12-2003

A Robust Two-Sample Procedure to Estimate a Shift Parameter

Feridun Tasdan
Western Michigan University

A ROBUST TWO-SAMPLE PROCEDURE TO ESTIMATE A SHIFT PARAMETER

by

Feridun Tasdan

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Statistics

Western Michigan University
Kalamazoo, Michigan
December 2003

A ROBUST TWO-SAMPLE PROCEDURE TO ESTIMATE A SHIFT PARAMETER

Feridun Tasdan, Ph.D.

Western Michigan University, 2003

This study estimates the location shift parameter in the two-sample problem. The classical method, Least Square(LS), obtains the shift parameter estimate under the normality assumption. A departure from normality assumption makes the estimate inefficient and unreliable. One alternative to the least square estimate is Hodges-Lehmann(HL) estimate which uses Wilcoxon ranks to estimate the shift parameter. This estimate is robust against contaminations and large outliers. The proposed method in this study combines two samples and uses convolution technique to find a density function for the combined sample. This new density function is later used in the construction of the log likelihood function. By the quasiconvexity of log likelihood function, a minimization procedure finds the estimate of the shift parameter. Asymptotic properties of this estimator is established under conditions that are similar to those used in LS and HL. Among those properties, the asymptotic linearity and asymptotic normality conditions are satisfied and found in the latter case. As shown in the study, the proposed estimator is highly efficient and robust against contaminations and outliers. This result is supported by the real data examples and by a bootstrap simulation study.

ACKNOWLEDGMENTS

I would like to sincerely express my gratitude to my advisor Dr. Gerald Sievers. I am especially indebted to him for his guidance, helpful comments and support that helped me accomplish this difficult task. My thanks go to all of my friends and family members for their support and encouragements that gave me confidence in completing my Ph.D. Finally, a special person deserves the most credit, I express deep gratitude to my wife, Muruvvet, for her love, sacrifice and support during my Ph.D study.

Feridun Tasdan

TABLE OF CONTENTS

Table of Contents—Continued

CHAPTER

Table of Contents—Continued

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Definition of the Problem

Let $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_2}$ be two independent i.i.d random samples with common distribution functions F(x) and G(y), respectively. We assume a relationship of G(y)=F(y-Δ_0) which is also called the location-shift model. In general, two problems are often studied in this set up. First problem is the testing of the hypotheses

$$H_0 : F(x) = G(y) \text{ vs } H_A : F(x) \neq G(y)$$

where H_0 implies that the sample X and the sample Y come from the same population distribution function under H_0. Equivalently, above null and alternative hypotheses can also be expressed in terms of shift parameter Δ_0

$$H_0 : \Delta_0 = 0 \, vs \, H_A : \Delta_0 \neq 0$$

Most of the procedures in the literature only deal with this problem, testing of H_0 vs H_A. The main interest is whether or not two samples come from the same population distribution function F(x).

Second type of problem often seen in the literature is the estimation of the shift parameter Δ_0. The shift parameter Δ_0 is sometimes called the treatment effect in the literature if the two random samples are classified as control and treatment groups. Δ_0 is the expected effect due to the treatment. If Δ_0 is positive, it is the expected increase due to the treatment. If Δ_0 is negative, it is the expected decrease due to the treatment.

We should note that, in many cases, neither of the random samples are classified as treatment or control group. The procedures we discuss here are applicable even if the samples are not related to drug testing problems.

1.2 Least Square Estimation

The problem of estimating shift parameter Δ_0 in the two-sample problem has been studied extensively in the past. The least square estimation method first introduced by Gauss in the 18th century can be used to derive the shift parameter estimate in the two-sample location problem. Hettmansperger and McKean(1998) described the two-sample location problem in terms of a linear model as follows:

$$Z_i = \Delta_0 c_i + e_i, \qquad 1 \leq i \leq n,$$

where $e_1, ..e_n$ are i.i.d with distribution function F(x), c_i is the indicator function

$$c_i = \begin{cases} 0, & \text{if } 1 \leq i \leq n_1 \\ 1, & \text{if } n_1 + 1 \leq i \leq n \end{cases}$$

and Z_i is the ith element of the combined sample of $(X_1, ..X_{n_1}, Y_1, ..Y_{n_2})$. The classical least squares estimation approach can be used to estimate the shift parameter Δ_0 based on the following square pseudo-norm,

$$D_{LS}(\Delta) = \| \mathbf{Z} - \mathbf{C}\Delta \|_{LS}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}(Z_i - c_i\Delta - Z_j + c_j\Delta)^2$$

