On Minimum Distance Estimation for Binomial Regression Models

by

Boxiao Liu

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences

University of Alberta

Abstract

This thesis investigates efficient and robust estimators for binomial regression models. For this purpose, I have made use of two minimum distance estimation methods developed for discrete data, namely Minimum Hellinger Distance Estimation (MHDE) and Symmetric Chi-squared Distance Estimation (SCDE) methods. These methods generally known to produce efficient estimators when the chosen model is correct and, at the same time, are robust against model misspecification and outliers. Asymptotic properties and robustness features of the proposed estimators are discussed through theoretical demonstrations and simulations. Furthermore, the performance of estimators is compared with the traditional estimation approach of the maximum likelihood estimation.

Binomial regression models generally requires a specified "link function." In this thesis, cumulative distribution functions of the logistic and standard normal distributions are primarily used as the link functions. From theoretical results, it is concluded that the proposed MHDE is asymptotically equivalent to the maximum likelihood estimator when the model is correctly chosen. Some asymptotic properties of the proposed SCDE estimator is studied. Monte Carlo simulations are carried out compare the estimators for small to moderate sample sizes. It is observed that both MHDE and SCDE estimators show some robustness against model contamination, and the MHDE and the SCDE outperform the MLE in various conditions. Optimal conditions are discussed through extensive simulations under different scenarios.

Acknowledgements

I would like to take this opportunity to thank my supervisor, Dr. Rohana J. Karunamuni, for his dedication of supervising me for the past two years. Dr. Karunamuni has been the most influential professor I have had, and the most inspiring person in my university studies. His tremendous amount of dedication to academia and teaching would further motivate me to achieve greatness in all aspects of my personal life.

I also would like to thank Dr. Kong for his generous help and guidance over the years. His inspiring perspective and critical thinking on learning illuminates my paths in life.

I would also like to thank my examining committee members, Dr. Linglong Kong, Dr. Yau Shu Wang and Dr. Bei Jiang, for taking time away from their busy lives to serve in my committee.

I would like to thank all faculty members, staff and students here at the University of Alberta. Working among greatness has been truly amazing and shapes me into a better individual.

I remember the days of me as an undergraduate student who was eager to learn more and now to present my work after five years later. This could not happen without the unconditional support from my family. I would like to thank my grandmother and (late) grandfather for every effort they have made in my schooling to make today a reality.

Table of Contents

Abstract ii

Acknowledgements iii

List of Tables vi

List of Figures ix

Chapter 1. Introduction 1

1.1. Motivation and model assumptions 1

1.2. Maximum likelihood estimation 4

1.3. Motivation of new estimation methods 5

1.4. Organization of Chapters 6

Chapter 2. Minimum Hellinger Distance Estimation 8

2.1. Definition of Hellinger Distance 8

2.2. Properties of Hellinger Distance 8

2.3. MHDE of discrete distributions 10

2.4. Properties MHDE 13

2.5. Simulation Study 22

Chapter 3. Symmetric Chi-squared Distance Estimation 31

3.1. Introduction 31

3.2. Properties of SCDE 32

3.3. Simulation study 38

Chapter 4. Comparison between MLE, MHDE and SCDE 45

4.1. Logistic CDF 45

4.2. Normal CDF 49

Chapter 5. Summary and Future work 66

References 68

List of Tables

2.1 Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV 24

2.2 Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV 24

2.3 Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII 25

2.4 Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII 25

2.5 Biases and MSEs of MHDE & MLE under a Normal CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV 28

2.6 Biases and MSEs of MHDE & MLE under a Normal CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV 28

2.7 Biases and MSEs of MHDE & MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII 29

2.8 Biases and MSEs of MHDE & MLE under a Normal CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII 30

3.1 Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV 40

3.2 Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV 40

3.3 Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII 40

3.4 Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII 41

3.5 Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV 42

3.6 Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII 43

3.7 Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV 44

3.8 Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII 44

4.1 Baises and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV 46

4.2 Biases and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII 48

4.3 Biases and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV 50

4.4 Biases and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII 51

4.5 Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV 52

4.6 Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model V to VII 53

4.7 Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I to IV 55

4.8 Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V to VII 65

List of Figures

4.1 Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I 46

4.2 Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II 47

4.3 Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III 48

4.4 Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV 49

4.5 Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I 50

4.6 Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II 51

4.7 Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III 52

4.8 Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV 53

4.9 Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V 54

4.10 Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI 55

4.11 Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII 56

4.12 Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V 56

4.13 Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI 57

4.14 Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII 57

4.15 Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I 58

4.16 Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II 58

4.17 Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III 59

4.18 Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV 59

4.19 Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I 60

4.20 Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II 60

4.21 Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III 61

4.22 Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV 61

4.23 Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V 62

4.24 Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI 62

4.25 Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII 63

4.26 Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V 63

4.27 Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI 64

4.28 Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII 64

CHAPTER 1

Introduction

1.1. Motivation and model assumptions

Maximum likelihood method is by far the most popular estimation technique employed in statistics. Maximum likelihood based estimators have many nice properties, including the consistency and asymptotic efficiency under the model. Yet, they can be highly unstable if the model is not totally correct, and they are not *robust* if the data is slightly contaminated. (Here the word 'robust' refers to the ability of a procedure to retain its validity under a model misspecification and/or when outliers are present.) Further, their efficiency may be significantly reduced for model mispecification; i.e., if the chosen model is slightly different from the chosen model.

The need for robust procedures in statistical inference has been widely recognized now. This is motivated by the common belief that statistical models are just approximations to reality and that real data never come from the specified model exactly. Furthermore, a common problem in applied statistics is the presence of outliers in the data. Many different approaches have been proposed for finding robust statistics in the literature. These methods have had varying degrees of success in dealing with "bad" data but they all suffer from a loss of efficiency if the postulated/chosen model is the correct one. The literature on robust estimation is too extensive to make a complete listing here, an interested reader is referred to the monographs of Huber (1964) and Maronna et al. (2006) and the references therein for techniques and properties of well-known robust methods.

The purpose of this thesis is to study efficient robust estimators for a *binomial regression model*, as efficient robust procedures are vital for effective data analysis. Specifically, when the outcomes of a series of experiments are binary, the following model is used. Let $Y_j \sim$ Bernoulli (p_j) for $j = 1, ..., n$. Furthermore, let $\mathbf{z}_j = (1, z_{j,1}, ..., z_{j,p})^T$ be the values of control variables or clinical variables or other discriptive variables, known as covariates for the i^{th}-individual. Finally, let $g(.)$ be a monotonic and differentiable function, known as the link function, which maps the interval probabilities $(0, 1)$ to the real numbers. Examples of such tranformations are the logit $g(p) = \ln(p/(1 - p))$ and the probit $g(p) = \Phi^{-1}(p)$, the inverse of the standard distribution function. The regression parameters $\theta = (\theta_0, \theta_1, ..., \theta_p)^T$ enter into the model through $g(p_j) = \mathbf{z}_j^T \theta$.

An excellent introduction to data analysis procedures based on the above *binary regression model* is the monograph of McCullagh and Nelder (1983, Chapter 4). One's primary concern is the joint dependency between binary response and independent explanatory variables. Here we consider a binomial regression model. That is, we consider a situation where replication within a specific combination is readily accessible. We assume that K independent binomial random variables are available for observation. Thus, K groups with a total of N data points are collected, and there are n_j subjects within each group. So, $N = \sum_{j=1}^{K} n_j$. A binary response is coded 1 if subject shows a response to a given covariate $\mathbf{z}_j = (1, z_{j,1}, ..., z_{j,p})^T$, and 0 if it shows no response. Equivalently, we assume that there are K independent binomial random variables; i.e., $Y_j \sim$ Binomial (n_j, p_j) for $j = 1, ..., K$. We further assume that p_j is related to covariate $\mathbf{z}_j$ as $P(\text{showing a response} | \, \mathbf{z}_j) = p_j = F(\mathbf{z}_j^T \theta)$, where F is some known distribution function and $\theta = (\theta_0, \theta_1, ..., \theta_p)^T$ is the unknown regression parameter of interest. A natural estimator of p_j is the sample propotion of successes given by $\hat{p}_j = \frac{m_j}{n_j}$, where m_j is the number of responses out of n_j trials, $j = 1, ..., K$. For a *generalized linear model* set up with binomial regression, one generally assumes that $g(\mu_j) = \mathbf{z}_j^T\theta$, where $\mu_j = E(Y_j) = n_j p_j$, $j = 1, ...,$ K, and g is a link function (McCullagh and Nelder,1983, Chapter 4).Thus, our model is slightly different from the usual generalized linear model set up, but the methodologies used in this thesis can be easily modified to this model as well.

An example of the above set up is in "effective dose-level" estimation of dose-response studies experiments; see, e.g. Stather (1981), Li and Wiens (2011), Karunamuni et al. (2015), among others. Specifically, in pharmacology or toxicology studies, experimenters are often interested in estimating the effective dose level ED_p, where $0 < p < 1$. The ED_P is the dose at which $100p\%$ of the subjects show a response. Generally, K groups of test subjects characterized by different dose levels x_j ($j = 1, ..., K$) are collected, where each subject in the corresponding group is collected independently. The number of test subjects for groups would be n_j ($1 \leq j \leq K$) and the number of test subjects showing a response at the dose level x_j ($1 \leq j \leq K$) would be m_j. In the dose-response context, it is generally assumed that $x_1 < x_2 < .... < x_K$. That is, the outcome of interest is usually measured at several increasing dose levels. For every subject, a binary response is produced. If the subject shows a response, such subject is denoted as "1", and a no response is denoted as "0". So, the model reduces to $p_j = F(\mathbf{z}_j^T\theta)$ with $\theta = (\alpha, \beta)^T$ and $\mathbf{z}_j = (1, z_j)^T$ for some parameters α and $\beta > 0$.

1.2. Maximum likelihood estimation

Maximum likelihood estimation is by far the well-known and frequently used method of estimation for binary regression models. The likelihood function in the present set of binomial regression models takes the form

$$L(\theta, \mathbf{Z}) = \prod_{j=1}^{K} \binom{n_j}{m_j} p_j^{m_j} (1-p_j)^{n_j - m_j},$$

where $p_j = F(\mathbf{z}_j^T \theta)$ and $\mathbf{Z} = (\mathbf{z}_1^T, ..., \mathbf{z}_K^T)$. A typical method of finding a maximum likelihood estimator (MLE) solution is by equating the first order derivative of the objective function to zero and then solve for the unknown parameter. We assume that the parameter is compact and the true parameters away from boundary points. Alternatively, if a solution is not available in an explicit form, then iteration method can be used to obtain a numerical solution. Taking the logarithm of the above likelihood function we obtain

$$\begin{aligned} \log L(\theta, \mathbf{Z}) &= \sum_{j=1}^{K} \{\log \binom{n_j}{m_j} + \log(F_j^{m_j}(1-F_j)^{n_j - m_j})\} \\ &= \sum_{j=1}^{K} \{m_j \log F_j + (n_j - m_j) \log(1 - F_j) + \log \binom{n_j}{m_j}\}, \end{aligned}$$

where $F_j = F(\mathbf{z}_j^T \theta)$. Now differentiating $\log L(\theta, \mathbf{Z})$ with respect to θ and setting it equal to zero, we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\theta, \mathbf{Z}) = \sum_{j=1}^{K} z_j^T F_j^{(1)} \{\frac{m_j}{F_j} - \frac{n_j - m_j}{1 - F_j}\} = 0. \tag{1.1}$$

The above log-likelihood equation is nonlinear in $\boldsymbol{\theta}$ and thus cannot be solved analytically. An iterative approach such as Newton-Raphson is typically required to find the maximum likelihood estimator for θ; see, e.g., Li and Wiens (2011).

1.3. Motivation of new estimation methods

It is well known that the estimates of regression coefficients for logistic and probit models are not robust to outliers (Pregibon, 1982). The lack of robustness of the maximum likelihood estimator for θ has been extensively discussed in the statistical literature for binary regression models. The preceding models are in fact an example of generalized linear models (McCullagh and Nelder, 1983). Robust alternatives for ordinary and generalized linear models are treated in many papers; see, for instance, the work of Stefanski et al. (1986), Kunsch et al (1989), Morgenthaler (1992), Marazzi and Ruffieux (1996) and Cantoni and Ronchetti (2001). In particular, notable papers that discuss diagnostic methods, resistant techniques and redescending M-estimates in binary regression models include Pregibon (1981), Pregibon (1982), Copas (1988), Carroll and Pederson (1993), Bianco and Yohai (1996), Markatou et al. (1997) and Croux and Haesbroeck (2003), among others.

From (1.1) it is clear that the MLE is not robust and could be greatly affected by many types of errors, among others they include

1. errors in the measurement or recording of the z_j values,

2. errors in the p_j values caused, for example, by subjects showing a response (e.g., dying) that is the result of other causes,

3. errors caused by choosing the wrong link function.

The robust estimators proposed in the literature are based on modifications of the log-likelihood and are inspired by the work of Huber (1964), Hampel (1971, 1974) and others in robust linear regression. In this thesis, I proposed two estimators based on minimum distance estimation techniques, namely minimum Hellinger distance estimation (Beran, 1977) and symmetric chi-squared distance estimation (Lindsay, 1994) methods. They generally provide alternative estimators of θ that are efficient and more robust estimators than the MLE, at least for some of the above errors. Various other distances such as chi-squared distance, total variation distance, etc. have been used in the literature, see Lindsay (1994) for more discussions on these distances and their applications.

1.4. Organization of Chapters

The thesis is organized as follows. In Chapter 2, I will derive the proposed minimum Hellinger distance estimator (MHDE) of θ and its statistical properties. Some simulation examples are also given there. The mathematical conditions such as continuity and differentiality on the link function guarantees the existence of such estimators, and statistical properties such as efficiency, consistency and asymptotic normality offer us more insights into the behaviour of the proposed MHDE. In Chapter 3, mathematical and statistical properties of the proposed symmetric chi-squared distance estimator (SCDE) will be discussed, followed by some simulation examples. The proposed MHDE and SCDE will be compared with the corerrsponding MLE using simulations in both Chapters 2 and 3. In Chapter 4, the proposed MHDE and SCDE will be compared with the MLE in simulations. Optimal conditions under different scenarios will be selected based on the presence of contamination, link functions and the sample size n. In Chapter 5, I will summarize my findings of the two proposed estimators and some suggestions will be provided for further investigation of future work.

CHAPTER 2

Minimum Hellinger Distance Estimation

2.1. Definition of Hellinger Distance

Given two probably measures F and G with densities f and g, respectively, with respect to a dominating σ-finite measure μ, Hellinger Distance $H(F,G)$ between F and G is defined by

$$H(F,G) = h(f,g) = \left\{ \int [f^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x)]^2 d\mu(x) \right\}^{\frac{1}{2}}.$$

2.2. Properties of Hellinger Distance

1. h is a metric on the space of probability measures with $0 \leq h \leq \sqrt{2}$.

2. The value of $h(F,G)$ is independent of the choice of the dominating measure μ.

3. $h^2(f,g) \leq \int |f(x) - g(x)| d\mu(x) \leq 2h(f,g)$. So, the topology induced on the space of probability measures by the Hellinger metric is the same as that induced by the total variation metric, since $\int |f(x) - g(x)| d\mu(x) = 2s\text{up}|F(A) - G(A)|$, where the supremum is taken over all measurable sets A.

4. If $\{f_\theta : \theta \in \Theta\}$ is a parametric family of densities, then minimizing $h(f_\theta, g)$ is equivalent to maximizing $\int f_\theta^{\frac{1}{2}} g^{\frac{1}{2}} d\mu(x)$.

Definition 2.1: (Shorack 2000, p. 66). Suppose both the signed measure ϕ and the measure μ are σ-finite on a measure space $(\Omega,\ \mathcal{A},\ \mu)$. Then $\phi \ll \mu$ if and only if there exists uniquely a.e. μ a finite-valued $\mathcal{A}$-measurable function Z_0 on Ω for which

$$\phi(A) = \int_A Z_0 d\mu, \quad \forall A \in \mathcal{A} \tag{2.1}$$

The function Z_0 in above equation is often denoted by $\frac{d\phi}{d\mu}$, so one can re-write above equation to be

$$\phi(A) = \int_A \frac{d\phi}{d\mu} d\mu, \quad \forall A \in \mathcal{A} \tag{2.2}$$

Lemma 2.1: (Change of variable). Let $\mu \ll v$ with μ a σ finite measure on $(\Omega, \mathcal{A})$. If $\int X \, d\mu$ is well-defined, then

$$\int_A X d\mu = \int_A X \frac{d\mu}{dv} dv, \quad \forall A \in \mathcal{A} \tag{2.3}$$

The definition of $\phi \ll \mu$ means that

$$\forall A \in \mathcal{A}\,,\ \mu(A) = 0 \Rightarrow \phi(A) = 0.$$

Definition of Randon-Nikodym provides that for σ-finite measure μ dominating both F and G, i.e., $F \ll \mu, G \ll \mu$, we can write

$$F(A) = \int_A \frac{dF}{d\mu}\, d\mu; \quad G(A) = \int_A \frac{dG}{d\mu}\, d\mu.$$

Denote $f = \frac{dF}{d\mu}, g = \frac{dG}{d\mu}$ as Randon-Nikodym derivatives and substituting in (2.1), we obtain

$$H^2(F,G) = h^2(f,g) = \int [f^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x)]^2 d\mu(x)$$

$$= \int \frac{dF}{d\mu}\, d\mu + \int \frac{dG}{d\mu}\, d\mu - \int \sqrt{\frac{dF}{d\mu}\frac{dG}{d\mu}} d\mu$$

$$= \int f\, d\mu + \int g\, d\mu - 2\int \sqrt{fg} = 2 - 2\int \sqrt{fg}.$$

This shows that our choice of σ-finite measure μ has no quantitative effect on $h^2(f,g)$. To show the value of $h^2(f,g)$ is in between 0 and $\frac{1}{\sqrt{2}}$, take $f = 1$ and $g = 0$.

2.3. MHDE of discrete distributions

We consider two discrete probability distributions $P = \{f_j : j \in S\}$ and $Q = \{g_j : j \in S\}$, where S is a discrete set, $\sum f_j = \sum g_j = 1$. For all $j \in S$, $f_j > 0, g_j > 0$. Then the squared Hellinger Distance (HD) between two discrete distributions P and Q is defined by $H^2(P,Q) = \sum\limits_{j\in S}(\sqrt{f_j} - \sqrt{g_j})^2 = \sum\limits_{j\in S}(\frac{f_j - g_j}{\sqrt{f}_j + \sqrt{g}_j})^2$. The idea of MHDE is to minimize the Hellinger distance to obtain an estimator of the unknown parameter.

