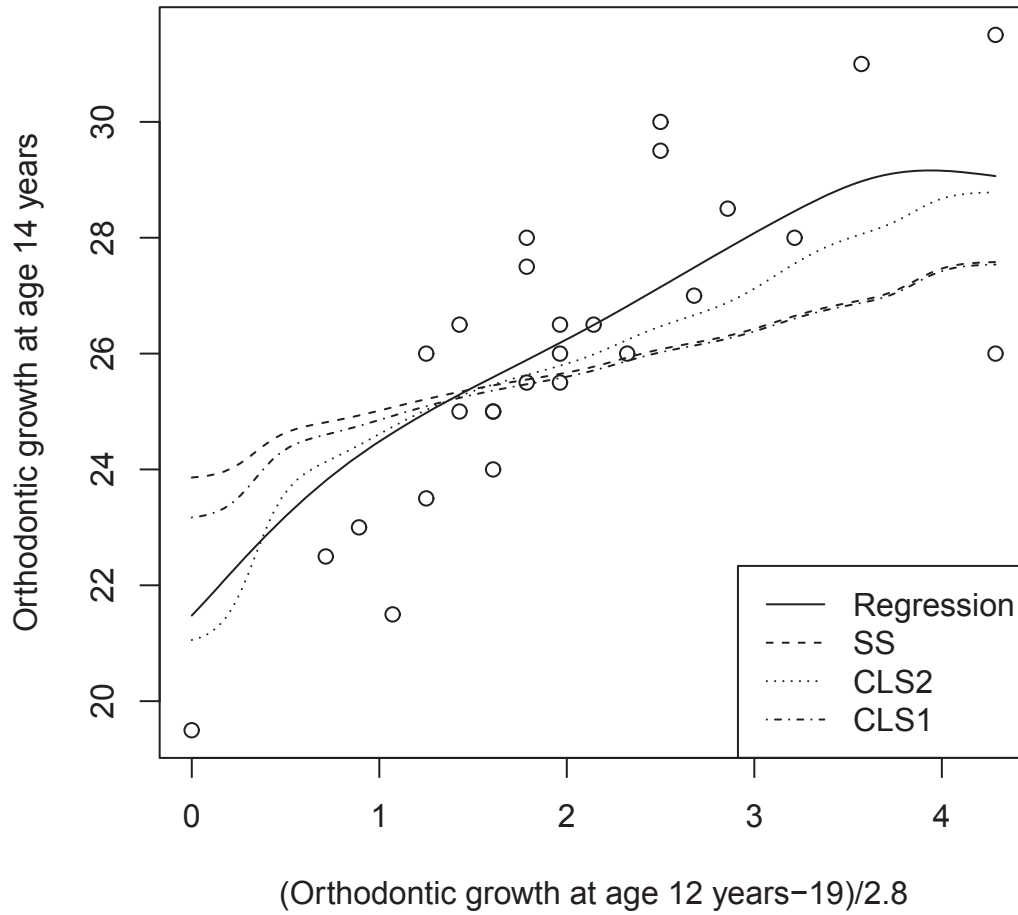


Figure 4.1: Regression of orthodontic growth data with KR method



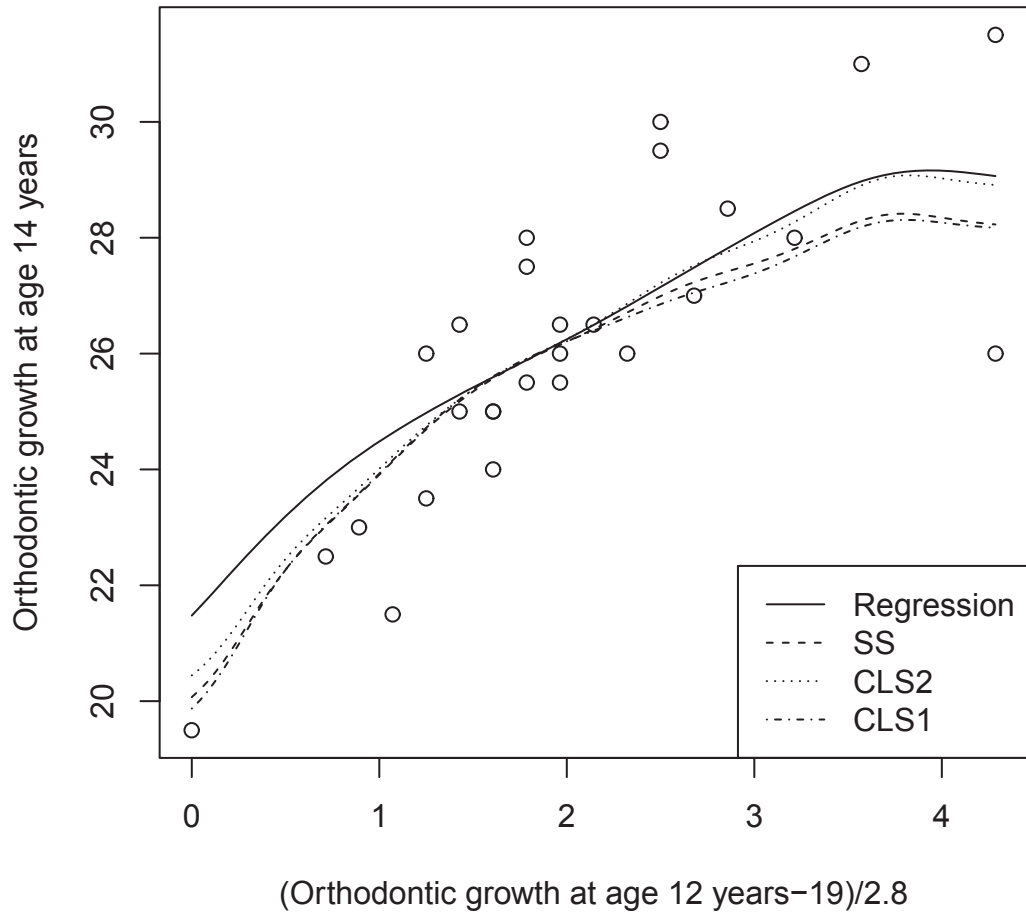


Figure 4.2: Regression of orthodontic growth data with DR method

## 4.3 Conclusions and future research

### 4.3.1 Conclusions

- The KR and DR methods show a better performance than HT method when both of the two Gamma kernels are applied into the nonparametric imputation. This is consistent with the results in the literature by [Ning and Cheng \(2012\)](#), in which HT method showed larger bias and MSE compared with both KR and DR methods.
- As the distribution of  $x$  is uniform, it is a bit more appropriate to choose  $\epsilon_n = 0$  when the function  $m(x) = 0$  at  $x = 0$ . When  $m(x) \neq 0$  at  $x = 0$ , there is no preference between two selections for  $\epsilon_n$  with KR and DR imputation methods.

### 4.3.2 Future research

Based on the limited conclusions of this study, some topics would be proposed for future research.

- More rigorous proof should be given to demonstrate the property on the application gamma kernel estimators into the nonparametric imputation methods. Also, more examples are needed to compare the performance of symmetric kernels with the asymmetric ones.
- The performance of both gamma kernels, including the two circumstances about the setting of the parameter  $\epsilon_n$  in CLS, should be compared within more nonparametric methods (e.g. nearest neighbor imputation) and different propensity score functions. Due to the small data size of the orthodontic grow data, other examples with larger data size are needed to observe the result.