By taking the first derivative of the above $D(\Delta)$ function with respect to Δ, we get the gradient function

$$S_{LS}(\Delta) = -4n_1n_2(\bar{Y} - \bar{X} - \Delta)$$

Therefore, setting $S_{LS}(\Delta)$ equal to zero, we get the classical least square estimate of Δ_0 which is

$$\widehat{\Delta}_{LS} = \bar{Y} - \bar{X}$$

It can be shown that

$$\widehat{\Delta}_{LS} \text{ is approximately } N(\Delta_0, \sigma^2(n_1^{-1} + n_2^{-1}))$$

By setting $\Delta = 0$ in $S_{LS}(\Delta)$, we can find the test statistic

$$S_{LS}(0) = \bar{Y} - \bar{X}$$

Then, under H_0, the classical two-sampled pooled t-test statistic based on the above results is defined by

$$t = \frac{(\bar{Y}-\bar{X})}{S_{pool}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

where

$$S_{pool} = \sqrt{\frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}$$

This is a well known testing procedure for comparing the locations of two samples. It can be found in almost every elementary statistics textbook. Under the normality of population distribution functions assumption, $\widehat{\Delta}_{LS}$ and pooled t-test are the best for estimating the shift parameter Δ_0 and testing the null hypothesis in the two-sample location problem. On the other hand, if there is contamination in the data or departure from the normality assumption, then the result of the estimation and naturally the result of the testing procedure could be less than optimal. Therefore, both of these methods, $\widehat{\Delta}_{LS} = \bar{Y} - \bar{Y}$ and the two-sample t-test, are vulnerable to the outliers and depend heavily on the normality assumption (Student (1927), Tukey (1960), Hoyland (1965) and Huber (1972)). The two-sample t test should be used with a caution if the underlying distribution function is unknown or if there are large outliers in the data.

1.3 Hodges-Lehmann Estimation

The robust estimation and testing procedures are on the big move toward replacing classical methods in recent years. The reason for this, as we have mentioned above, is that the classical estimation method (Least Square) needs the data to be normally distributed (at least approximately normal or outlier free) but in the real world, the data does not follow these assumptions. Hample, Ronchetti, Rousseew, and Stahel (1986) argue that real data contain 1% to 10% gross errors in real life. Further studies by Huber(1972), and Hettmansperger and McKean(1998) also indicate that real life data can be contaminated. On the other hand, the robust estimation and testing procedures do not depend on the normality assumption and robust against outliers. Mostly, they are distribution free under the null hypothesis. The most widely used and well known robust procedure in the two sample problem is the Mann-Whitney-Wilcoxon(MWW) method. This method uses ranks and based on the pseudo-norm(L_1 norm) defined by

$$D_R(\Delta) = \| \mathbf{Z} - \mathbf{C}\Delta \|_R = \sum_{i=1}^{n}\sum_{j=1}^{n} | Z_i - c_i\Delta - Z_j + c_j\Delta |$$

By taking the first derivative of the dispersion function $D_R(\Delta)$ with respect to Δ, we get the MWW gradient function

$$S_R(\Delta_0) = -2\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} sgn(Y_j - X_i - \Delta_0)$$

By setting $S_R(\Delta) = 0$ and solving for Δ, we get

$$\widehat{\Delta}_R = med\{Y_j - X_i\}$$

which is the median of the pairwise differences. This is the pseudo-norm(L_1 norm) based on the estimate of the shift parameter Δ_0. The above estimator is called the Hodges and Lehmann(1963) estimator. Hodges and Lehmann(1963) and Hettmansperger and McKean(1998) showed that $\widehat{\Delta}_R$ is approximately normal with mean Δ_0 and variance $\sigma^2 = \frac{1}{12\lambda_1\lambda_2(\int f^2(x)dx)^2}$. Based on the MWW gradient function, Wilcoxon (1945) proposed rank based test statistic

$$W(\Delta) = \sum_{i=1}^{n_2} R(Y_i - \Delta)$$

Later, Mann and Whitney (1947) proposed the test statistic

$$S_R(\Delta) = \#(Y_j - X_i > \Delta).$$

Mann and Whitney(1947) showed that the linear relationship between Wilcoxon and Mann-Whitney is

$$S_R(\Delta) = W(\Delta) - \frac{n_2(n_2+1)}{2}$$

The statistic $S_R(\Delta)$ has been referred as the Mann-Whitney-Wilcoxon(MWW) test statistic. This derivation of the linear relationship can be found in Hettmansperger and McKean (1998).