For our model, we compute $H^2(P,Q)$ with the following two discrete probability distributions:

$$P = (r_1\,\hat{p}_1, .., r_k\,\hat{p}_k, r_1(1-\hat{p}_1), .., r_k(1-\hat{p}_k))^T$$

$$Q = (r_1\,p_1, .., r_k\,p_k, r_1(1-p_1), .., r_k(1-p_k))^T$$

where $\hat{p}_j = \frac{m_j}{n_j}$, $p_j = F(\boldsymbol{z_j^T\theta})$, $r_{j,N} = \frac{n_j}{N}$, $N = \sum_{j=1}^k n_j$. (Note that $\hat{p}_j$ is an estimator of p_j, $1 \le j \le k$.) Using a simple algebraic expansion, we obtain the squared Hellinger Distance $H^2(P,Q)$ as

$$\sum_{j\in S}(\sqrt{P} - \sqrt{Q})^2 = r_{1,N}(\sqrt{\hat{p}}_1 - \sqrt{p}_1)^2 + r_{2,N}(\sqrt{\hat{p}}_2 - \sqrt{p}_2)^2 + ... + r_{k,N}(\sqrt{\hat{p}}_k - \sqrt{p}_k)^2 +$$

$$r_{1,N}(\sqrt{1-\hat{p}_1} - \sqrt{1-p_1})^2 + r_{2,N}(\sqrt{1-\hat{p}_2} - \sqrt{1-p_2})^2 + ... + r_{k,N}(\sqrt{1-\hat{p}_k} - \sqrt{1-p_k})^2$$

$$= \sum_{j\in S}\left\{r_{j,N}(\sqrt{\hat{p}}_j - \sqrt{p}_j)^2 + r_{j,N}(\sqrt{1-\hat{p}_j} - \sqrt{1-p_j})^2\right\}$$

$$= \sum_{j=1}^{k} r_{j,N}\left\{\left(\sqrt{\hat{p}_j} - \sqrt{F(\boldsymbol{z_j^T\theta})}\right)^2 + \left(\sqrt{1-\hat{p}_j} - \sqrt{1-F(\boldsymbol{z_j^T\theta})}\right)^2\right\}.$$

An MHDE is obtained by minimizing

$$\sum_{j=1}^{K} r_{j,N}\left\{\left(\sqrt{\hat{p}_j}-\sqrt{F(\boldsymbol{z_j}^T\boldsymbol{\theta})}\right)^2+\left(\sqrt{1-\hat{p}_j}-\sqrt{1-F(\boldsymbol{z_j}^T\boldsymbol{\theta})}\right)^2\right\} \tag{2.4}$$

w.r.t. $\boldsymbol{\theta}$ for a given link function F. Equivalently, an MHDE is equal to

$$\max_{\boldsymbol{\theta}\in\Theta^K}\sum_{j=1}^{K} r_{j,N}\left\{\sqrt{\hat{p}_jF(\boldsymbol{z_j}^T\boldsymbol{\theta})}+\sqrt{(1-\hat{p}_j)(1-F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}\right\}. \tag{2.5}$$

This can be shown from using simple algebra as follows:

$$\sum_{j=1}^{K} r_{j,N}\left\{(\sqrt{\hat{p}_j}-\sqrt{F(\boldsymbol{z_j}^T\boldsymbol{\theta})})^2+(\sqrt{1-\hat{p}_j}-\sqrt{1-F(\boldsymbol{z_j}^T\boldsymbol{\theta})})^2\right\}$$

$$=\sum_{j=1}^{K} r_{j,N}\left\{\left(\hat{p}_j+F(\boldsymbol{z_j}^T\boldsymbol{\theta})-2\sqrt{\hat{p}_jF(\boldsymbol{z_j}^T\boldsymbol{\theta})}\right)+\left(1-\hat{p}_j+1-F(\boldsymbol{z_j}^T\boldsymbol{\theta})-2\sqrt{(1-\hat{p}_j)(1-F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}\right)\right\}$$

$$=\sum_{j=1}^{K} r_{j,N}\left\{2-2\sqrt{\hat{p}_jF(\boldsymbol{z_j}^T\boldsymbol{\theta})}-2\sqrt{(1-\hat{p}_j)(1-F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}\right\}$$

$$=2-2\sum_{j=1}^{K} r_{j,N}\left\{\sqrt{\hat{p}_jF(\boldsymbol{z_j}^T\boldsymbol{\theta})}+\sqrt{(1-\hat{p}_j)(1-F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}\right\}.$$

Hence, to minimize the previous is to maximize the latter. Now, let us differentiate with respect to $\boldsymbol{\theta}$, in order to examine if a solution can be obtained by mathematical derivation:

$$\frac{\partial}{\partial\boldsymbol{\theta}}\sum_{j=1}^{K} r_{j,N}\left\{\sqrt{\hat{p}_jF(\boldsymbol{z_j}^T\boldsymbol{\theta})}+\sqrt{(1-\hat{p}_j)(1-F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}\right\}$$

$$=\sum_{j=1}^{K} r_{j,N}\frac{\partial}{\partial\boldsymbol{\theta}}\left\{\sqrt{\hat{p}_jF(\boldsymbol{z_j}^T\boldsymbol{\theta})}+\sqrt{(1-\hat{p}_j)(1-F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}\right\}$$

$$= \sum_{j=1}^{K} r_{j,N} \left\{ \frac{\boldsymbol{z}_j^T}{2} f(\boldsymbol{z_j}^T \boldsymbol{\theta}) \right\} \left\{ \sqrt{\hat{p}_j} F^{-\frac{1}{2}}(\boldsymbol{z_j}^T \boldsymbol{\theta}) - \sqrt{1-\hat{p}_j} [1 - F(\boldsymbol{z_j}^T \boldsymbol{\theta})]^{-\frac{1}{2}} \right\},$$

which is a non-linear function in $\boldsymbol{\theta}$. Thus, as in the maximum likelihood estimation in this context, an itaration method must be used to obtain an MHDE for $\boldsymbol{\theta}$.

2.3.1. Some notation

We now introduce some notation to state a few theorems on the proposed MHDE defined at (2.5). Let us denote $I^K = [0,1]^K$, as the product space, namely [0,1] with K copies. Let $\boldsymbol{\pi}_N$ and $\boldsymbol{r_N}$ denote vectors of length K with $\boldsymbol{\pi}_N = (\pi_{j,N})_{j=1}^K = (\frac{m_j}{n_j})_{j=1}^K$ and $\boldsymbol{r}_N = (\frac{n_j}{N})_{j=1}^K$. Define $E = \{\boldsymbol{r} : \boldsymbol{r} \in I^K, \sum_{j=1}^K r_j = 1, r_j > 0, 1 \leq j \leq K\}$ and $\mathcal{G}_K = I^K \times E$. Then, $\boldsymbol{\pi}_N \times \boldsymbol{r}_N \in I^K \times E$.

Let Θ be the parameter space for $\boldsymbol{\theta}$, and we assume that Θ is a compact subset of $\mathbb{R}^p$. A Hellinger distance functional for estimating true unknown parameter value $\boldsymbol{\theta}_0$ is a functional T: $\mathcal{G}_K \to \Theta$ such that $T(\boldsymbol{\pi}, \boldsymbol{r})$ is a value of $\boldsymbol{\theta}$ given by

$$\max_{\boldsymbol{\theta} \in \Theta} \sum_{j=1}^{K} \{r_j\} \left\{ \sqrt{\pi_j F(\boldsymbol{z_j}^T \boldsymbol{\theta})} + \sqrt{(1-\pi_j)(1 - F(\boldsymbol{z_j}^T \boldsymbol{\theta}))} \right\}, \tag{2.6}$$

where, $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_K)^T \in I^K$ and $\boldsymbol{r} = (r_1, r_2, ..., r_K)^T \in E$. Then the proposed MHDE defined by (2.5) of $\boldsymbol{\theta}$ is equal to $T(\boldsymbol{\pi}_N, \boldsymbol{r}_N)$.

2.4. Properties MHDE

2.4.1. Existence of MHDE

***Theorem* 2.1**: (i) If Θ is compact and F is continuous, then $T(\boldsymbol{\pi}, \boldsymbol{r})$ exists for all $(\boldsymbol{\pi}, \boldsymbol{r}) \in \mathcal{G}_K$; (ii) If F is continuous and strictly increasing on $\mathbb{R}$ and $\pi_j = F(\boldsymbol{z}_j^T \boldsymbol{\theta}_0)$, with not all $\boldsymbol{z}_j$'s equal, then $T(\boldsymbol{\pi}, \boldsymbol{r}) = \boldsymbol{\theta}_0$ uniquely, for every $\boldsymbol{r} \in E$.

Proof: Observe that

1. $F(\boldsymbol{z_j^T \theta})$ is a continuous function of $\boldsymbol{\theta}$; therefore $\sqrt{\pi_j F(\boldsymbol{z_j^T \theta})}$, $\sqrt{(1-\pi_j)(1-F(\boldsymbol{z_j^T \theta}))}$ are continuous.
2. $g(\boldsymbol{\theta}) = \sum_{j=1}^{K} \{r_j\}\{\sqrt{\pi_j F(\boldsymbol{z_j^T \theta})} + \sqrt{(1-\pi_j)(1-F(\boldsymbol{z_j^T \theta}))}\}$ is continuous of $\boldsymbol{\theta}$.
3. $g(\boldsymbol{\theta})$ is bounded (when K is fixed), because $|r_j| \leq 1, 0 \leq \pi_j \leq 1, 0 \leq F(\boldsymbol{z_j^T \theta}) \leq 1$.

Part (i) now follows from the above facts since the maximum of function

$$\sum_{j=1}^{K} \{r_j\}\{\sqrt{\pi_j F(\boldsymbol{z_j^T \theta})} + \sqrt{(1-\pi_j)(1-F(\boldsymbol{z_j^T \theta}))}\}$$

exists, that is $T(\boldsymbol{\pi}, \boldsymbol{r})$ exists.

For Part (ii), we use the Cauchy-Schwarz inequality as follows: for $\boldsymbol{\pi} = (\pi_1, .., \pi_j, .., \pi_N), \boldsymbol{F} = (F_1, .., F_j, .., F_N)$, it follows that maximum of $\boldsymbol{\pi} \cdot \boldsymbol{F}$ is obtained when $\boldsymbol{\pi} = k\boldsymbol{F}$. Assume $\pi_j = F(\boldsymbol{z_j^T \theta})$ can be obtained and $T(\boldsymbol{\pi}, \boldsymbol{r}) = \boldsymbol{\theta}_0$ is a solution of $\boldsymbol{\theta}$. Then with $\pi_j = F(\boldsymbol{z}_j^T \boldsymbol{\theta}_0)$ and F strictly increasing (one-to-one correspondence), we have for every j, $F(\boldsymbol{z_j^T \theta}) = F(\boldsymbol{z}_j^T \theta_0)$ implies $\boldsymbol{z}_j^T \boldsymbol{\theta} = \boldsymbol{z}_j^T \boldsymbol{\theta}_0$, which further implies $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, when $\boldsymbol{z}_j$ are not all equal.

Notice that if $\boldsymbol{z}_j$'s are identical copies for $j = 1, 2, ..., K$, then $T(\boldsymbol{\pi}, \boldsymbol{r})$ is not unique, as $\boldsymbol{\theta}$ is not a unique solution. Therefore, the condition that not all $\boldsymbol{z}_j$'s equal guarantees the parameterization is identifiable.

2.4.2. Continuity

***Theorem* 2.2**: Suppose Θ is compact, F is continuous and strictly increasing on $\mathbb{R}$, and $(\boldsymbol{\pi}, \boldsymbol{r})$ is such that $T(\boldsymbol{\pi}, \boldsymbol{r})$ is unique with $0 < \pi_j < 1$, $1 \leq j \leq K$. Then T is continuous at $(\boldsymbol{\pi}, \boldsymbol{r})$.

Proof: Let $\boldsymbol{\pi}_n = (\pi_{1,n}, \pi_{2,n}, ..., \pi_{K,n})^T$ and $\boldsymbol{r}_n = (r_{1,n}, r_{2,n}, ..., r_{K,n})^T$. Assume $\boldsymbol{\pi}_n \times \boldsymbol{r}_n \in \mathcal{G}_K$ and $(\boldsymbol{\pi}_n, \boldsymbol{r}_n) \to (\boldsymbol{\pi}, \boldsymbol{r})$ as $n \to \infty$. Denote $\boldsymbol{\theta} = T(\boldsymbol{\pi}, \boldsymbol{r})$ and $\boldsymbol{\theta}_n = T(\boldsymbol{\pi}_n, \boldsymbol{r}_n)$. We will show $T(\boldsymbol{\pi}_n, \boldsymbol{r}_n) \to T(\boldsymbol{\pi}, \boldsymbol{r})$ as $n \to \infty$. Then T is continuous at $(\boldsymbol{\pi}, \boldsymbol{r})$.

Define

$$g_{j,n}(\boldsymbol{\theta}) = log\{\sqrt{\pi_{j,n} F(\boldsymbol{z_j^T \theta})} + \sqrt{(1 - \pi_{j,n})(1 - F(\boldsymbol{z_j^T \theta}))}\}$$

$$g_n(\boldsymbol{\theta}) = \sum_{j=1}^{K} \{r_{j,n}\} \{g_{j,n}(\boldsymbol{\theta})\}$$

$$g_j(\boldsymbol{\theta}) = log\{\sqrt{\pi_j F(\boldsymbol{z_j^T \theta})} + \sqrt{(1 - \pi_j)(1 - F(\boldsymbol{z_j^T \theta}))}\}$$

$$g(\boldsymbol{\theta}) = \sum_{j=1}^{K} r_j g_j(\boldsymbol{\theta}).$$

We will prove that $g(\boldsymbol{\theta}_n) \longrightarrow g(\boldsymbol{\theta})$. Then from the facts that g is a continuous function of $\boldsymbol{\theta}$, uniqueness of $T(\boldsymbol{\pi}, \boldsymbol{r})$ and the comapctness of Θ, it follows that $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ as $n \to \infty$. By the triangular inequality,

$$|g(\boldsymbol{\theta}_n) - g(\boldsymbol{\theta})| \leq |g(\boldsymbol{\theta}_n) - g_n(\boldsymbol{\theta}_n)| + \sup_{\boldsymbol{\theta} \in \Theta} |g_n(\boldsymbol{\theta}_n) - g(\boldsymbol{\theta})|$$

Once $g(\boldsymbol{\theta}_n) \to g(\boldsymbol{\theta})$, $\max g(\boldsymbol{\theta}_n) \to \max g(\boldsymbol{\theta})$ because if a sequence is convergent, then all the subsequences are convergent as well, and Θ is compact so maximum exists. Then $\arg\max g(\boldsymbol{\theta}_n) \to \arg\max g(\boldsymbol{\theta})$.

It is our objective to show that $\sup_{\boldsymbol{\theta}\in\Theta} |g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})| \to 0$, Then the rest of the proof is as follows:

$$\sup_{\boldsymbol{\theta}\in\Theta} |g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})| \geq |\sup_{\boldsymbol{\theta}\in\Theta} g_n(\boldsymbol{\theta}) - \sup_{\boldsymbol{\theta}\in\Theta} g(\boldsymbol{\theta})| = |\max_{\boldsymbol{\theta}\in\Theta} g_n(\boldsymbol{\theta}) - \max_{\boldsymbol{\theta}\in\Theta} g(\boldsymbol{\theta})| \to 0$$

for every $\boldsymbol{\theta} \in \Theta$, and therefore $\max_{\boldsymbol{\theta}\in\Theta} g_n(\boldsymbol{\theta}) \xrightarrow{n\to\infty} \max_{\boldsymbol{\theta}\in\Theta} g(\boldsymbol{\theta})$ i.e., $g_n(\boldsymbol{\theta}_n) \to g(\boldsymbol{\theta})$. On the other hand, $\sup_{\boldsymbol{\theta}\in\Theta} |g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})| \to 0$ implies $|g_n(\boldsymbol{\theta}_n) - g(\boldsymbol{\theta})| \to 0$. This is clear since $|g_n(\boldsymbol{\theta}_n) - g(\boldsymbol{\theta}_n)| \leq \sup_{\boldsymbol{\theta}\in\Theta} |g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})|$.

We now show that

$$\sup_{\boldsymbol{\theta}\in\Theta} |g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})| \to 0. \tag{2.7}$$

We first show the pointwise convergence for every j, i.e., $g_{j,n}(\boldsymbol{\theta}) \xrightarrow{n\to\infty} g_j(\boldsymbol{\theta})$, then use the uniform continuity of $g_{j,n}$ for every j. Pointwise convergence and uniform continuity implies $|g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})| \to 0$, $\forall \boldsymbol{\theta} \in \Theta$. Since Θ is compact, it follows that $\sup_{\boldsymbol{\theta}\in\Theta} |g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})| \to 0$, i.e., g_n is uniformly convergent.

Using some algebraic manipulation, we have

$$|g_n(\boldsymbol{\theta}) - g(\boldsymbol{\theta})| = \left| \sum_{j=1}^{K} \{ r_{j,n} g_{j,n}(\boldsymbol{\theta}) - r_j g_j(\boldsymbol{\theta}) \} \right|$$

$$= \left| \sum_{j=1}^{K} \{ r_{j,n} g_{j,n}(\boldsymbol{\theta}) + r_{j,n} g_j(\boldsymbol{\theta}) - r_{j,n} g_j(\boldsymbol{\theta}) - r_j g_j(\boldsymbol{\theta}) \} \right|$$

$$= \left| \sum_{j=1}^{K} \{ r_{j,n} \left[g_{j,n}(\boldsymbol{\theta}) - g_j(\boldsymbol{\theta}) \right] + (r_{j,n} - r_j) g_j(\boldsymbol{\theta}) \} \right|$$

$$\leq \sum_{j=1}^{K} |r_{j,n}| |g_{j,n}(\boldsymbol{\theta}) - g_j(\boldsymbol{\theta})| + \sum_{j=1}^{K} |r_{j,n} - r_j| |g_j(\boldsymbol{\theta})|,$$

by applying the triangular inequality. Since $g_j(\boldsymbol{\theta})$ is bounded and $r_{j,n} \xrightarrow{n\to\infty} r_j$, the second term converges to zero. For the first term, first we replace $F(\boldsymbol{z_j}^T\boldsymbol{\theta})$ with F_j. Then,

$$g_{j,n}(\boldsymbol{\theta}) - g_j(\boldsymbol{\theta})$$

$$= log \left\{ \sqrt{\pi_{j,n} F_j} + \sqrt{(1 - \pi_{j,n})(1 - F_j)} \right\} - log \left\{ \sqrt{\pi_j F_j} + \sqrt{(1 - \pi_j)(1 - F_j)} \right\}$$

$$= log \left\{ \frac{\sqrt{\pi_{j,n} F_j} + \sqrt{(1 - \pi_{j,n})(1 - F_j)}}{\sqrt{\pi_j F_j} + \sqrt{(1 - \pi_j)(1 - F_j)}} \right\}$$

$$= log \left\{ \frac{\sqrt{\pi_{j,n} F_j} + \sqrt{(1 - \pi_{j,n})(1 - F_j)} - \sqrt{\pi_j F_j} - \sqrt{(1 - \pi_j)(1 - F_j)}}{\sqrt{\pi_j F_j} + \sqrt{(1 - \pi_j)(1 - F_j)}} + 1 \right\}.$$

It is then sufficient to show that

$$\left\{ \sqrt{\pi_{j,n} F_j} + \sqrt{(1 - \pi_{j,n})(1 - F_j)} \right\} - \left\{ \sqrt{\pi_j F_j} + \sqrt{(1 - \pi_j)(1 - F_j)} \right\} \to 0,$$

which is again, after simple algebraic rearrangement, equivalent to show that

$$\sqrt{F_j} \left(\sqrt{\pi_{j,n}} - \sqrt{\pi_j} \right) + \sqrt{1 - F_j} (\sqrt{1 - \pi_{j,n}} - \sqrt{1 - \pi_j}) \to 0.$$

Let $\delta_{j,n} = \sqrt{F_j}\left(\sqrt{\pi_{j,n}} - \sqrt{\pi_j}\right) + \sqrt{1-F_j}(\sqrt{1-\pi_{j,n}} - \sqrt{1-\pi_j})$, then using some algebraic manipulation, we have

$$|\delta_{j,n}| = (\pi_{j,n} - \pi_j)\left\{\frac{\sqrt{F_j}}{\sqrt{\pi_{j,n}} + \sqrt{\pi_j}} - \frac{\sqrt{1-F_j}}{\sqrt{1-\pi_{j,n}} - \sqrt{1-\pi_j}}\right\}$$

$$\leq |\pi_{j,n} - \pi_j|\left\{\left|\frac{\sqrt{F_j}}{\sqrt{\pi_{j,n}} + \sqrt{\pi_j}}\right| + \left|\frac{\sqrt{1-F_j}}{\sqrt{1-\pi_{j,n}} + \sqrt{1-\pi_j}}\right|\right\} = \epsilon_{j,n}.$$

Now, $\log\left\{\frac{|\delta_{j,n}|}{\sqrt{\pi_j F_j} + \sqrt{(1-\pi_j)(1-F_j)}} + 1\right\} \leq \log\left\{\frac{|\epsilon_{j,n}|}{\sqrt{\pi_j F_j} + \sqrt{(1-\pi_j)(1-F_j)}} + 1\right\}$, and the latter is bounded by

$$\max_{\boldsymbol{\theta}\in\Theta}\left\{\frac{\epsilon_{j,n}}{\sqrt{\pi_j F_j} + \sqrt{(1-\pi_j)(1-F_j)}} + 1, -log\left\{1 - \frac{\epsilon_{j,n}}{\sqrt{\pi_j F_j} + \sqrt{(1-\pi_j)(1-F_j)}}\right\}\right\}$$

Since $\pi_{j,n} \to \pi_j$, we have $|\pi_{j,n} - \pi_j\ | \xrightarrow{n\to\infty} 0$, which implies $\epsilon_{j,n} \to 0$. We know $g_{j,n}$ is uniformly continuous for all j, so $|g_{j,n}(\boldsymbol{\theta}) - g_j(\boldsymbol{\theta})\ | \to 0$ for every $\boldsymbol{\theta}$. Then from the compactness of Θ, we have $\sup_{\boldsymbol{\theta}\in\Theta}\ |g_{j,n}(\boldsymbol{\theta}) - g_j(\boldsymbol{\theta})\ | \to 0$. This completes the proof.