# Bibliography

- Bagai, I. and Rao, B. L. S. Kernel type density estimates for positive valued random variables. *Sankhya*, 57(1):56–67, 1996.
- Brunner, E., Munzel, U., and Puri, M. L. Rank-score tests in factorial designs with repeated measures. *Journal of Multivariate Analysis*, 70:286–317, 1999.
- Chaubey, Y. P. and Sen, P. K. On smooth estimation of survival and density function. *Statistics and Decision*, 14:1–22, 1996.
- Chaubey, Y. P., Laïb, N., and Sen, A. Generalized kernel smoothing for non-negative stationary ergodic processes. *Journal of Nonparametric Statistics*, 50:973–997, 2010.
- Chaubey, Y. P., Li, J., Sen, A., and Sen, P. K. A new smooth density estimator for non-negative random variables. *Journal of the Indian Statistical Association*, 50:83–104, 2012.
- Chen, S. X. Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, 52(3):471–480, 2000.
- Cheng, P. E. Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, 89(425):81–86, 1994.
- Cheng, P. E. and Wei, L. J. Nonparametric inference under ignorable missing data process and treatment assignment. In *International Statistical Symposium*, volume 1, pages 97–112, Taipei, 1986.

- Dempster, A. P., Laird, N. M., and Rubin, D. B. Maximum likelihood estimation from incomplete data via em algorithm. *Journal of the Royal Statistical Society Series*, B39 (1):1–38, 1977.
- Feller, W. *An introduction to Probability Theory and its Applications*. John Wiley and Sons, New York, 1965.
- Hille, E. *Functional Analysis And Semigroups*. Am. Math. Colloq. Pub 31, New York, 1948.
- Horvitz, D. G. and Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952.
- Little, R. J. A. and Rubin, D. B. *Statistical Analysis with Missing Data*. Wiley, New York, 2 edition, 2002.
- Marron, J. S. and Ruppert, D. Transformations to reduce boundary bias in kernel density estimation. *Journal of Royal Statistical Society*, 56(4):653–671, 1994.
- Nadaraya, E. A. On estimating regression. *Theory of Probability and its Applications*, 9 (1):141–142, 1964.
- Neyman, J. contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, 33(201):101–116, 1938.
- Ning, J. and Cheng, P. E. A comparison study of nonparametric imputation methods. *Statistics and Computing*, 22(1):273–285, 2012.
- Potthoff, R. F. and Roy, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51:313–326, 1964.

- Robin, J. M. and Rotnitzky, J. M. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429): 122–129, 1995.
- Robin, J. M., M, R. J., and Zhao, L. P. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89 (427):846–866, 1994.
- Rosenbaum, P. R. and Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
- Rubin, D. B. *Multiple imputation for nonresponse in surveys*. Wiley, New York, 1987.
- Scaillet, D. O., Rotnitzky, A., and Robin, J. M. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448):1096–1120, 1999.
- Scaillet, O. Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics*, 40:217–226, 2004.
- Shi, J. H. and Song, W. X. Asymptotic results in gamma kernel regression. *Communications in Statistics Theory and Methods*, 2013.
- Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
- Titterton, D. M. and Mill, G. M. Kernel-based density estimates from incomplete data. *Journal of the Royal Statistical Society.*, 45(2):258–266, 1983.
- Wald, M. P. and Jones, M. C. *Kernel Smoothing*. Chapman and Hall, London, 1995.
- Wand, M. P., Marron, J. S., and Ruppert, D. Transformations in density estimation. *Journal of the American Statistical Association*, 86(414):343–353, 1991.

Watson, G. S. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics*, 26  
(4):359–372, 1964.

# Appendix A

## R Code for Simulation Study

```
S <- matrix(rep(0,12),nrow=3)
for (k in seq(200))
{
Smse=matrix(rep(0,12),nrow=3)
ri <- 1
for (n in c(100,150,200))
{
t <- runif(n,0,1)
delta <- seq(n)
for (i in 1:n)
{
if(t[i]<0.4*log(1+exp(2.5))-0.4*log(2))
delta[i] <- 1
else
delta[i] <- 0
}

ci <- 1
```

```

for(model in c(1,2,3,4))
{
x <- runif(n,0,1)
y1 <- 2*x+rnorm(n,0,0.5)
y2 <- 3-12*(x-0.5)^2+rnorm(n,0,0.5)
y3 <- x+2*exp(-x^2)+rnorm(n,0,0.5)
y4 <- sin(2*x)+2*exp(-x^2)+rnorm(n,0,0.5)
y=y1*(model==1)+y2*(model==2)+y3*(model==3)+y4*(model==4)