Theorem 2.3: Suppose $\boldsymbol{\Theta}$ is compact and F is strictly increasing and continuous on $\mathbb{R}$. Define length K vectors $\boldsymbol{r}_N = (r_{j,N})_{j=1}^K = (\frac{n_j}{N})_{j=1}^K$, $\boldsymbol{\pi}_N = (\pi_{j,N})_{j=1}^K = (\frac{m_j}{n_j})_{j=1}^K$, $\boldsymbol{r} = (r_j)_{j=1}^K$ and $\boldsymbol{\pi} = (\pi_j)_{j=1}^K = (F(\boldsymbol{z_j}^T\boldsymbol{\theta}))_{j=1}^K$, where $0 < r_j < 1$, $1 \leq j \leq K$. Assume that $r_{j,N} \to r_j > 0$ as $N \to \infty$, $1 \leq j \leq K$. Then MHDE is consistent, i.e., $T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) \xrightarrow{p} T(\boldsymbol{\pi}, \boldsymbol{r})$.

Proof : The proof follows from the continuity of $T(.,.)$ and $(\boldsymbol{\pi}_N, \boldsymbol{r}_N) \xrightarrow{p} (\boldsymbol{\pi}, \boldsymbol{r})$ as $N \to \infty$.

2.4.3. Asymptotics

***Theorem* 2.4**: Suppose Θ is compact and let $C = \{\boldsymbol{z}_j^T\boldsymbol{\theta} : \boldsymbol{\theta} \in \Theta, 1 \leq j \leq K\}$. Suppose F is thrice differentiable with derivatives $f, f^{(1)}, f^{(2)}$ bounded on C, F is strictly increasing and $F(C) \subseteq [\delta, 1-\delta]$ for some $\delta > 0$. Let $(\boldsymbol{\pi}, \boldsymbol{r}) \in \mathcal{G}_K$ be such that $T(\boldsymbol{\pi}, \boldsymbol{r})$ is unique, and $\{(\boldsymbol{\pi}_n, \boldsymbol{r}_n)\} = \{(\pi_{j,n})_{j=1}^K, (r_{j,n})_1^K\}$ be a sequence in $\mathcal{G}_K$ such that $(\boldsymbol{\pi}_n, \boldsymbol{r}_n) \to (\boldsymbol{\pi}, \boldsymbol{r})$. Let W_n be a $(p+1)\times(p+1)$ matrix whose components converge to zero as $n \to \infty$. Let $\Sigma = \sum_{j=1}^K r_j \boldsymbol{z}_j \boldsymbol{z}_j^T G_j^{(1)}(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta})$ and $\lambda(\boldsymbol{\pi}, \boldsymbol{r}, \boldsymbol{\theta}) = \sum_{j=1}^K r_j \boldsymbol{z}_j G_j(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta})$ with $G_j(y) = \frac{\partial}{\partial y}\log\Big\{\sqrt{\pi_j F(y)} + \sqrt{(1-\pi_j)(1-F(y))}\Big\}$. If Σ is non-singular, then

$$T(\boldsymbol{\pi}_n, \boldsymbol{r}_n) - T(\boldsymbol{\pi}, \boldsymbol{r}) = -\lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta})(\Sigma^{-1} + W_n) \tag{2.8}$$

as $n \to \infty$, where $\lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta})$ is obtained from $\lambda(\boldsymbol{\pi}, \boldsymbol{r}, \boldsymbol{\theta})$ by replacing $(\boldsymbol{\pi}, \boldsymbol{r})$ with $(\boldsymbol{\pi}_n, \boldsymbol{r}_n)$.

Proof : Let $T(\boldsymbol{\pi}_n, \boldsymbol{r}_n) = \boldsymbol{\theta}_n$ and $T(\boldsymbol{\pi}, \boldsymbol{r}) = \boldsymbol{\theta}$. Then $\boldsymbol{\theta}_n$ satisfies the following equation:

$$0 = \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{j=1}^K r_{j,n} \log\{\sqrt{\pi_{j,n} F(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta})} + \sqrt{(1-\pi_{j,n})(1-F(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta}))}\}$$

$$= \sum_{j=1}^K r_{j,n} \boldsymbol{z}_j \frac{1}{2} f_j \frac{\sqrt{\frac{\pi_{j,n}}{F_j}} - \sqrt{\frac{1-\pi_{j,n}}{1-F_j}}}{\sqrt{\pi_{j,n}F_j} + \sqrt{(1-\pi_{j,n})(1-F_j)}}.$$

Let $A_{j,n}(y) = \sqrt{\pi_{j,n}F(y)} + \sqrt{(1-\pi_{j,n})(1-F(y))}$, $A_{j,n}^{(1)}(y) = \frac{1}{2} f(y)\{\sqrt{\frac{\pi_{j,n}}{F(y)}} - \sqrt{\frac{1-\pi_{j,n}}{1-F(y)}}\}$

$$G_{j,n}(y) = \frac{\partial}{\partial y}\log\left\{\sqrt{\pi_{j,n}F(y)} + \sqrt{(1-\pi_{j,n})(1-F(y))}\right\} = \frac{A_{j,n}^{(1)}(y)}{A_{j,n}(y)}.$$

$G_j(y)$ is defined similarly as

$$G_j(y) = \frac{\partial}{\partial y}\log\left\{\sqrt{\pi_j F(y)} + \sqrt{(1-\pi_j)(1-F(y))}\right\} = \frac{A_j^{(1)}(y)}{A_j(y)}.$$

The first derivative of $G_{j,n}(y)$ is

$$G_{j,n}^{(1)}(y) = \frac{f^{(1)}(y)}{f(y)}\frac{A_{j,n}^{(1)}(y)}{A_{j,n}(y)} - \frac{f(y)^2}{4A_{j,n}(y)}\left(\sqrt{\frac{\pi_j}{F(y)^3}} + \sqrt{\frac{1-\pi_j}{(1-F(y))^3}}\right) - \left(\frac{A_{j,n}^{(1)}(y)}{A_{j,n}(y)}\right)^2.$$

and $\lambda(\boldsymbol{\theta}_n, \boldsymbol{\pi}_n, \boldsymbol{r}_n) = \sum_{j=1}^{K} r_{j,n}\boldsymbol{z}_j G_{j,n}(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta}_n)$.

Since F is thrice differentiable, $G_{j,n}$ and $G_{j,n}^{(1)}$ are continuous. Together with the boundedness condition in the theorem, $G_{j,n}$ and $G_{j,n}^{(1)}$ are continuous and bounded, $G_{j,n}^{(2)}$ is bounded provided in the condition that "... $F(C) \subseteq [\delta, 1-\delta]$ for some $\delta > 0$", F is thrice differentiable with f, $f^{(1)}$ and $f^{(2)}$ are bounded on C "...". Note that $G_{j,N}^{(2)}$ is not necessarily continuous; F has bounded derivatives; and F does not take values 0 and 1, for $\delta > 0$. Then, by a Taylor expansion,

$$G_{j,n}(\boldsymbol{z}_j^T\boldsymbol{\theta}_n) = G_{j,n}(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta}) + G_{j,n}^{(1)}(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta})\ \boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta}) + \frac{1}{2}G_{j,n}^{(2)}(\boldsymbol{z}_j\boldsymbol{\theta}_n^*)\ [\boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta})]^2, \tag{2.9}$$

where $\boldsymbol{\theta}_n^*$ is a value between $\boldsymbol{\theta}_n$ and $\boldsymbol{\theta}$.

If $a_n \to a$ and f is continuous on a compact set D, then $f(a_n) \xrightarrow{n\to\infty} f(a)$. Also $\pi_{j,n} \to \pi_j$ means $G_{j,n}^{(1)}(y) \to G_j^{(1)}(y)$ uniformly with respect to y. Now, since Theorem 5.2 implies $\boldsymbol{\theta}_n - \boldsymbol{\theta} \to 0$, plugging expression (2.9) into $0 = \sum_{j=1}^{K} r_{j,n}\boldsymbol{z}_j G_{j,n}(\boldsymbol{z}_j^T\boldsymbol{\theta}_n)$ gives

$$0 = \sum_{j=1}^{K} r_{j,n}\boldsymbol{z}_j[G_{j,n}(\boldsymbol{z}_{\boldsymbol{j}}^T\boldsymbol{\theta}) + \boldsymbol{z}_j^T G_{j,n}^{(1)}(\boldsymbol{z}_j^T\boldsymbol{\theta})\ (\boldsymbol{\theta}_n - \boldsymbol{\theta}) + \frac{1}{2}\boldsymbol{z}_j^T G_{j,n}^{(2)}(\boldsymbol{z}_j^T\boldsymbol{\theta}_n^*)\ (\boldsymbol{\theta}_n - \boldsymbol{\theta})^2]$$

$$= \sum_{j=1}^{K} r_{j,n} \boldsymbol{z}_j G_{j,n}(\boldsymbol{z}_{\boldsymbol{j}}^T \boldsymbol{\theta}) + \left\{ \sum_{j=1}^{K} r_{j,n} \boldsymbol{z}_j \boldsymbol{z}_j^T \left[G_{j,n}^{(1)}(\boldsymbol{z}_j^T \boldsymbol{\theta}) + \frac{1}{2} G_{j,n}^{(2)}(\boldsymbol{z}_j^T \boldsymbol{\theta}_n^*)\ \boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta}) \right] \right\} (\boldsymbol{\theta}_n - \boldsymbol{\theta})$$

Now, since $r_{j,n} \to r_j$, $\sum_{j=1}^K r_{j,n} \boldsymbol{z}_j \boldsymbol{z}_j^T G_j^{(1)}(\boldsymbol{z}_j^T \boldsymbol{\theta}) \to \sum_{j=1}^K r_j \boldsymbol{z}_j \boldsymbol{z}_j^T G_j^{(1)}(\boldsymbol{z}_j^T \boldsymbol{\theta}) = \Sigma$ and the elements of $\Sigma_1 = \sum_{j=1}^K r_j \boldsymbol{z}_j \boldsymbol{z}_j^T \frac{1}{2} G_{j,n}^{(2)}(\boldsymbol{z}_{\boldsymbol{j}}^T \boldsymbol{\theta}_n^*) \boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta})$ goes to zero.

Through algebraic manipulation, we then have $0 = \lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta}) + (\Sigma + \Sigma_1)(\boldsymbol{\theta}_n - \boldsymbol{\theta})$. When Σ_1 goes to zero as $n \to \infty$, we have $\boldsymbol{\theta}_n - \boldsymbol{\theta} = -\lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta})(\Sigma^{-1} + W_n)$, where elements of matrix W_n goes to zero, i.e., $W_n \propto \Sigma_1^{-1} \to 0$ as $n \to \infty$. This completes the proof of Theorem 2.4.

A special case that $G_j(y) = 0$ takes place when $\pi_j = F(y)$, and $G_j^{(1)}(y) = -\frac{1}{4} \frac{f^2(y)}{F(y)[1-F(y)]}$. Consequently, $\Sigma = \sum_{j=1}^K r_j \boldsymbol{z}_j \boldsymbol{z}_j^T G_j^{(1)}(\boldsymbol{z}_{\boldsymbol{j}}^T \boldsymbol{\theta}) = \sum_{j=1}^K r_j \boldsymbol{z}_j \boldsymbol{z}_j^T \left(-\frac{1}{4}\right) \frac{f^2(\boldsymbol{z}_{\boldsymbol{j}}^T \boldsymbol{\theta})}{F(\boldsymbol{z}_{\boldsymbol{j}}^T \boldsymbol{\theta})(1-F(\boldsymbol{z}_{\boldsymbol{j}}^T \boldsymbol{\theta}))}$. Let $\Sigma^* = \sum_{j=1}^K r_j \boldsymbol{z}_j \boldsymbol{z}_j^T \frac{f^2(\boldsymbol{z}_j^T \boldsymbol{\theta})}{F(\boldsymbol{z}_j^T \boldsymbol{\theta})[1-F(\boldsymbol{z}_j^T \boldsymbol{\theta})]}$. Now Σ^* has a less complicated form and can be readily expressed. Replacing Σ^{-1} by $(-4)\Sigma^{*^{-1}}$ gives

$$T(\boldsymbol{\pi}_n, \boldsymbol{r}_n) - T(\boldsymbol{\pi}, \boldsymbol{r}) = 4\lambda(\boldsymbol{\theta}_n, \boldsymbol{\pi}_n, \boldsymbol{r}_n)(\Sigma^{*^{-1}} + W_n).$$

A note here is that Σ^* is singular only when $F(\boldsymbol{z}_{\boldsymbol{j}}^T \boldsymbol{\theta}) = 0$ for some j values and $\boldsymbol{z}_j = \boldsymbol{z}$ for all other j values. This means $\boldsymbol{z}_j$ are not linearly independent from each other and hence rank less than the dimension. Thus Σ^* is non-singular except in trivial cases.

The next theorem gives the asymptotic distribution of the proposed MHDE. First we state a lemma.

***Lemma* 1** : If $\pi_{j,n} \to \pi_j$ as $n \to \infty$, then, for any value of y with $0 < F(y) < 1$,

$$G_{j,n}(y) - G_j(y) = \frac{1}{4} f(y)(\pi_{j,n} - \pi_j) \left[\pi_j(1 - \pi_j)F(y)(1 - F(y))\right]^{-\frac{1}{2}}$$

***Theorem* 2.5** : Suppose that the MHDE is consistent and the expansion (2.8) holds for $T(\boldsymbol{\pi}_N, \boldsymbol{r}_N)$. Suppose that the probability of a response at $\boldsymbol{z}_j$ is $\pi_j, 1 \leq j \leq K$, and that $T(\boldsymbol{\pi}, \boldsymbol{r})$ is uniquely defined. Let W_N be a $(p+1)\times(p+1)$ matrix whose components converge to zero in probability as $N \to \infty$. Let $\boldsymbol{\theta} = T(\boldsymbol{\pi}, \boldsymbol{r})$. Then, as $N \to \infty$, we have

$$\sqrt{N}\,[T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) - T(\boldsymbol{\pi}, \boldsymbol{r})] \xrightarrow{\mathcal{D}} N(0, \frac{1}{16}\Sigma^{-1}\Sigma^*\Sigma^{-1}) \tag{2.10}$$

where Σ^* is defined by $\sum_{j=1}^K r_j \boldsymbol{z}_j \boldsymbol{z}_j^T \frac{f^{(2)}(\boldsymbol{z_j^T\theta})}{F(\boldsymbol{z_j^T\theta})[1-F(\boldsymbol{z_j^T\theta})]}$ and Σ is as defined in Theorem 2.4.

Proof : From (2.8), we have

$$T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) - T(\boldsymbol{\pi}, \boldsymbol{r}) = -\lambda(\boldsymbol{\pi}_N, \boldsymbol{r}_N, \boldsymbol{\theta})(\Sigma^{-1} + W_N),$$

and so

$$T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) - T(\boldsymbol{\pi}, \boldsymbol{r}) = -\left\{\sum_{j=1}^K r_j \boldsymbol{z}_j \left[G_{j,N}(\boldsymbol{z_j^T\theta}) - G_j(\boldsymbol{z_j^T\theta})\right]\right\}\left\{\Sigma^{-1} + W_N\right\}.$$

Theorem 2.4 also have the condition that $\pi_{j,n}$ converges to π_j in probability. From the lemma above, we have $G_{j,N}(y) - G_j(y) = \frac{1}{4} f(y)(\boldsymbol{\pi}_{j,N} - \boldsymbol{\pi}_j)\left[\boldsymbol{\pi}_j(1-\boldsymbol{\pi}_j)F(y)(1-F(y))\right]^{-\frac{1}{2}}$. Using this expression, we obtain

$$T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) - T(\boldsymbol{\pi}, \boldsymbol{r}) = -\frac{1}{4}\left\{\Sigma^{-1} + W_N\right\}\sum_{j=1}^K r_j \frac{f(\boldsymbol{z_j^T\theta})(\pi_{j,N} - \pi_j)\boldsymbol{z}_j\{1 + o_P(1)\}}{\left[\pi_j(1-\pi_j)F(\boldsymbol{z_j^T\theta})\left(1 - F(\boldsymbol{z_j^T\theta})\right)\right]^{\frac{1}{2}}}$$

since $W_N = o_P(1)$ by assumptions. For $1 \leq j \leq K$, we have $\sqrt{n_j}(\pi_{j,N} - \pi_j) \xrightarrow{\mathbb{D}} N(0, \pi_j(1-\pi_j))$. Also $\sqrt{Nr_j} = \sqrt{n_j}$, and thus we complete the proof.

***Theorem* 2.6** : Suppose the consistency condition and the expansion (2.8) hold for $T(\boldsymbol{\pi}_N, \boldsymbol{r}_N)$. Let W_N be a $(p+1)\times(p+1)$ matrix whose components converge to zero in probability as $N \to \infty$. Then, as $N \to \infty$, we have

$$T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) - T(\boldsymbol{\pi}, \boldsymbol{r}) = \left\{\Sigma^{*^{-1}} + W_N\right\} \sum_{j=1}^{K} \left[\frac{r_j f(\boldsymbol{z_j}^T \boldsymbol{\theta}) \boldsymbol{z}_j (\pi_{j,N} - \pi_j)[1 + o_p(1)]}{F(\boldsymbol{z_j}^T \boldsymbol{\theta})[1 - F(\boldsymbol{z_j}^T \boldsymbol{\theta})]} \right]$$

and

$$\sqrt{N}\,[T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) - T(\boldsymbol{\pi}, \boldsymbol{r})] \xrightarrow{\mathcal{D}} N(0, \Sigma^{*^{-1}}).$$

Theorem 2.6 states that when the model is chosen correctly, MHDE and the MLE are asymptotically equivalent. Therefore, the MHDE is asymptotically efficient.

2.5. Simulation Study

2.5.1. Logistic CDF

An experiment testing 20 subjects at each of 10 levels was carried out. For each level, I generated 20 data points based on a binomial distribution with $n = 20$ and probability of success as $F(\boldsymbol{z_j}^T \boldsymbol{\theta})$, $j = 1, ..., 10$. Take $\boldsymbol{\theta} = (\theta_0, \theta_1)$ to be the *true* parameter vector and $\boldsymbol{z}_j = (1, x_j)^T$ where $x_j = j$, $\forall j = 1, ..., 10$. For Models I-IV studied below, the cumulative distribution function (CDF) of *Logistic* (1.2, 1.21) distribution is chosen as for the CDF F and the true parameter is taken as $\boldsymbol{\theta} = (\theta_0, \theta_1) = (-1,\ 0.4)$. Then I varied the true parameter vector $\boldsymbol{\theta}$ to be $(\theta_0 = -3, \theta_1 = 0.75)$ to see the effect of the proposed MHDE for the value of the true parameter. For model V, VI, VII studied below, I used the same CDF with $\boldsymbol{\theta} = (\theta_0 = -1, \theta_1 = 0.4)$, and then varied the true parameter vector $\boldsymbol{\theta}$ to be $(\theta_0, \theta_1) = (-3,\ 0.75)$ with the *Logistic*$(1.125, 2.27)$ distribution. *R*-function *optim*(.) (Nelder and Mead method) is used to search for numerical solutions of the MLE and MHDE defined by (1.1) and (2.5), respectively. Then the MLE and MHDE are obtained by averaging 1000 replications. To compare the performance between the MLE and the MHDE, I computed the bias and mean squared errors (MSEs) of the estimators based on 1000 replications as follows:

$$Bias(\hat{\theta}_{m,i}) = \frac{\sum_{i=1}^{1000}(\hat{\theta}_{m,i} - \theta_m)}{1000}, \quad MSE(\hat{\theta}_{m,i}) = \frac{\sum_{i=1}^{1000}(\hat{\theta}_{m,i} - \theta_m)^2}{1000}, \quad m = 0, 1.$$

The performance of MLE and MHDE was compared under each of the following four models:

Model I: $F(y) = \frac{1}{1+e^{-(y-\mu_y)/\sigma_y}} = L(y)$

Model II: $F(y) = 0.9\,L(y) + 0.1\,L(0.5\,y)$

Model III: $F(y) = 0.9\,L(y) + 0.1\,L(2\,y)$

Model IV: $F(y) = 0.9\,L(y) + 0.1$

Model I is the clean model (i.e., there is no contamination); Model II and Model III are classical Huber contamination models with 10% elongated tails and 10% shortened tails, respectively. In doing so, I flatten out or steepen up true distribution to examine the robustness of MHDE with respect to contamination of the clean model. Model IV represents the overall increase of response for 10% of the observation.

Simulation results are given Tables 2.1 and 2.2. The values in Tables 2.1 and 2.2 show that if we only consider the bias of these estimators for comparison, then the MLE has less bias under Models I and III. On the other hand, if we only consider the MSE of these estimators, then the MHDE has mean squared errors slightly greater or equals to the MLE, except in Model IV. Overall, if the true distribution has a longer tail, then the MLE loses its advantage (for example in the cases of Models II and IV), and we recommend the MHDE as our estimator of choice. An example of this could be that the actual

Table 2.1. Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	I	-0.1573	0.0261	0.2681	0.0071
	II	0.0301	-0.0063	0.2118	0.0058
	III	-0.2764	0.0510	0.3034	0.0086
	IV	0.4297	-0.0238	0.3793	0.0062
MLE	I	-0.080	0.0125	0.2118	0.0056
	II	0.0920	-0.0175	0.1946	0.0054
	III	-0.1961	0.0364	0.2327	0.0063
	IV	0.4789	-0.0347	0.4080	0.0062

Table 2.2. Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	I	-0.1580	0.0269	0.8361	0.0224
	II	0.0716	-0.0112	0.6922	0.0189
	III	-0.5146	0.0944	1.1528	0.0320
	IV	0.8268	-0.0476	1.4035	0.0225
MLE	I	-0.0260	0.0029	0.7008	0.0186
	II	0.1892	-0.0326	0.6477	0.0175
	III	-0.3645	0.0671	0.892	0.024
	IV	0.9248	-0.0689	1.5058	0.0224

distribution function F is logistic CDF, but the CDF of a normal distribution is used instead. On the other hand, if true distribution is with shortened tails, the MHDE does not offer significant protection against contamination. Furthermore, the MHDE offers protection against overall increase of response showing. An overall increase of response showing might occur, say, when increase of response is due to some other variables outside of this simulation study. In such cases, the MHDE has both smaller biases and smaller mean squared errors.

To further study the behaviour of proposed MHDE and to compare it with the MLE, I increased the contamination rate to 20% from 10%. That is, I considered following three models:

Model V: $F(y) = 0.8\, L(y) + 0.2\, L(0.5\, y)$

Model VI: $F(y) = 0.8\, L(y) + 0.2\, L(2\, y)$

Model VII: $F(y) = 0.8\, L(y) + 0.2$

Model V is with 20% contamination rate with elongated tails; Model VI is with 20% contamination rate with shortened tails; and Model VII is the strictly increasing response. Simulations were carried out based on these models and the results are presented in Tables 2.3 and 2.4.

Table 2.3. Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	V	0.1587	-0.0295	0.2210	0.0058
MHDE	VI	-0.4755	0.0851	0.4956	0.0146
	VII	0.9462	-0.0666	1.0760	0.0101
	V	0.2151	-0.0397	0.222	0.006
MLE	VI	-0.3777	0.0673	0.3634	0.0105
	VII	0.9843	-0.0773	1.1336	0.0109

Table 2.4. Biases and MSEs of MHDE & MLE under a Logistic CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	V	0.2405	-0.0444	0.7795	0.0221
MHDE	VI	-0.8827	0.1623	1.6992	0.0517
	VII	1.7458	-0.1155	3.6462	0.0322
	V	0.3506	-0.0644	0.7678	0.0220
MLE	VI	-0.7010	0.1294	1.2395	0.0372
	VII	1.8151	-0.1351	3.8444	0.0347

The comparison of estimators MLE and MHDE based on a larger contamination rate (20%) shows the following observations. If we are to compare based on the bias, then with true parameter vector ($\theta_0 = -1$, $\theta_1 = 0.4$), the MHDE has smaller biases than those of the MLE under Models V and VII, and the MLE has a smaller bias than that of the MHDE under Model VI. If we are to compare estimators based on the MSE only, then the MHDE has smaller MSEs under Models V and VII comapred to the MLE. This suggests that the MHDE offers consistent protection when the long-tailed contamination gets larger, as well as the strictly increasing contamination. Under the true parameter vector ($\theta_0 = -3$, $\theta_1 = 0.4$), we observed the same findings compared to the first four models (i.e., models I to IV) considered above. If we are to compare mean squared errors only, then under the true parameter vector ($\theta_0 = -1$, $\theta_1 = 0.4$), the MHDE has slightly smaller MSE than that of the MLE for Model VII. We reach the same conclusion when $\boldsymbol{\theta}$ is changed to ($\theta_0 = -3, \theta_1 = 0.75$). This suggests my findings obtained earlier that the MHDE outperforms the MLE in terms of MSE when contamination has strict increase in response, as demonstrated in Models IV and VII, and further it outperforms the MLE in terms of MSE when contamination gets larger ($\geq 20\%$). On the other hand, the MHDE does not perform well under short-tailed contamination, at least compared to the MLE. Note that the MHDE outperforms the MLE consistently in terms of bias when exposed under long-tailed or strictly increasing contaminations. As for the performance between two sets of true parameter vectors, it can be seen that the first set has a moderately smaller output in terms of the absolute value, compared to the second set. This suggests that the location and scale parameters of the logistic CDF may have an effect on the performance of both MHDE and MLE.

2.5.2. Normal CDF

In this section, the CDF of a normal distribution is used as the CDF for F. In particular, the CDFs of $N(1.2, 1.21)$ and $N(1.125, 2.27)$ are employed as the CDF for F. The idea is to build a better relationship between $F^{-1}(\pi_j)$ and $\boldsymbol{z_j^T\theta}$, so that the scattering is linearly scattered, rather than clustered together (creating boundary problems, for example). This can be solved by adjusting the location and scale of $\boldsymbol{z_j^T\theta}$. In a way, it is under the "true" distribution. If we use the CDF of $N(0,1)$ distibution for F, then it produces a less ideal fit and therefore not a *true* CDF.

Again, 20 subjects at each of 10 levels were simulated, by generating 20 data points within each level with binomial $(n = 20, F(\boldsymbol{z_j^T\theta}))$ distribution. Also, $\boldsymbol{\theta} = (\theta_0 = -1, \theta_1 = 0.4)$ is again used as the *true* parameter, and I continued to use $\boldsymbol{z}_j = (1,\ x_j)^T$, where $x_j = j,\ j = 1, ..., 10$.Then I varied true parameter to be $\boldsymbol{\theta} = (\theta_0, \theta_1) = (-3,\ 0.75)$ and $N(1.125, 2.27)$, as in the case of the logistic CDF. The performance of MHDE and MLE were compared for the following four models:

Model I: $F(y) = \Phi(y)$

Model II: $F(y) = 0.9\,\Phi(y) + 0.1\,\Phi(0.5\,y)$

Model III: $F(y) = 0.9\,\Phi(y) + 0.1\,\Phi(2\,y)$

Model IV: $F(y) = 0.9\,\Phi(y) + 0.1$

Model I is the clean model with normal CDF; Model II is a contaminated model with elongate tails; Model III is with shortened tails; Model IV represents the overall increase of response. The MLE and MHDE were again estimated by averaging 1000 replications. Biases and mean squared errors were obtained and the simulation analysis are given in Tables 2.5 and 2.6

Table 2.5. Biases and MSEs of MHDE & MLE under a Normal CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	I	-0.1650	0.0312	0.1697	0.0051
MHDE	II	0.0139	-0.0314	0.1221	0.0042
	III	-0.1941	0.0475	0.1683	0.0062
	IV	0.422	-0.036	0.2772	0.0043
	I	-0.0262	0.0062	0.1000	0.0029
MLE	II	0.1187	-0.0484	0.1046	0.0047
	III	-0.0497	0.0206	0.0951	0.0031
	IV	0.5025	-0.0535	0.3319	0.0051

Table 2.6. Biases and MSEs of MHDE & MLE under a Normal CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	I	-0.3367	0.0597	0.5997	0.0174
MHDE	II	0.1526	-0.0538	0.4173	0.0136
	III	-0.4657	0.0972	0.7002	0.0234
	IV	0.7696	-0.0618	0.9304	0.0145
	I	-0.0796	0.0132	0.3463	0.0097
MLE	II	0.3298	-0.0844	0.4079	0.0152
	III	-0.1980	0.0475	0.3771	0.0118
	IV	0.9239	-0.0953	1.1314	0.0174

Simulation results in Tables 2.5 and 2.6 based on the normal CDF show that if we are to compare the two estimators based on the bias only, then the MLE has the best performance under Models I and III. Otherwise, the MHDE has the best performance. If we are to compare the MLE and MHDE based on the mean squared error, then the MLE has a smaller MSE, except for Model IV, in which case the MHDE has a smaller MSE. This matches up with the obsevation on the bias above. This suggests that it might be best to use the MLE under clean model and for short tailed contaminations, whereas it might be best to use the MHDE when the true distribution has longer tails. If there is a strict-increase in response, then the MHDE dominates MLE in terms of both the bias and MSE.

To further study the behaviour of the MHDE, I examined three more models with contamination rate of 20%:

Model V: $F(y) = 0.8\,\Phi(y) + 0.2\,\Phi(0.5\,y)$

Model VI: $F(y) = 0.8\,\Phi(y) + 0.2\,\Phi(2\,y)$

Model VII: $F(y) = 0.8\,\Phi(y) + 0.2$

Model V is of long-tailed contamination; Model VI is of short-tailed contamination; and Model VII is of strict-increase contamination. Simulation results for these three models are presented in Tables 2.7 and 2.8. They are somewhat similar to the simulation results for Models I, II, III and IV with the normal CDF case.

Table 2.7. Biases and MSEs of MHDE & MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	V	0.0236	-0.0317	0.1179	0.0040
	VI	-0.1944	0.0623	0.169	0.008
	VII	0.8386	-0.0748	0.7831	0.0083
MLE	V	0.1253	-0.0484	0.1039	0.0046
	VI	-0.0474	0.0335	0.0924	0.0038
	VII	0.9033	-0.0916	0.8847	0.0105

Table 2.8. Biases and MSEs of MHDE & MLE under a Normal CDF with $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	V	0.1554	-0.0550	0.4147	0.0142
	VI	-0.7252	0.1558	1.0826	0.0408
	VII	1.5648	-0.1410	2.7482	0.0299
MLE	V	0.3308	-0.0850	0.4045	0.0156
	VI	-0.4043	0.0957	0.5555	0.0205
	VII	1.6884	-0.1727	3.1070	0.0376

CHAPTER 3

Symmetric Chi-squared Distance Estimation

3.1. Introduction

Let P and Q be two discrete distributions with equal length, that is $P = \{f_j : j \in S\}$ and $Q = \{g_j : j \in S\}$. The *squared Symmetric Chi-squared Distance* between P and Q is then defined as

$$S^2(P, Q) = \sum_{j \in S} \left(\frac{f_j - g_j}{f_j + g_j} \right)^2 . \tag{3.1}$$

We consider the following two distributions for P and Q :

$$P = (r_1 \hat{p}_1, r_2 \hat{p}_2, .., r_k \hat{p}_k, r_1(1 - \hat{p}_1), r_2(1 - \hat{p}_2), .., r_k(1 - \hat{p}_k))^T$$

$$Q = (r_1 p_1, r_2 p_2, .., r_k p_k, r_1(1 - p_1), r_2(1 - p_2), .., r_k(1 - p_k))^T.$$

Then $S^2(P, Q)$ reduces to

$$S^2(P, Q) = (\frac{\hat{p}_1 - p_1}{\hat{p}_1 + q_1})^2 + .. + (\frac{\hat{p}_k - p_k}{\hat{p}_k + p_k})^2 + (\frac{\hat{p}_1 - p_1}{2 - \hat{p}_1 - p_1})^2 + .. + (\frac{\hat{p}_k - p_k}{2 - \hat{p}_k - p_k})^2.$$

For the binomial regression model defined in Chapter 2, $S^2(P, Q)$ becomes

$$S^2(P, Q) = \sum_{j=1}^{K} \left\{ \left(\frac{\hat{p}_j - p_j}{\hat{p}_j + p_j} \right)^2 + \left(\frac{\hat{p}_j - p_j}{2 - \hat{p}_j - p_j} \right)^2 \right\}, \tag{3.2}$$

where $\hat{p}_j = \frac{m_j}{n_j}$, $p_j = F(\boldsymbol{z_j}^T \boldsymbol{\theta})$, $r_{j,N} = \frac{n_j}{N}$ and $N = \sum_{j=1}^{K} n_j$. Again, note that $\hat{p}_j$ is an estimator of p_j, $1 \leq j \leq K$. Then a symmetric chi-squared distance estimator (SCDE) of $\boldsymbol{\theta}$ obtained by minimizing $S^2(P, Q)$ w.r.t. $\boldsymbol{\theta}$.

To find a SCDE, we set the gradient of $S^2(P,Q)$ equal to zero to find a possible solution from a mathematical derivation. The derivarive of $S^2(P,Q)$ defined by (3.2) with respect to $\boldsymbol{\theta}$ gives

$$\begin{aligned}
&\frac{\partial}{\partial\theta}\sum_{j=1}^{K}\left\{\left(\frac{\hat{p}_j-p_j}{\hat{p}_j+p_j}\right)^2+\left(\frac{\hat{p}_j-p_j}{2-\hat{p}_j-p_j}\right)^2\right\}\\
&=\frac{\partial}{\partial\theta}\sum_{j=1}^{K}\left\{(\hat{p}_j-p_j)^2\ \left[(\hat{p}_j+p_j)^{-2}+(2-\hat{p}_j-p_j)^{-2}\right]\right\}.
\end{aligned} \tag{3.3}$$

If we let the bracket term {.} to zero, then we have

$$(\hat{p}_j-p_j)\left[(2-\hat{p}_j-p_j)^{-3}-(\hat{p}_j+p_j)^{-3}\right]=\frac{\hat{p}_j-p_j}{(2-\hat{p}_j-p_j)^3}-\frac{\hat{p}_j-p_j}{(\hat{p}_j+p_j)^3},$$

and this implies that

$$\frac{1}{(\hat{p}_j+p_j)^2}+\frac{1}{(2-\hat{p}_j-p_j)^2}=\frac{2(\hat{p}_j-1)}{(2-\hat{p}_j-p_j)^3}.$$

The preceding expression cannot be solved explicitely for $\boldsymbol{\theta}$. Thus, one needs to use a numerical metheod to find a solution for SCDE. We will attempt to minimize the objective function (3.2) numerically later in the simulation section.

3.2. Properties of SCDE

Some notation is needed to state a few theorems on the proposed SCDE. For $1\le j\le K$, let $\boldsymbol{\pi}_N=(\pi_{j,N})_{j=1}^K=\left(\frac{m_j}{n_j}\right)_{j=1}^K$ and $\boldsymbol{r}_N=\left(\frac{n_j}{N}\right)_{j=1}^K$. Define $E=\{\boldsymbol{r}:\boldsymbol{r}\in I^K,\ \sum_{j=1}^K r_j=1,\ r_j>0, 1\le j\le K\}$ and $\mathcal{G}_K=I^K\times E$. Denote $I^K=[0,1]^K$. Then $\boldsymbol{\pi}_N\times\boldsymbol{r}_N\in I^K\times E$. Let Θ be the parameter space for $\boldsymbol{\theta}$, and Θ is assumed to be a compact subset of $\mathbb{R}^p$. A symmetric chi-squared distance functional for estimating an unknown true parameter $\boldsymbol{\theta}_0$ is a functional $T:\ \mathcal{G}_K \to \Theta$ such that $T(\boldsymbol{\pi}, \boldsymbol{r})$ is a value of $\boldsymbol{\theta}$ given by

$$\min_{\boldsymbol{\theta}\in\Theta} \sum_{j=1}^{K} \left\{ \left(\frac{\pi_j - F(\boldsymbol{z_j^T\theta})}{\pi_j + F(\boldsymbol{z_j^T\theta})} \right)^2 + \left(\frac{\pi_j - F(\boldsymbol{z_j^T\theta})}{2 - \pi_j - F(\boldsymbol{z_j^T\theta})} \right)^2 \right\}, \tag{3.4}$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_K)^T \in I^K$ and $\boldsymbol{r} = (r_1, r_2, ..., r_K)^T \in E$. The proposed SCDE defined by (3.3) is then equal to $T(\boldsymbol{\pi}_N, \boldsymbol{r}_N)$.