```

```

#KR imputation method with SS kernel

```

```

smse <- function(h)
{
xij <- matrix(kronecker(x,x/h,"^"), nrow=n)
Xiy <- xij%%diag(delta*y*exp(-x/h))
Xi <- xij%%diag(delta*exp(-x/h))
Asum <- apply(Xiy,1,sum);
Bsum <- apply(Xi,1,sum);
mean((y-Asum/Bsum)^2)
}

```

```

#HT imputation method with SS kernel

```

```

smse=function(h)
{
xij <- matrix(kronecker(x,x/h,"^"), nrow=n)
Xiy <- xij%%diag(delta*exp(-x/h))
Xi <- xij%%diag(exp(-x/h))
Asum <- apply(Xiy,1,sum)
Bsum <- apply(Xi,1,sum)

```



```

mean((y-delta*y/(Asum/Bsum))^2)
}

```

```

#DR imputation method with SS kernel
smse=function(h)
{
xij <- matrix(kronecker(x,x/h,"^"), nrow=n)
Xiy <- xij%*%diag(delta*y*exp(-x/h))
Xi <- xij%*%diag(delta*exp(-x/h))
Yi <- xij%*%diag(exp(-x/h))
Asum <- apply(Xiy,1,sum)
Bsum <- apply(Xi,1,sum)
Csum <- apply(Yi,1,sum)
Wi <- Bsum/Csum
MXi <- Asum/Bsum
mean((y-MXi-delta*(y-MXi)/Wi)^2)
}

```

```

#KR imputation method with CLS1 kernel, set en=0 for CLS2 kernel
smse <- function(aa)
{
vn <- aa;
en <- aa^2;
xij <- matrix(kronecker(x,vn^2*(x+en),"/"), nrow=n)
Xiy <- exp(-xij)%*%diag(delta*y*x^{1/vn^2-1})
Xi <- exp(-xij)%*%diag(delta*x^{1/vn^2-1})
Asum <- apply(Xiy,1,sum)
Bsum <- apply(Xi,1,sum)

```

```
mean((y-Asum/Bsum)^2)
}
```

```
#HT imputation method with CLS1 kernel, set en=0 for CLS2 kernel
smse=function(aa)
{
vn <- aa;
en <- aa^2;
xij <- matrix(kronecker(x,vn^2*(x+en),"/"), nrow=n)
Xiy <- exp(-xij)%*%diag(delta*x^{1/vn^2-1})
Xi <- exp(-xij)%*%diag(x^{1/vn^2-1})
Asum <- apply(Xiy,1,sum)
Bsum <- apply(Xi,1,sum)
mean((y-delta*y/(Asum/Bsum))^2)
}
```

```
#DR imputation method with CLS1 kernel, set en=0 for CLS2 kernel
smse=function(aa)
{
vn <- aa
en <- 0
xij <- matrix(kronecker(x,vn^2*(x+en),"/"), nrow=n)
Xiy <- exp(-xij)%*%diag(delta*y*x^{1/vn^2-1})
Xi <- exp(-xij)%*%diag(delta*x^{1/vn^2-1})
Yi <- exp(-xij)%*%diag(x^{1/vn^2-1})
Asum <- apply(Xiy,1,sum)
Bsum <- apply(Xi,1,sum)
Csum <- apply(Yi,1,sum)
```

```

Wi <- Bsum/Csum
MXi <- Asum/Bsum
mean((y-MXi-delta*(y-MXi)/Wi)^2)
}

hseq <- seq(0.001,1,length=200);
fsmse <- rep(0,length(hseq))
i <- 1;
for(h in hseq)
  {
    fsmse[i]=smse(h)
    i=i+1
  }

Smse[ri,ci]=min(fsmse[fsmse!="NaN"&fsmse!="Inf"])
ci=ci+1
}
ri=ri+1
}
S=S+Smse
}
S/200
dimnames(S)=list(c(100,150,200),c("model 1","model 2","model 3","model 4"))
S

```