3.2.1. Existence of SCDE

***Theorem* 3.1 :** (i) If Θ is compact and F is continuous, then if $\pi_j \neq 1$ for all j, $T(\boldsymbol{\pi}, \boldsymbol{r})$ exists for all $(\boldsymbol{\pi}, \boldsymbol{r}) \in \mathcal{G}_K$. (ii) If F is continuous and strictly increasing on $\mathbb{R}$ and $\pi_j = F(\boldsymbol{z_j^T\theta}_0), 1 \leq j \leq K$, with not all $\boldsymbol{z}_j$ equal, then $T(\boldsymbol{\pi}, \boldsymbol{r}) = \boldsymbol{\theta}_0$ uniquely.

Proof : Let $\Delta(\boldsymbol{\theta}) = \sum_{j=1}^{K} \Delta_j(\boldsymbol{\theta})$, where

$$\Delta_j(\boldsymbol{\theta}) = \left(\frac{\pi_j - F(\boldsymbol{z_j^T\theta})}{\pi_j + F(\boldsymbol{z_j^T\theta})} \right)^2 + \left(\frac{(1-\pi_j) - (1 - F(\boldsymbol{z_j^T\theta}))}{(1-\pi_j) + (1 - F(\boldsymbol{z_j^T\theta}))} \right)^2 .$$

Let $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$. Since $F(\boldsymbol{z_j^T\theta})$ is a continuous function with respect to $\boldsymbol{\theta}$, $\Delta(\boldsymbol{\theta}_n) \to \Delta(\boldsymbol{\theta})$ when $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$. Therefore, Δ is continuous w.r.t. $\boldsymbol{\theta}$ on a compact set, which means a minimum can be obtained. As for (ii), again we note the condition that not all $\boldsymbol{z}_j$'s are equal. Let us denote $\boldsymbol{\theta}_t$ as another solution different from $\boldsymbol{\theta}_0$. Since F is one-to-one and continuous, this implies $F(\boldsymbol{z}_j^T\boldsymbol{\theta}_t) \neq F(\boldsymbol{z}_j^T\boldsymbol{\theta}_0)$. Thus, $T(\boldsymbol{\pi}, \boldsymbol{r})$ is uniquely defined.

3.2.2. Continuity of SCDE

***Theorem* 3.2 :** Suppose Θ is compact, F is continuous and strictly increasing on $\mathbb{R}$, and $(\boldsymbol{\pi}, \boldsymbol{r})$ is such that $T(\boldsymbol{\pi}, \boldsymbol{r})$ is unique with $0 < \pi_j < 1, 1 \leq j \leq K$. Then T is continuous at $(\boldsymbol{\pi}, \boldsymbol{r})$.

Proof : Let $(\boldsymbol{\pi}_n, \boldsymbol{r}_n) \to (\boldsymbol{\pi}, \boldsymbol{r})$. Denote $\Delta(\boldsymbol{\theta}) = \sum_{j=1}^{K} \Delta_j(\boldsymbol{\theta})$ and $\Delta_n(\boldsymbol{\theta}) = \sum_{j=1}^{K} \Delta_{j,n}(\boldsymbol{\theta})$, where

$$\Delta_j(\boldsymbol{\theta}) = \left(\frac{\pi_j - F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}{\pi_j + F(\boldsymbol{z_j}^T\boldsymbol{\theta})}\right)^2 + \left(\frac{(1-\pi_j) - (1 - F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}{(1-\pi_j) + (1 - F(\boldsymbol{z_j}^T\boldsymbol{\theta}))}\right)^2,$$

$$\Delta_{j,n}(\boldsymbol{\theta}) = \left(\frac{\pi_{j,n} - F(\boldsymbol{z_j}^T\boldsymbol{\theta})}{\pi_{j,n} + F(\boldsymbol{z_j}^T\boldsymbol{\theta})}\right)^2 + \left(\frac{\pi_{j,n} - F(\boldsymbol{z_j}^T\boldsymbol{\theta})}{2 - \pi_{j,n} - F(\boldsymbol{z_j}^T\boldsymbol{\theta})}\right)^2.$$

It is sufficient to show, as $n \to \infty$,

$$\sup_{\boldsymbol{\theta} \in \Theta} |\Delta_n(\boldsymbol{\theta}) - \Delta(\boldsymbol{\theta})| \to 0 \tag{3.5}$$

The rest of the proof follows: from uniform continuity $\Delta_n \to \Delta$ for all $\boldsymbol{\theta}$ we get $|\Delta(\boldsymbol{\theta}_n) - \Delta_n(\boldsymbol{\theta}_n)| \to 0$, and it also implies that

$$\left|\max_{\boldsymbol{\theta} \in \Theta} \Delta_n(\boldsymbol{\theta}) - \max_{\boldsymbol{\theta} \in \Theta} \Delta(\boldsymbol{\theta})\right| = |\Delta_n(\boldsymbol{\theta}_n) - \Delta(\boldsymbol{\theta})| \to 0.$$

By squeeze theorem,

$$|\Delta(\boldsymbol{\theta}_n) - \Delta(\boldsymbol{\theta})| = |\Delta(\boldsymbol{\theta}_n) - \Delta_N(\boldsymbol{\theta}_n) + \Delta_n(\boldsymbol{\theta}_n) - \Delta(\boldsymbol{\theta})|$$

$$\leq |\Delta(\boldsymbol{\theta}_n) - \Delta_n(\boldsymbol{\theta}_n)| + \sup_{\boldsymbol{\theta} \in \Theta} |\Delta_n(\boldsymbol{\theta}_n) - \Delta(\boldsymbol{\theta})| \to 0,$$

and thus we conclude that $\Delta(\cdot)$ is continuous.

To show that

$$\sup_{\boldsymbol{\theta}\in\Theta} |\Delta_n(\boldsymbol{\theta}) - \Delta(\boldsymbol{\theta})| \to 0 \tag{3.6}$$

write

$$\Delta_n(\boldsymbol{\theta}) - \Delta(\boldsymbol{\theta}) =$$

$$\sum_{j=1}^{K}\left\{(\frac{\pi_{j,n}-F_j}{\pi_{j,n}+F_j})^2 + (\frac{\pi_{j,n}-F_j}{2-\pi_{j,n}-F_j})^2\right\} - \sum_{j=1}^{K}\left\{(\frac{\pi_j-F_j}{\pi_j+F_j})^2 + (\frac{\pi_j-F_j}{2-\pi_j-F_j})^2\right\}.$$

By triangular inequality,

$$|\Delta_n(\boldsymbol{\theta}) - \Delta(\boldsymbol{\theta})|$$

$$\leq \sum_{j=1}^{K}\left|(\frac{\pi_{j,n}-F_j}{\pi_{j,n}+F_j})^2 - (\frac{\pi_j-F_j}{\pi_j+F_j})^2\right| + \sum_{j=1}^{K}\left|(\frac{\pi_{j,n}-F_j}{2-\pi_{j,n}-F_j})^2 - (\frac{\pi_j-F_j}{2-\pi_j-F_j})^2\right|. \tag{3.7}$$

The RHS of (3.7) can be simplified by algebraic expansion as

$$= \left|\frac{4F_j(\pi_{j,n}\pi_j)}{(\pi_{j,n}+F_j)^2(\pi_j+F_j)^2}\right| |(\pi_{j,n}-\pi_j)|$$

$$+ \left|\frac{4(1-F_j)\left[(1-F_j)^2-(1-\boldsymbol{\pi}_j)(1-\boldsymbol{\pi}_{j,n})\right]}{(2-\boldsymbol{\pi}_{j,n}-F_j)^2(2-\boldsymbol{\pi}_j-F_j)^2}\right| |(\boldsymbol{\pi}_{j,n}-\boldsymbol{\pi}_j)|$$

$$= |A_n| \cdot |(\boldsymbol{\pi}_{j,n}-\boldsymbol{\pi}_j)|,$$

where $A_n = \left|\frac{4F_j(\boldsymbol{\pi}_{j,n}\boldsymbol{\pi}_j)}{(\boldsymbol{\pi}_{j,n}+F_j)^2(\boldsymbol{\pi}_j+F_j)^2}\right| + \left|\frac{4(1-F_j)\left[(1-F_j)^2-(1-\boldsymbol{\pi}_j)(1-\boldsymbol{\pi}_{j,n})\right]}{(2-\boldsymbol{\pi}_{j,n}-F_j)^2(2-\boldsymbol{\pi}_j-F_j)^2}\right|$. We note that A_n is bounded since $F_j, 0 < \boldsymbol{\pi}_j < 1, 0 < \boldsymbol{\pi}_{j,n} < 1$ are all probabilities. Then $|A_n| \cdot |(\boldsymbol{\pi}_{j,n} - \boldsymbol{\pi}_j)| \to 0$ since $\pi_{j,n} \to \pi_j \to 0$.This shows $\sup_{\boldsymbol{\theta}\in\Theta} |\Delta_n(\boldsymbol{\theta}) - \Delta(\boldsymbol{\theta})| \to 0$. By uniqueness of $T(\boldsymbol{\pi}, \boldsymbol{r})$ and the compactness of Θ we then obtain that $T(\boldsymbol{\pi}_n, \boldsymbol{r}_n) \to T(\boldsymbol{\pi}, \boldsymbol{r})$ as $n \to \infty$; i.e., $T(.)$ is continuous at $(\boldsymbol{\pi}, \boldsymbol{r})$. This completes the proof.

3.2.3. Consistency of SCDE

***Theorem* 3.3 :** Suppose Θ is compact and F is strictly increasing and continuous on $\mathbb{R}$. Define length K vectors $\boldsymbol{r}_N = (r_{j,N})_{j=1}^K = \left(\frac{n_j}{N}\right)_{j=1}^K$, $\boldsymbol{\pi}_N = (\pi_{j,N})_{j=1}^K = \left(\frac{m_j}{n_j}\right)_{j=1}^K$, $\boldsymbol{r} = (r_j)_{j=1}^K$ and $\boldsymbol{\pi} = (\pi_j)_{j=1}^K = \left(F(\boldsymbol{z_j^T\theta})\right)_{j=1}^K$, where $0 < r_j < 1, 1 \leq j \leq K$. Assume that $r_{j,N} \to r_j > 0$ as $N \to \infty$. Then SCDE is consistent, i.e., $T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) \xrightarrow{p} T(\boldsymbol{\pi}, \boldsymbol{r})$ as $N \to \infty$.

Proof : The proof follows from the continuity of $T(.,.)$ and $(\boldsymbol{\pi}_N, \boldsymbol{r}_N) \to^P (\boldsymbol{\pi}, \boldsymbol{r})$ as $N \to \infty$.

3.2.4. Asymptotic results

***Theorem* 3.4 :** Suppose Θ is compact and let $C = \left\{\boldsymbol{z_j^T\theta} : \boldsymbol{\theta} \in \Theta, 1 \leq j \leq K\right\}$. Suppose F is thrice differentiable with derivatives $f, f^{(1)}, f^{(2)}$ bounded on C, F is strictly increasing and $F(C) \subseteq [\delta, 1-\delta]$ for some $\delta > 0$. Let $(\boldsymbol{\pi}_n, \boldsymbol{r}_n) = \left\{(\pi_{j,n})_{j=1}^K, (r_{j,n})_{j=1}^K\right\}$ be a sequence in $\mathcal{G}_K$ such that $(\boldsymbol{\pi}_n, \boldsymbol{r}_n) \to (\boldsymbol{\pi}, \boldsymbol{r})$ as $n \to \infty$. Let W_n be a $(p+1) \times (p+1)$ matrix whose components converge to zero as $n \to \infty$. Let $\Sigma = \sum_{j=1}^K \boldsymbol{z}_j \boldsymbol{z}_j^T G_j^{(1)}(\boldsymbol{z_j^T\theta})$ and $\lambda(\boldsymbol{\pi}, \boldsymbol{r}, \boldsymbol{\theta}) = \sum_{j=1}^K \boldsymbol{z}_j G_j(\boldsymbol{z_j^T\theta})$ with $G_j(y) = \frac{\partial}{\partial y}\sum_{j=1}^K \{(\frac{\pi_j - F(y)}{\pi_j + F(y)})^2 + (\frac{\pi_j - F(y)}{2-\pi_j - F(y)})^2\}$. If Σ is non-singular, then

$$T(\boldsymbol{\pi}_n, \boldsymbol{r}_n) - T(\boldsymbol{\pi}, \boldsymbol{r}) = -\lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta})(\Sigma^{-1} + W_n) \tag{3.8}$$

as $n \to \infty$, where $\lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta})$ is obtained from $\lambda(\boldsymbol{\pi}, \boldsymbol{r}, \boldsymbol{\theta})$ by replacing $(\boldsymbol{\pi}, \boldsymbol{r})$ with $(\boldsymbol{\pi}_n, \boldsymbol{r}_n)$.

Proof : Let $\boldsymbol{\theta}_n = T(\boldsymbol{\pi}_n, \boldsymbol{r}_n)$, $\boldsymbol{\theta} = T(\boldsymbol{\pi}, \boldsymbol{r})$. Define

$$G_{j,n}(y) = \frac{\partial}{\partial y}\{(\frac{\pi_{j,n} - F(\boldsymbol{z_j^T\theta})}{\pi_{j,n} + F(\boldsymbol{z_j^T\theta})})^2 + (\frac{\pi_{j,n} - F(\boldsymbol{z_j^T\theta})}{2 - \pi_{j,n} - F(\boldsymbol{z_j^T\theta})})^2\}$$

similarly to $G_j(y)$ with π_j replaced by $\pi_{j,n}$. Note that $\boldsymbol{\theta}_n$ is a solution to the equation

$$0 = \frac{\partial}{\partial \boldsymbol{\theta}} \sum_{j=1}^{K} \left\{ \left(\frac{\pi_{j,n} - F(\boldsymbol{z_j}^T \boldsymbol{\theta})}{\pi_{j,n} + F(\boldsymbol{z_j}^T \boldsymbol{\theta})} \right)^2 + \left(\frac{\pi_{j,n} - F(\boldsymbol{z_j}^T \boldsymbol{\theta})}{2 - \pi_{j,n} - F(\boldsymbol{z_j}^T \boldsymbol{\theta})} \right)^2 \right\}$$

$$= 4 \left[F(\boldsymbol{z_j}^T \boldsymbol{\theta}) - \pi_{j,n} \right] f(\boldsymbol{z_j}^T \boldsymbol{\theta}) \boldsymbol{z}_j^T \left[\frac{\pi_{j,n}}{(F(\boldsymbol{z_j}^T \boldsymbol{\theta}) + \pi_{j,n})^3} - \frac{1 - \pi_{j,n}}{(F(\boldsymbol{z_j}^T \boldsymbol{\theta}) + \pi_{j,n} - 2)^3} \right].$$

Since differentiability implies continuity, $G_{j,n}, G_{j,n}^{(1)}$ are continuous and bounded, $G_{j,n}^{(2)}$ is bounded but not necessarily continuous, because of the condition that "...F is thrice differentiable and bounded on C...." Then by a Taylor expansion, we have

$$G_{j,n}(\boldsymbol{z_j}^T \boldsymbol{\theta}_n) = G_{j,n}(\boldsymbol{z_j}^T \boldsymbol{\theta}) + G_{j,n}^{(1)}(\boldsymbol{z_j}^T \boldsymbol{\theta})(\boldsymbol{\theta}_n - \boldsymbol{\theta}) + \frac{1}{2} G_{j,n}^{(2)}(\boldsymbol{z_j}^T \boldsymbol{\theta}^*)[\boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta})]^2,$$

where $\boldsymbol{z_j}^T \boldsymbol{\theta}^*$ lies between $\boldsymbol{z_j}^T \boldsymbol{\theta}_n$ and $\boldsymbol{z_j}^T \boldsymbol{\theta}$. Note that $G_{j,n}(.) \to G_j$ uniformly in y, since $\boldsymbol{\pi}_n \to \boldsymbol{\pi}$. We replace $G_{j,n}^{(1)}$ with $G_j^{(1)}$ and rewrite $G_{j,n}(\boldsymbol{z_j}^T \boldsymbol{\theta}_n)$ as

$$G_{j,n}(\boldsymbol{z_j}^T \boldsymbol{\theta}_n) = G_{j,n}(\boldsymbol{z_j}^T \boldsymbol{\theta}) + G_j^{(1)}(\boldsymbol{z_j}^T \boldsymbol{\theta})\, \boldsymbol{z}_j(\boldsymbol{\theta}_n - \boldsymbol{\theta}) + \frac{1}{2} G_{j,n}^{(2)}(\boldsymbol{z_j}^T \boldsymbol{\theta}^*)\, [\boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta})]^2. \tag{3.9}$$

Then substituting the rhs of (3.9) in the equation $0 = \sum_{j=1}^{K} \boldsymbol{z}_j G_{j,n}(\boldsymbol{z_j}^T \boldsymbol{\theta}_n)$ gives

$$0 = \sum_{j=1}^{K} \boldsymbol{z}_j G_j(\boldsymbol{z_j}^T \boldsymbol{\theta}) + \sum_{j=1}^{K} \boldsymbol{z}_j \boldsymbol{z}_j^T \left[G_j^{(1)}(\boldsymbol{z_j}^T \boldsymbol{\theta}) + \frac{1}{2} G_{j,n}^{(2)}(\boldsymbol{z_j}^T \boldsymbol{\theta}^*)\, \boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta}) \right] (\boldsymbol{\theta}_n - \boldsymbol{\theta})$$

$$= \lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta}) + \left\{ \sum_{j=1}^{K} \boldsymbol{z}_j \boldsymbol{z}_j^T \left[G_j^{(1)}(\boldsymbol{z_j}^T \boldsymbol{\theta}) + \frac{1}{2} G_{j,n}^{(2)}(\boldsymbol{z_j}^T \boldsymbol{\theta}^*)\, \boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta}) \right] \right\} (\boldsymbol{\theta}_n - \boldsymbol{\theta})$$

and

$$\Sigma_{1,n} = \sum_{j=1}^{K} \boldsymbol{z}_j \boldsymbol{z}_j^T \frac{1}{2} G_{j,n}^{(2)}(\boldsymbol{z_j}^T \boldsymbol{\theta}^*) \boldsymbol{z}_j^T(\boldsymbol{\theta}_n - \boldsymbol{\theta}),$$

a matrix whose elements go to zero as $n \to \infty$. Therefore $0 = \lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta}) + (\Sigma + \Sigma_{1,n})(\boldsymbol{\theta}_n - \boldsymbol{\theta})$. Then we obtain $(\boldsymbol{\theta}_n - \boldsymbol{\theta}) = T(\boldsymbol{\pi}_n, \boldsymbol{r}_n) - T(\boldsymbol{\pi}, \boldsymbol{r}) = -\lambda(\boldsymbol{\pi}_n, \boldsymbol{r}_n, \boldsymbol{\theta})(\Sigma^{-1} + W_n)$, where $W_n \propto \Sigma_{1,n} \to 0$. This completes the proof.

***Theorem* 3.5 :** Suppose SCDE is consistent and the expansion (3.8) holds for $T(\boldsymbol{\pi}_N, \boldsymbol{r}_N)$. Suppose the probability of a response at $\boldsymbol{z}_j$ is π_j, $1 \leq j \leq K$, and that $T(\boldsymbol{\pi}, \boldsymbol{r})$ is uniquely defined. Let W_N be a $(p+1) \times (p+1)$ matrix whose components converge to zero in probability as $N \to \infty$. Let $\boldsymbol{\theta} = T(\boldsymbol{\pi}, \boldsymbol{r})$. Then, as $N \to \infty$, we have

$$\sqrt{N}\,[T(\boldsymbol{\pi}_N, \boldsymbol{r}_N) - T(\boldsymbol{\pi}, \boldsymbol{r})] \xrightarrow{\mathcal{D}} N(0, \Sigma^{-1}\Sigma\Sigma^{-1}), \tag{3.10}$$

where $\Sigma = \frac{1}{4}\sum_{j=1}^{K} f(\boldsymbol{z_j^T\theta})(1 - F(\boldsymbol{z_j^T\theta}))\left(\frac{1}{(F(\boldsymbol{z_j^T\theta}))^2} + \frac{1}{(1-F(\boldsymbol{z_j^T\theta}))^2}\right)^2 \boldsymbol{z}_j\boldsymbol{z}_j^T$.

3.3. Simulation study

I have kept the true parameters and distribution functions to be the same throughout the simulations in Chapters 2 and 3, so a fair comparison can be carried out in Chapter 4. An experiment testing 30 subjects at 10 levels was carried out. For each level, 30 random binomial realizations were generated based on *Binomial* $(n = 30, p = F(z_j^T\boldsymbol{\theta}))$ distribtion $j = 1, .., 10$. Again $\boldsymbol{\theta} = (\theta_0, \theta_1)$ denotes the true parameter and $\boldsymbol{z}_j = (1, x_j)^T$. I used the true parameter as $(\theta_0 = -1,\ \theta_1 = 0.4)$ for all models, and then varied the true parameter to be $(\theta_0 = -3,\ \theta_1 = 0.75)$ for all models. In order to obtain the proposed SCDE given circa (3.2) numerically, R-function *optim*(.) (Nelder and Mead method) was used to find the minimum of the objective function (3.2). Estimates of SCDE and MLE were then obtained based averaging 1000 replications. To compare the performance between the SCDE and MLE estimators, I obtained biases and mean squared errors as measures of comparison defined as follows:

$$Bias(\hat{\theta}_{m,i}) = \frac{\sum_{i=1}^{1000}(\hat{\theta}_{m,i} - \theta_m)}{1000}, \quad MSE(\hat{\theta}_{m,i}) = \frac{\sum_{i=1}^{1000}(\hat{\theta}_{m,i} - \theta_m)^2}{1000}, \quad m = 0, 1,$$

where i represents the sequential order of replication and m the number the parameters. For example, $\hat{\theta}_{0,1}$ represents the value of parameter estimator θ_0 at its first replication.

3.3.1. Logistic CDF

In this section, the CDFs of logistic density functions *Logistic*(1.2, 1.21) and *Logistic*(1.12, 2.27) are used as the cumulative distribution function for F. Simulation results were compared for Huber's contamination models. In particular, the four models were selected to compare the biases and mean squared errors (MSEs) of SCDE and MLE:

Model I: $F(y) = \frac{1}{1+e^{-(y-\mu_y)/\sigma_y}} = L(y)$

Model II: $F(y) = 0.9\,L(y) + 0.1\,L(0.5\,y)$

Model III: $F(y) = 0.9\,L(y) + 0.1\,L(2\,y)$

Model IV: $F(y) = 0.9\,L(y) + 0.1$

Model I is the clean model (i.e., there is no contamination); Models II and III represent classical Huber contamination models with 10% elongated tails 10% shortened tails, respectively. Model IV represents a model with an overall increase of response for 10% of the observation. Simulated results of biases and MSEs of SCDE and MLE for true parameters ($\theta_0 = -1$, $\theta_1 = 0.4$) are presented in Tables 3.1 and 3.2.

The simulation results based on the *Logistic*(*1.2*, *1.21*) distribution with true parameters ($\theta_0 = -1$, $\theta_1 = 0.4$) demonstrate that if we only compare biases between the two estimators SCDE and MLE, then the SCDE has a smaller bias under Models II &

Table 3.1. Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
SCDE	I	-0.1002	0.0185	0.1629	0.0047
	II	0.0224	-0.0041	0.1443	0.0038
	III	-0.2558	0.0460	0.2260	0.0065
	IV	0.4555	-0.0266	0.3435	0.0048
MLE	I	-0.0152	0.0031	0.1284	0.0036
	II	0.1051	-0.0192	0.1297	0.0035
	III	-0.1666	0.0301	0.1643	0.0046
	IV	0.5147	-0.0414	0.3853	0.0052

Table 3.2. Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
SCDE	I	-0.2630	0.0444	0.6153	0.0168
	II	-0.0030	-0.0011	0.5333	0.0141
	III	-0.4575	0.0844	0.8014	0.0235
	IV	0.8252	-0.0436	1.1069	0.0153
MLE	I	-0.0808	0.0122	0.4585	0.0122
	II	0.1666	-0.0319	0.4520	0.0123
	III	-0.2861	0.0533	0.5613	0.0161
	IV	0.9458	-0.0737	1.2688	0.0165

Table 3.3. Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
SCDE	V	0.1397	-0.0250	0.1641	0.0045
	VI	-0.4457	0.0819	0.3595	0.0114
	VII	0.9249	-0.0593	0.9766	0.0075
MLE	V	0.2239	-0.0402	0.1732	0.0049
	VI	-0.3613	0.0656	0.2683	0.0082
	VII	0.9841	-0.0768	1.0753	0.0091

IV, otherwise the MLE has a smaller bias under the clean model and for Model III. On the other hand, if we only compare mean squared errors of the two estimators, then the

Table 3.4. Biases and MSEs of SCDE & MLE under the Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	V	0.2379	-0.0443	0.5478	0.0158
SCDE	VI	-0.8304	0.1516	1.3085	0.0392
	VII	1.6973	-0.1073	3.2907	0.0257
	V	0.3891	-0.0725	0.5577	0.0165
MLE	VI	-0.6472	0.1190	0.8887	0.0268
	VII	1.8118	-0.1415	3.6497	0.0313

SCDE has a smaller MSE than the MLE under the model IV only. The MLE has a smaller MSE than the SCDE for Models I, II & III. This may suggest that the SCDE has better performance than the MLE with respect to both bias and MSE under strict-increase contaminations; and the SCDE is less biased under 10% elongated contamination models, but it has a slightly higher mean squared error value. This could due to, for example, the bias-variance tradeoff.

I further studied the behaviour of SCDE under greater contamination levels. In particular, three more models were selected under a 20% contamination rate:

Model V: $F(y) = 0.8\,L(y) + 0.2\,L(0.5\,y)$

Model VI: $F(y) = 0.8\,L(y) + 0.2\,L(2\,y)$

Model VII: $F(y) = 0.8\,L(y) + 0.2$

Tables 3.3 and 3.4 present the simulation results for biases and MSEs of SCDE and MLE under the Models V, VI and VII. By comparing bises of the SCDE and MLE, we see that the SCDE has a smaller bias than the MLE under Models V & VII. By comparing the two estimators based on the MSE, we note that the MLE has a smaller MSE than the SCDE, except for the Model V &VII. From this simulation, one can thus conclude that if bias is the primary criteria used for comparison, then the SCDE is better than the

MLE when a long-tailed contamination is present; and the SCDE does not offer protection when a short-tailed contamination is present. In the case of ($\theta_0 = -3, \theta_1 = 0.75$), the mean and variance of $\boldsymbol{z}_j^T \boldsymbol{\theta}$ is different from that of ($\theta_0 = -1, \theta_1 = 0.4$). This might be the reason that the deviation when ($\theta_0 = -3, \theta_1 = 0.75$) is used is generally bigger than when ($\theta_0 = -1, \theta_1 = 0.4$).

3.3.2. Normal CDF

In this section, the CDF of the normal density $N(1.2,\ 1.21)$ is used as F. The following four models were considered first:

Model I: $F(y) = \Phi(y)$

Model II: $F(y) = 0.9\,\Phi(y) + 0.1\,\Phi(0.5\,y)$

Model III: $F(y) = 0.9\,\Phi(y) + 0.1\,\Phi(2\,y)$

Model IV: $F(y) = 0.9\,\Phi(y) + 0.1$

Table 3.5. Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	I	-0.0347	0.0055	0.0709	0.0018
SCDE	II	0.0881	-0.0436	0.0684	0.0035
	III	-0.0461	0.0175	0.0675	0.0021
	IV	0.4868	-0.0462	0.2957	0.0038
	I	-0.0207	0.0028	0.0670	0.0018
MLE	II	0.1399	-0.0511	0.0714	0.0040
	III	-0.0135	0.0150	0.0608	0.0020
	IV	0.4986	-0.0528	0.3031	0.0043

The simulation results for biases and MSEs of SCDE and MLE are given in Tables 3.5 and 3.6. We observe that if we are to compare biases only, then the SCDE has a smaller bias for Models II & IV. If we use the mean squared error for comparison, then we can see that the SCDE has smaller MSE values for Models II & IV as well. It is expected that the MLE has both a smaller bias and smaller MSE when the model is free from contamination. Notice that the difference between SCDE and MLE is not too big, and this might be due to normal CDF has a narrower distribution. Under a small contamination rate, the SCDE performs better than the MLE when contamination has long tails, as well as in the strict-increase case.

In order to see the behavior of SCDE under increased contamination, the contamination rate increased to 20% and then compared the performance between the SCDE and the MLE again for the following three models:

Model V: $F(y) = 0.8\,\Phi(y) + 0.2\,\Phi(0.5\,y)$

Model VI: $F(y) = 0.8\,\Phi(y) + 0.2\,\Phi(2\,y)$

Model VII: $F(y) = 0.8\,\Phi(y) + 0.2$

Table 3.6. Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
SCDE	V	0.0706	-0.0419	0.0718	0.0034
	VI	-0.0708	0.0298	0.0774	0.0028
	VII	0.8759	-0.0812	0.8102	0.0080
MLE	V	0.1228	-0.0491	0.0756	0.0039
	VI	-0.0520	0.0341	0.0674	0.0031
	VII	0.8907	-0.0897	0.8369	0.0094

Under 20% contamination, the SCDE has less bias for Models V & VII (i.e., same as the previous case), and the MLE has less bias in the Model VI again. Based on the simulation of this chapter, the SCDE should be preferred over the MLE when the potential contamination has longer tails, or strictly increasing. It is also worth to note that, the bias and MSE of SCDE for parameter θ_1 is smaller than those of the MLE.

Simulation results given in Tables 3.7 and 3.8 based on ($\theta_0 = -3, \theta_1 = 0.75$) reveal that the SCDE is superior in every category compared to the MLE. This may suggest that the SCDE is robust against misspecification of CDF when data $\boldsymbol{z}_j$ has a wider scattering (i.e., larger sample variance). Such cases may happen when the values of $\boldsymbol{z}_j^T\boldsymbol{\theta}$ are clustered around boundary probabilities that are close to 0 or 1. We will further investigate the behaviour of SCDE in comparison with the MLE and MHDE in the next chapter.

Table 3.7. Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	I	-0.0351	0.0078	0.2113	0.0057
SCDE	II	0.2576	-0.0763	0.2889	0.0115
	III	-0.2155	0.0465	0.2715	0.0078
	IV	0.8818	-0.0800	0.9793	0.0120
	I	-0.0202	0.0039	0.2201	0.0061
MLE	II	0.3420	-0.0899	0.3216	0.0135
	III	-0.2281	0.0536	0.2802	0.0095
	IV	0.9119	-0.0950	1.0168	0.0143

Table 3.8. Biases and MSEs of SCDE & MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	V	0.2607	-0.0745	0.3042	0.0117
SCDE	VI	-0.3244	0.0728	0.3349	0.0113
	VII	1.6522	-0.1553	2.9065	0.0293
	V	0.3562	-0.0901	0.3297	0.0135
MLE	VI	-0.3466	0.0853	0.3614	0.0144
	VII	1.6895	-0.1724	3.0230	0.0345

CHAPTER 4

Comparison between MLE, MHDE and SCDE

In this chapter, I compare the performance of the proposed estimators, MHDE and SCDE, with the traditional estimator MLE under the same conditions. Again, the behaviors of MHDE and SCDE are studied under some contamination models. Further, K (= 10) groups of Bernoulli random variables are generated. Within each group, $n_j(= 30)$ Bernoulli random variables are independently and identically distributed with corresponding probability of success p_j equal to $F(\boldsymbol{z_j}^T\boldsymbol{\theta})$, $j = 1, ..., 10$, where $F(.)$ is a CDF.

4.1. Logistic CDF

In this section, the CDF of *Logistic*(1.2, 1.21) distribution is used as the CDF for F to represent the true distribution function to study the estimators. A parameter vectors $(\theta_0 = -1, \theta_1 = 0.4)$ and $(\theta_0 = -3, \theta_1 = 0.75)$ are used as the "*true*″ parameter vector. Again, the same four models used in Chapters 2 & 3 are studied in this chapter, namely the *clean*, *10% short-tailed*, *10% long-tailed*, *10% strict-increase* models. Biases and mean squared errors of the three estimators are presented in Table 4.1 under these four different models.

From Table 4.1, we can observe the following. If we use bias as the measure of comparison, then the SCDE has a smaller bias in absolute value than the MLE and MHDE for Models II and IV. On the other hand, if we use the mean squared error as the measure of comparison, then the SCDE again has a slightly smaller MSE than the MLE and MHDE as well for IV. So, this suggests that the SCDE is an overall better estimator if

Table 4.1. Baises and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	I	-0.0692	0.0116	0.1451	0.0039
	II	0.0543	-0.0106	0.1293	0.0036
	III	-0.2076	0.0371	0.1908	0.0053
	IV	0.4682	-0.0330	0.3487	0.0047
SCDE	I	-0.1215	0.0210	0.1779	0.0049
	II	0.0017	-0.0012	0.1476	0.0041
	III	-0.2548	0.0457	0.2346	0.0066
	IV	0.4354	-0.0236	0.3299	0.0047
MLE	I	-0.0288	0.0042	0.1333	0.0036
	II	0.0927	-0.0175	0.1282	0.0036
	III	-0.1622	0.0288	0.1631	0.0045
	IV	0.4988	-0.0397	0.3706	0.0049

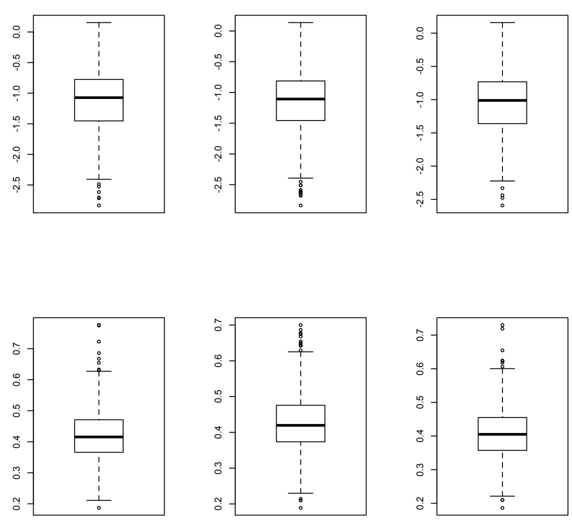

Figure 4.1. Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I

the model has longer tails; otherwise both MHDE and SCDE are subpar compared to the MLE when short-tailed models are present. Boxplots based on different contamination

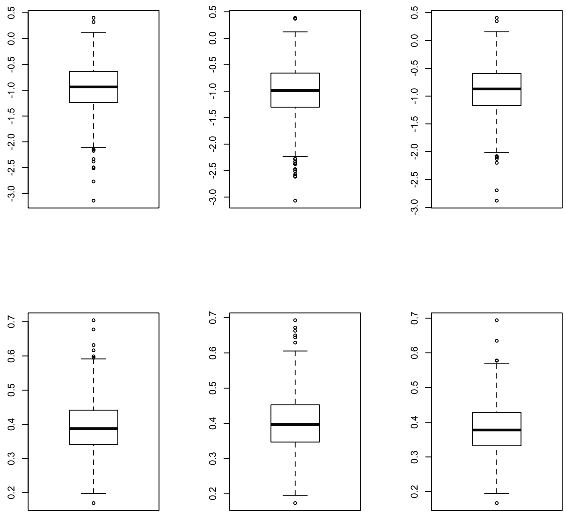

Figure 4.2. Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II

models suggests a systematic trends of estimators θ_0 and θ_1. It appears that θ_0 tend to be under-estimated, and θ_1 tend to be over-estimated for the clean and short-tailed models, whereas the othe way around for the long and strict-increase models. The skewness of histograms suggests the systematic deviation. To further study the behaviour of the estimators, three more models with a higher contamination rate of 20% as described in Chapter 1 are again used. Table 4.2 present the corresponding simulation results for biases and MSEs for the SCDE, MHDE and MLE.

Under the higher contamination rate, simulation results show that if we wish to compare biases only, then the SCDE has a clear advantage over both MHDE and MLE for Models V and VII. On the other hand, if we compare based on the mean squared error, then the MHDE and SCDE has smaller MSEs over MLE. The MLE outperforms both the SCDE and MHDE for the Model VI. In summary, both estimators MHDE and SCDE

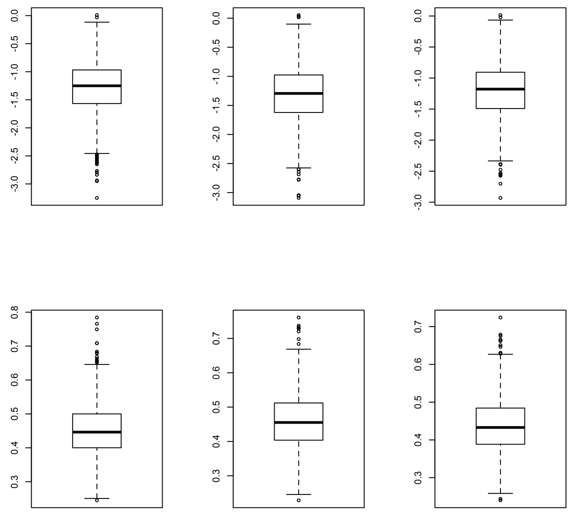

Figure 4.3. Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III

Table 4.2. Biases and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	V	0.1973	-0.0353	0.1588	0.0045
MHDE	VI	-0.4018	0.0734	0.3288	0.0097
	VII	0.9609	-0.0687	1.0349	0.0081
	V	0.1466	-0.0271	0.1605	0.0044
SCDE	VI	-0.4422	0.0808	0.3839	0.0114
	VII	0.9234	-0.057	0.9723	0.0072
	V	0.2309	-0.0415	0.1656	0.0048
MLE	VI	-0.3501	0.0640	0.2749	0.0080
	VII	0.9841	-0.0751	1.0745	0.0088

have some protection when the postulated models have long-tailed contamination and the SCDE is slightly better. When a strict-increasing contamination is present, the MHDE provides a smaller MSE for θ_0 in model V only.

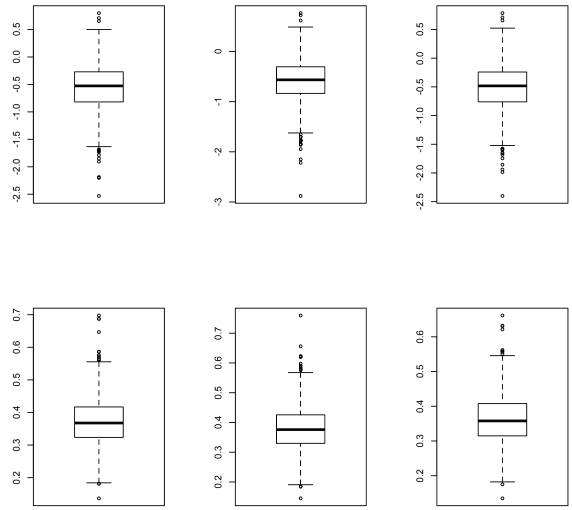

Figure 4.4. Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV

The bahavior of these three estimators are displayed in boxplots and histograms as well, see Figures 4.1 to 4.11. Boxplots and histograms show that all three estimators are subject to different skewness. In particular, they are left-skewed for θ_0 and right-skewed for θ_1. The MHDE has less skewed distributions than the SCDE overall. Histograms further validate this observation.

Biases and MSEs are also studied with the true parameter set equal to $\theta_0 = -3, \theta_1 = 0.75$ with adjusted location and scale for the logistic CDF, see Tables 4.3 and 4.4. Again, the observations are similar in nature to the previous case of the true parameter.

4.2. Normal CDF

In this section, the CDFs of $N(1.2, 1.21)$ and $N(1.125, 2.27)$ distributions are used as the CDF for F in the model. Again, the Models I to VII used in the previous section are

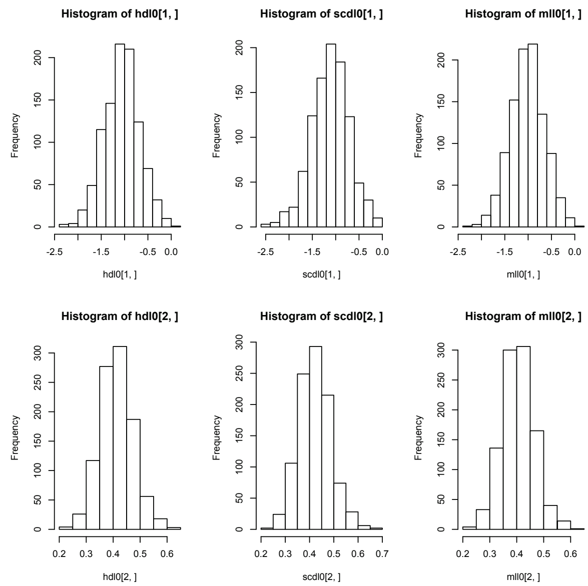

Figure 4.5. Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I

Table 4.3. Biases and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	I	-0.0994	0.0167	0.4552	0.0124
	II	0.1449	-0.0276	0.4648	0.0127
	III	-0.4581	0.0825	0.7287	0.0210
	IV	0.8809	-0.0584	1.1919	0.0157
SCDE	I	-0.1827	0.0316	0.5363	0.0145
	II	0.0590	-0.0122	0.4932	0.0131
	III	-0.5375	0.0958	0.8676	0.0247
	IV	0.8108	-0.0399	1.1137	0.0155
MLE	I	-0.0262	0.0035	0.4156	0.0112
	II	0.2166	-0.0406	0.4584	0.0126
	III	-0.3704	0.0664	0.6202	0.0175
	IV	0.9356	-0.0704	1.2726	0.0165

used to compare the MHDE, SCDE and MLE here as well. We observed the following: the SCDE performs better than the MHDE and MLE in Model II, but it has a higher

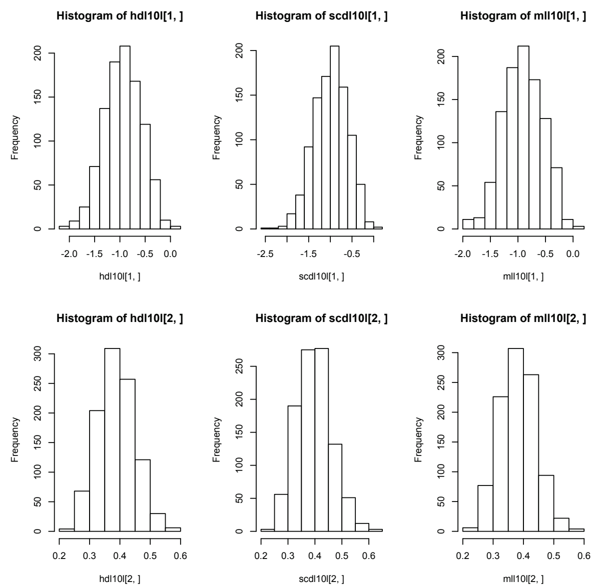

Figure 4.6. Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II

Table 4.4. Biases and MSEs of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V-VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	V	0.3107	-0.0591	0.5693	0.0164
MHDE	VI	-0.7624	0.1360	1.1223	0.0335
	VII	1.7920	-0.1311	3.5995	0.0287
	V	0.2308	-0.0449	0.5818	0.0162
SCDE	VI	-0.8410	0.1507	1.3237	0.0392
	VII	1.7168	-0.1085	3.3741	0.0258
	V	0.3780	-0.0712	0.5843	0.0171
MLE	VI	-0.6693	0.1193	0.9536	0.0281
	VII	1.8360	-0.1433	3.7422	0.0313

bias in Model IV. For Model IV, the MHDE has the edge over the other two. A higher contamination rate case (i.e., Models V, VI and VII) shows the difference in terms of bias and mean squared error begin to diminish, while MHDE and SCDE still perform superior

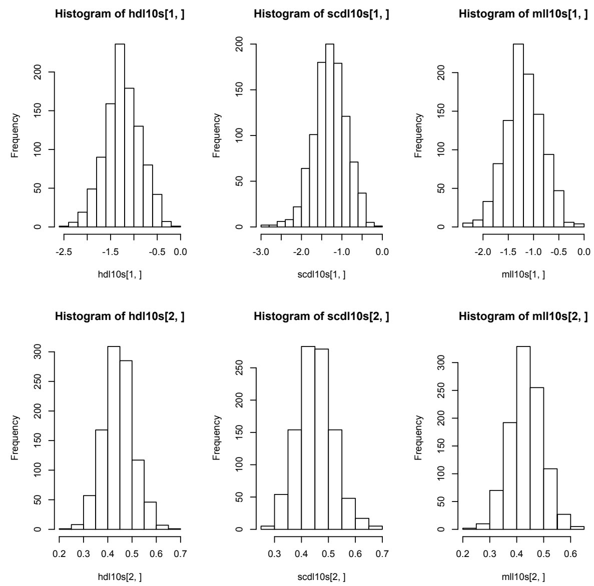

Figure 4.7. Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III

to the MLE in Models V and VII. This may suggest that both the MHDE and SCDE offer greater protection if the postulated model is of long-tailed contamination.

Table 4.5. Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I-IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	I	-0.1026	0.0189	0.0882	0.0026
	II	0.0701	-0.0405	0.0740	0.0034
	III	-0.1103	0.0324	0.0969	0.0035
	IV	0.4566	-0.0436	0.2662	0.0037
SCDE	I	-0.0509	0.0096	0.0704	0.0019
	II	0.0696	-0.0414	0.0686	0.0033
	III	-0.0477	0.0168	0.0702	0.0021
	IV	0.4906	-0.0475	0.2962	0.0038
MLE	I	-0.0294	0.0056	0.0638	0.0018
	II	0.1257	-0.0495	0.0731	0.0039
	III	-0.0288	0.0169	0.0659	0.0022
	IV	0.5049	-0.0541	0.3051	0.0044

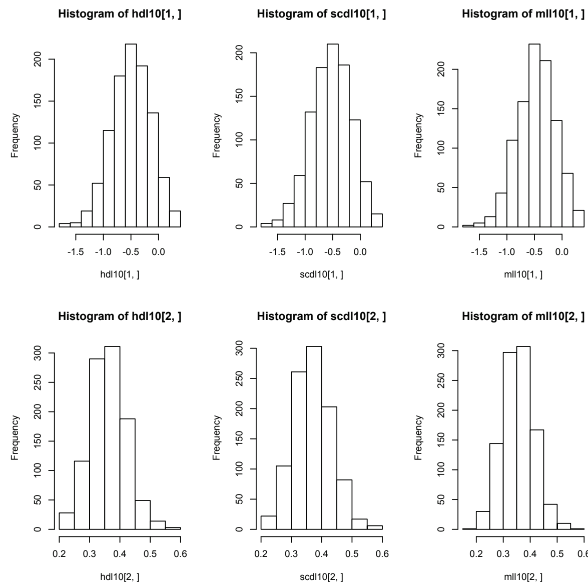

Figure 4.8. Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV

Table 4.6. Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF with $\theta_0 = -1, \theta_1 = 0.4$, Model V to VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	V	0.0901	-0.0427	0.0730	0.0035
MHDE	VI	-0.1398	0.0516	0.1111	0.0055
	VII	0.8729	-0.0844	0.8088	0.0087
	V	0.0910	-0.0439	0.0752	0.0036
SCDE	VI	-0.0522	0.0277	0.0737	0.0026
	VII	0.8907	-0.0837	0.8362	0.0084
	V	0.1440	-0.0515	0.0752	0.0041
MLE	VI	-0.0441	0.0330	0.0707	0.0031
	VII	0.9134	-0.0947	0.8760	0.0102

When a sample size in each group is less than 20, the SCDE shows good robust property with a price of unstable performance and indistinguishable advantage compared to the MLE. However, with a moderate sample size of $n = 30$, the performance of SCDE is quite stable and is comparable to the MLE and MHDE. When $\theta_0 = -3, \theta_1 = 0.75$,

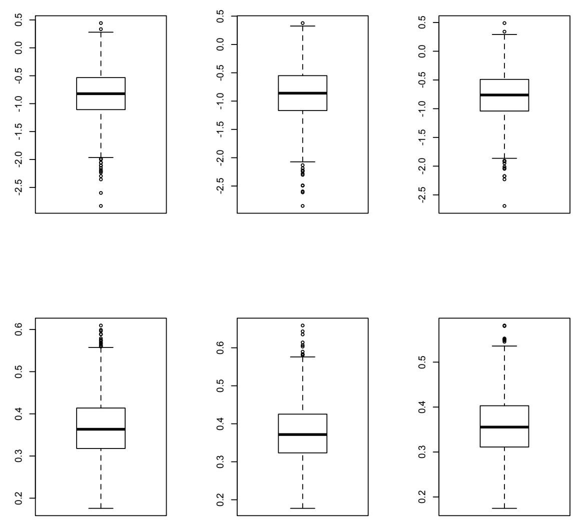

Figure 4.9. Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V

both bias and MSE of MHDE is better than those of SCDE, which are both better than those of MLE. In particular, the SCDE has bias and MSE smaller than those of MLE at a short-tailed contamination rate of 20%. It suggests that the SCDE might offer some protection at short-tail as well when contamination gets larger.

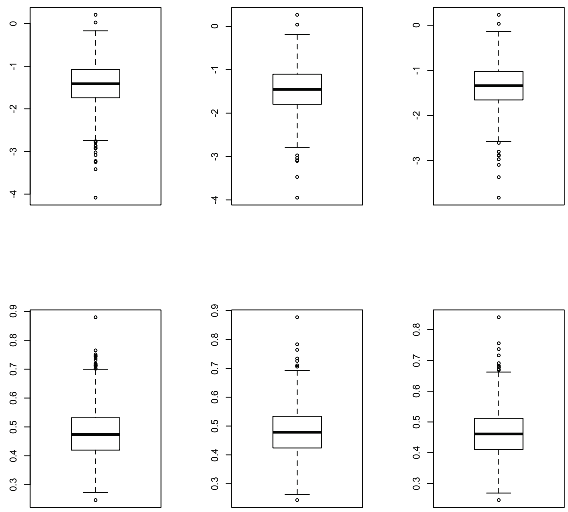

Figure 4.10. Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI

Table 4.7. Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model I to IV

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
	I	-0.1824	0.0347	0.3002	0.0085
MHDE	II	0.2168	-0.0676	0.2892	0.0110
	III	-0.3095	0.0687	0.3921	0.0130
	IV	0.8424	-0.0806	0.9185	0.0129
	I	-0.0887	0.0166	0.2246	0.0058
SCDE	II	0.2381	-0.0710	0.2894	0.0111
	III	-0.1711	0.0396	0.2638	0.0075
	IV	0.9199	-0.0890	1.0403	0.0136
	I	-0.0478	0.0100	0.2163	0.0058
MLE	II	0.3262	-0.0861	0.3068	0.0128
	III	-0.1564	0.0401	0.2553	0.0079
	IV	0.9320	-0.0997	1.0506	0.0153

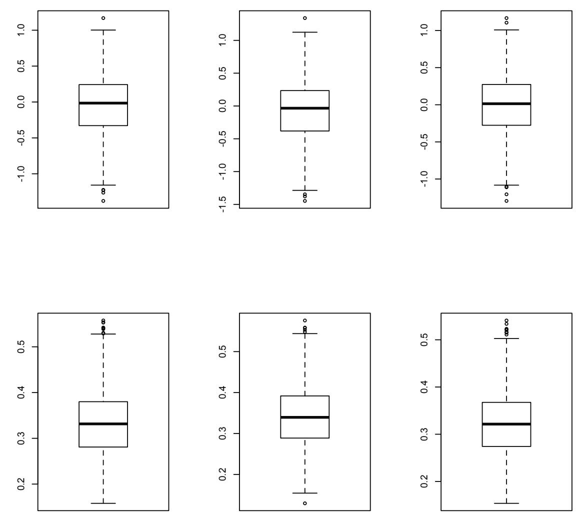

Figure 4.11. Boxplots of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII

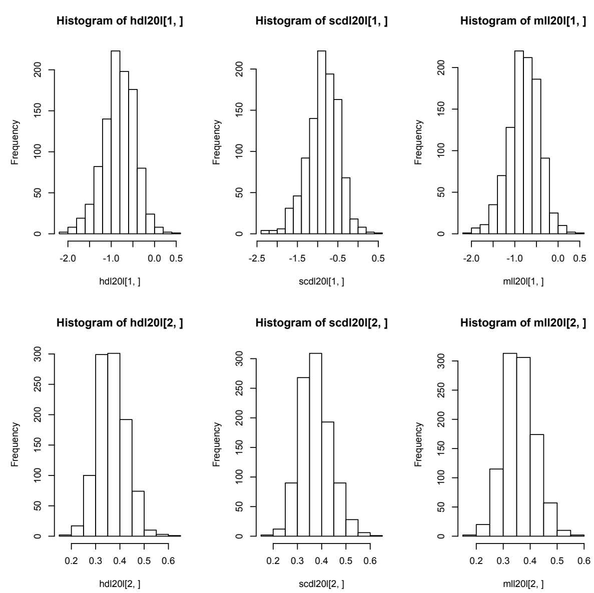

Figure 4.12. Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V

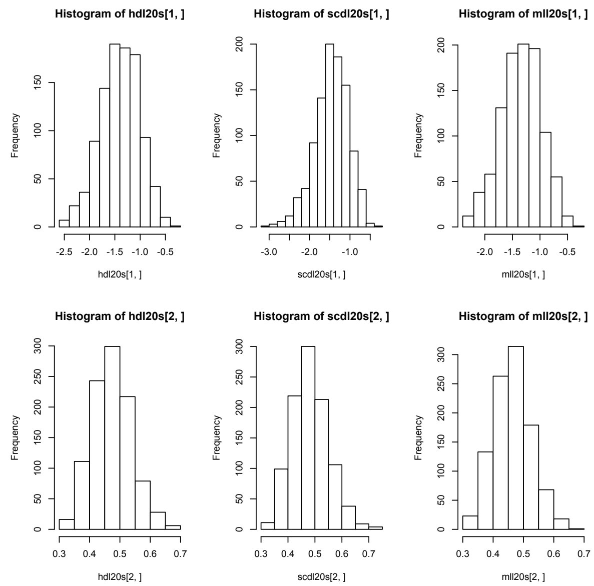

Figure 4.13. Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI

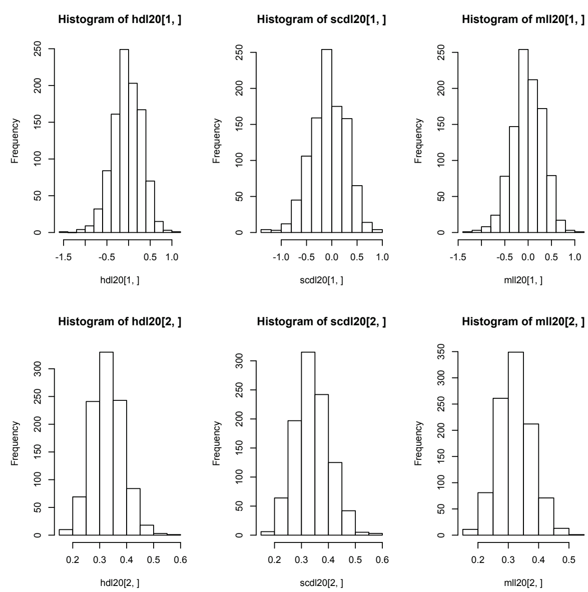

Figure 4.14. Histograms of MHDE, SCDE and MLE under a Logistic CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII

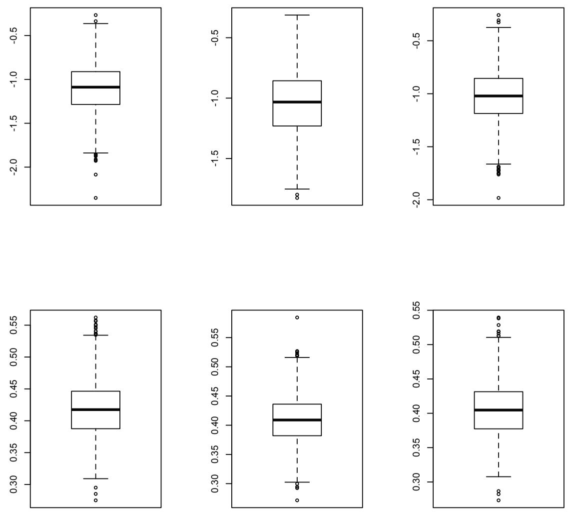

Figure 4.15. Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I

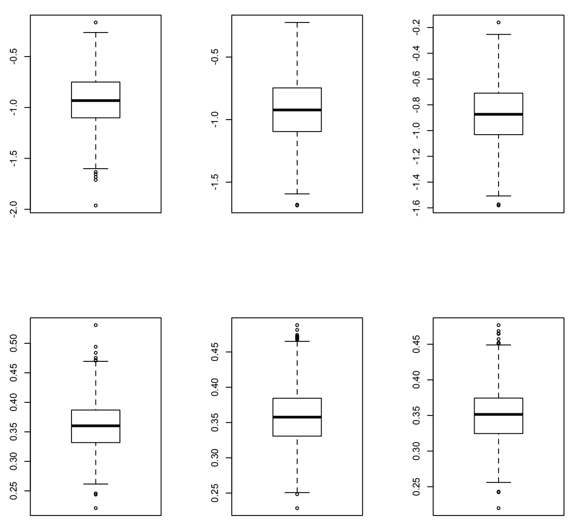

Figure 4.16. Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II

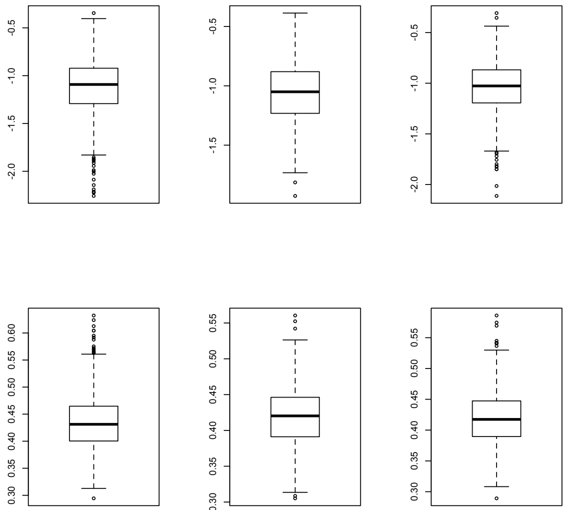

Figure 4.17. Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III

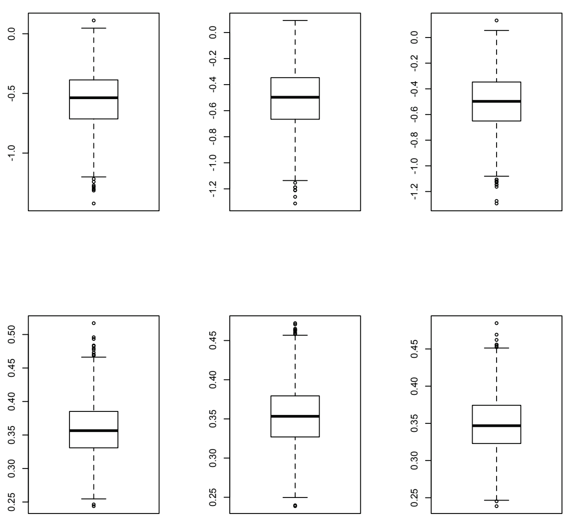

Figure 4.18. Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV

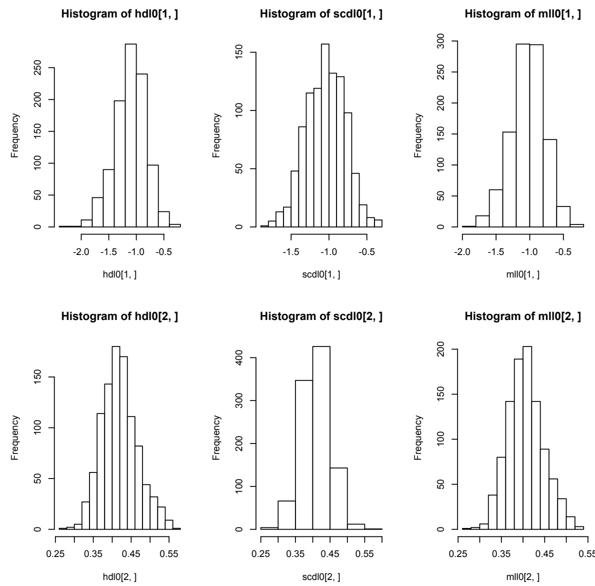

Figure 4.19. Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model I

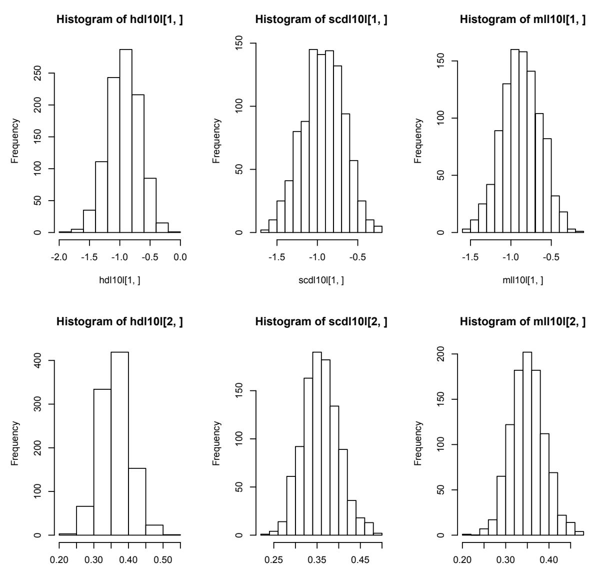

Figure 4.20. Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model II

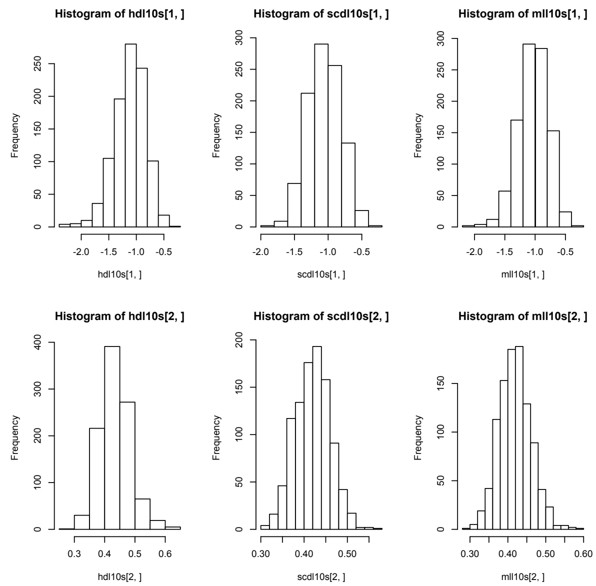

Figure 4.21. Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model III

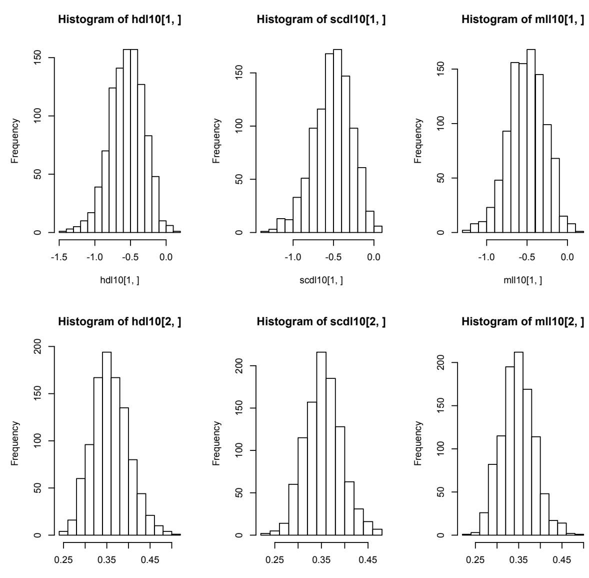

Figure 4.22. Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model IV

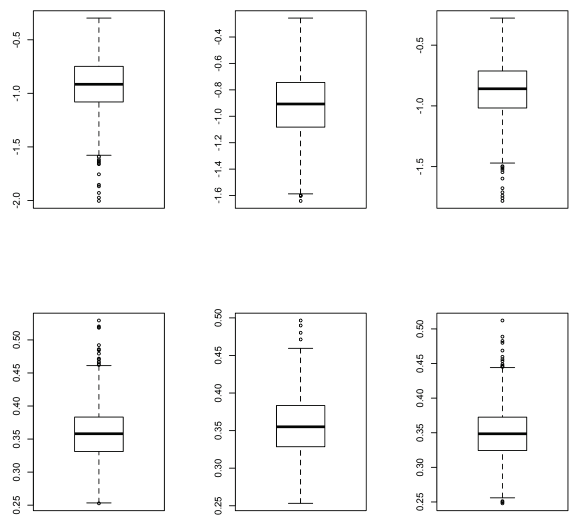

Figure 4.23. Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V

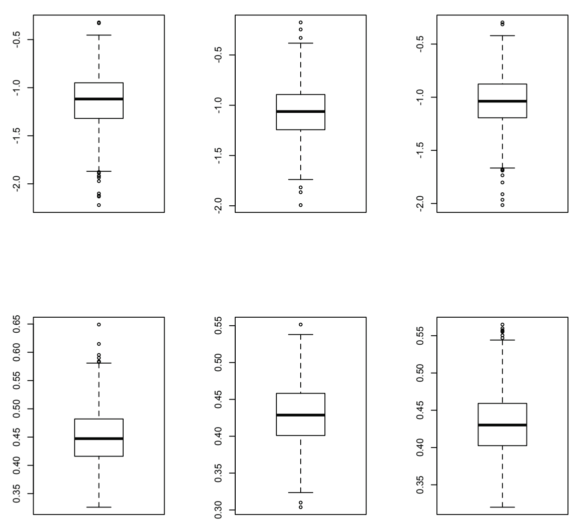

Figure 4.24. Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI

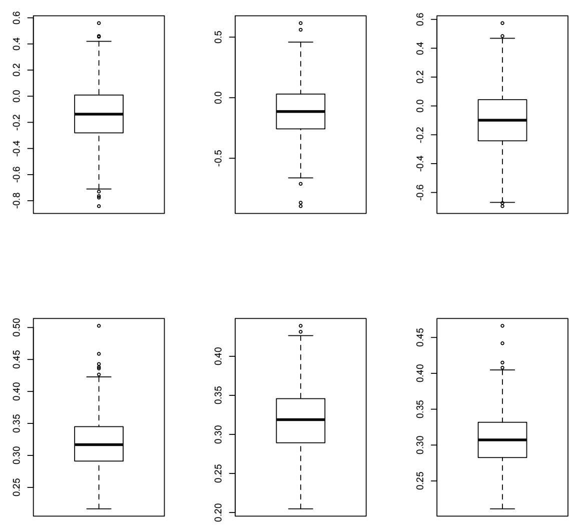

Figure 4.25. Boxplots of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII

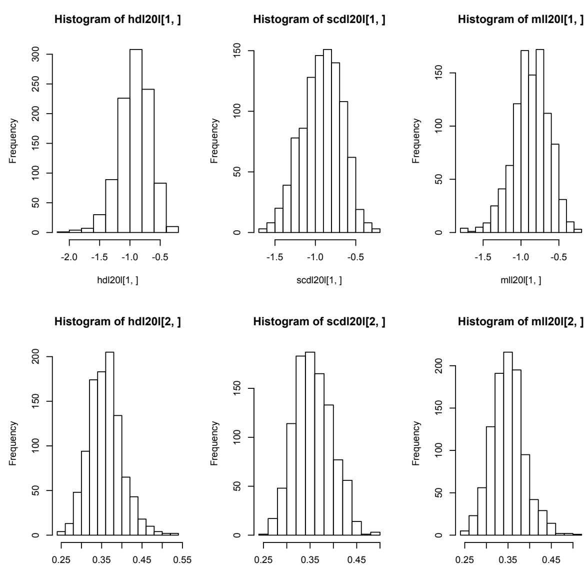

Figure 4.26. Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model V

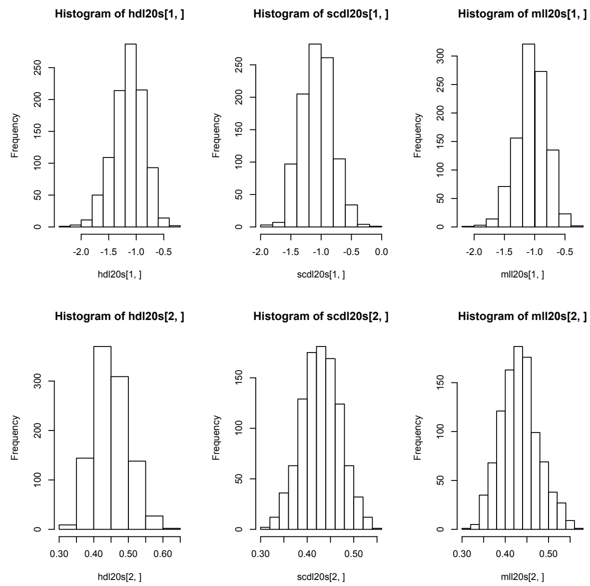

Figure 4.27. Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VI

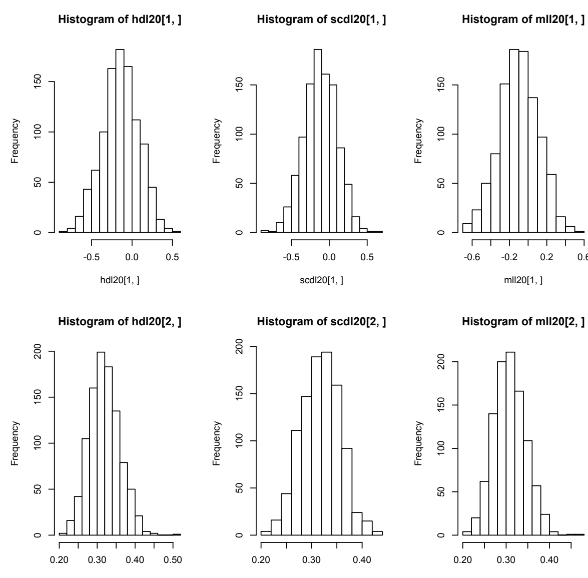

Figure 4.28. Histograms of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -1, \theta_1 = 0.4$, Model VII

Table 4.8. Biases and MSEs of MHDE, SCDE and MLE under a Normal CDF and $\theta_0 = -3, \theta_1 = 0.75$, Model V to VII

Method	Model	$Bias(\theta_0)$	$Bias(\theta_1)$	$MSE(\theta_0)$	$MSE(\theta_1)$
MHDE	V	0.2465	-0.0703	0.2886	0.0111
	VI	-0.4752	0.1107	0.5363	0.0215
	VII	1.5980	-0.1508	2.7292	0.0284
SCDE	V	0.2478	-0.0710	0.3013	0.0113
	VI	-0.2545	0.0630	0.3015	0.0102
	VII	1.6432	-0.1534	2.8590	0.0285
MLE	V	0.3490	-0.0878	0.3136	0.0130
	VI	-0.2962	0.0769	0.3222	0.0127
	VII	1.6729	-0.1697	2.9552	0.0335

CHAPTER 5

Summary and Future work

In this thesis, I have put emphasis on analyzing the asymptotic behaviour and robust property of two minimum distance estimators for a discrete binomial model. These two estimators, namely MHDE and SCDE, are based on minimizing under different metric measures other than the Euclidean metric. These estimators are generally known to be robust to outliers and model misspecifications.

In Chapters 2 and 3, consistency and asymptotic properties of the proposed MHDE and SCDE have been discussed through series of mathematical approaches and followed by simulation examples comparing with the corresponding MLE in order to show robustness properties of MHDE and SCDE. In Chapter 4, these two estimators along with the MLE are compared by simulation studies.

Understanding the behaviour of an estimator requires a full investigation of combination of link function, data frame $\underline{\boldsymbol{X}}$ and the parameters chosen. This is plausible for a discrete case analysis, given the condition that distribution of X is known and number of covariates is small. But as one can see, even in a case assuming independence between any of two covariates, exhausting combination of all k covariates can quickly explode and massive calculation would be cumbersome to compute. This is also not possible in a clinical trial, where absent observations (or not enough observations, i.e. less than 15) take place within some of the categories. Distribution of covariates is known in a simulated setting, but it is not known in a real-life setting. As for future work, one should try to fit various distributions (i.e., continuous and discrete), to further examine the behaviour of proposed estimators. Also, location and scale parameters for a "*correct*" cumulative density function matters, and the relationship of θ and the location and scale parameter is worth investigating.

References

[1] Banerjee, H. (2010). *Estimation of parameters for logistic regression model in dose response study with a single compound or mixture of compounds*. Ph.D dissertation, University of California Riverside, California.

[2] Beran, R. (1977). Minimum Hellinger Distance Estimates for Parametric Models. *The Annals of Statistics*, **5**, 445-463.

[3] Bianco, A. M., and Yohai, V. J. (1996). Robust estimation in the logistic regression model. In *Robust Statistics, Data Analysis, and Computer Intensive Methods* (pp. 17-34). Springer, New York.

[4] Cantoni, E., and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, **96**, 1022-1030.

[5] Carroll, R. J., and Pederson, S. (1993). On robustness in the logistic regression model. *Journal of the Royal Statistical Society*, **55**, 693-706.

[6] Casella, G., and Berger, R. L. (2002). *Statistical Inference*. Duxbury, California.

[7] Croux, C., and Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics and Data Analysis*, **44**, 273-295.

[8] Copas, J. B. (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society*, **50**, 225-265.

[9] Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, **42**, 1887-1896.

[10] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383-393.

[11] Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, **35**, 73-101.

[12] Karunamuni, R. J., Tang, Q., and Zhao, B. (2015). Robust and efficient estimation of effective dose. *Computational Statistics and Data Analysis*, **90**, 47-60.

[13] Künsch, H. R., Stefanski, L. A., and Carroll, R. J. (1989). Conditionally unbiased bounded-influence estimation in general regression models, with applications to generalized linear models. *Journal of the American Statistical Association*, **84**, 460-466.

[14] Li, P., and Wiens, D. P. (2011). Robustness of design in dose-response studies. *Journal of the Royal Statistical Society*, **73**, 215-238.

[15] Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *The Annals of Statistics*, **22**, 1081-1114.

[16] Ellis, S. P., and Morgenthaler, S. (1992). Leverage and breakdown in L_1 regression. *Journal of the American Statistical Association*, **87**, 143-148.

[17] Maronna, R. A. R. D., Martin, D., and Yohai, V. (2006). *Robust Statistics*. John Wiley & Sons, New York.

[18] Markatou, M., Basu, A., and Lindsay, B. (1997). Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *Journal of Statistical Planning and Inference*, **57**, 215-232.

[19] Marazzi, A., and Ruffieux, C. (1996). Implementing M-estimators of the Gamma Distribution. In *Robust Statistics, Data Analysis, and Computer Intensive Methods* (pp. 277-297). Springer, New York.

[20] McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, New York.

[21] Shorack, G. R. (2000). *Probability For Statisticians*. Springer, New York.

[22] Stather, C. R. (1981). *Robust statistical inference using Hellinger distance methods*. Ph.D dissertation, LaTrobe University.

[23] Stefanski, L. A., Carroll, R. J., and Ruppert, D. (1986). Optimally hounded score functions for generalized models with applications to logistic regression. *Biometrika*, **73**, 413-424.