Hettmansperger and McKean(1998) showed that the test statistic $S_R(\Delta)$ is distribution free under the null hypothesis H_0. Asymptotically, it is normally distributed with the mean $\mu = n_1n_2/2$ and standard deviation $\sigma = \sqrt{n_1n_2(n_1 + n_2 + 1)/12}$. Hence, based on the asymptotic results, an α level Z test statistic has been developed

$$Z = \frac{S_R(0)-(n_1n_2/2)}{\sqrt{n_1n_2(n_1+n_2+1)/12}}$$

The benefit of using robust procedures is that they don't lose efficiency and power under the contamination and presence of large outliers. If the normality assumption is satisfied, then robust methods still hold their ground against the least square estimation with asymptotic relative efficiency of 0.95, which was found by Hodges and Lehmann(1956). In some cases, even one outlier is enough to impair a two-sample t-test procedure and $\widehat{\Delta}_{LS}$, but MWW test and $\widehat{\Delta}_R$ remain robust and less sensitive to gross errors, see Hettmansperger and McKean(1998, Example 2.3.1).

1.4 General Rank Scores

The general rank score function estimate of the Δ_0 and testing procedure has been developed by Hettmansperger and McKean(1998). The pseudo norm for the scores is denoted by

$$D_\varphi(\Delta) = \| \mathbf{Z} - \mathbf{C}\Delta \|_\varphi = \sum_{i=1}^{n} a_\varphi(R(Z_i - c_i\Delta))(Z_i - c_i\Delta)$$

By taking the negative derivative of the $\| Z - C\Delta \|_\varphi$, we find gradient function

$$S_\varphi(\Delta) = \sum_{j=1}^{n_2} a_\varphi(R(Y_j - \Delta))$$

where a(i) are scores such that $a(1) \leq ... \leq a(n)$ and $\sum a(i) = 0$. This score function can be taken as $a(i) = \varphi(i/(n+1))$. If we take $\varphi_R = \sqrt{12}(u - 1/2)$, it is called Wilcoxon score and $\varphi_S = sgn(2u-1)$ is called sign score.

By setting $S_\varphi(\Delta) = 0$ and solving iteratively, we find $\widehat{\Delta}_\varphi$. This solution exist since the pseudo norm is a convex function of Δ. Hettmansperger and McKean(1998) found that

$$\widehat{\Delta}_\varphi \text{ is approximately } N(\Delta_0, \tau_\varphi^2(n_1^{-1} + n_2^{-1}))$$

where $\tau_\varphi^2 = \int \varphi(u)\varphi_f(u)du$

The test statistics, under H_0, is

$$S_\varphi(0) = \sum_{j=1}^{n_2} a_\varphi(R(Y_j))$$

This test statistics is distribution free and depends only on the ranks of the sample under the null hypothesis. The expected value and variance of S_φ are

$$E[S_\varphi] = 0$$

and

$$\sigma_\varphi^2 = \frac{n_1 n_2}{n-1}\left\{\sum_{i=1}^{n} a(i)\frac{1}{n}\right\} \doteq \frac{n_1 n_2}{n-1}$$

Then, for testing previously defined H_0 vs H_A, an asymptotic level α test is to

$$\text{Reject } H_0 \text{ in favor of } H_A \text{ if } S_\varphi(0) \geq z_\alpha \sigma_\varphi$$

1.5 Other Important Methods

Some other nonparametric procedures for the two-sample location problem are Fisher's Exact Test by Fisher (1936), Median Test by Mood(1950), Tukey's Quick Test(1959), Kolmogrov and Simirnov etc.

CHAPTER II

PROPOSED SOLUTION

2.1 Method and Notation

Let $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_2}$ be two independent i.i.d random samples with distribution functions F(x) and G(y)=F(x-Δ_0), respectively. As we know from the Chapter I, the parameter Δ_0 represents the true shift parameter between the locations of the two samples. The main interest in this dissertation is to estimate the true shift parameter Δ_0 and find asymptotic properties of the its estimator. Therefore, the shift in the locations of the two samples is our main focus. If we state our basic assumption:

Assumption (A1): $X_1, ..., X_{n_1}$ and $Y_1, ..., Y_{n_2}$ are two independent i.i.d random samples from continuous distribution functions F(x) and G(y)=F(x-Δ_0) with equal scale parameters.

In a linear model notation, the following representation is also used by Doksum(1974) and Hollander and Wolfe(1999).

$$Y - \Delta_0 \stackrel{d}{=} X$$

The notation $\stackrel{d}{=}$ indicates that $Y - \Delta_0$ and X have the same distribution function as F(x). If $\Delta_0 = \mu_y - \mu_x$ (the shift between the locations), then we have

$$Y - \mu_y \stackrel{d}{=} X - \mu_x$$

which indicates that by shifting, the data is centered.

The following figure shows the relationship between two distributions and the true shift parameter Δ_0.

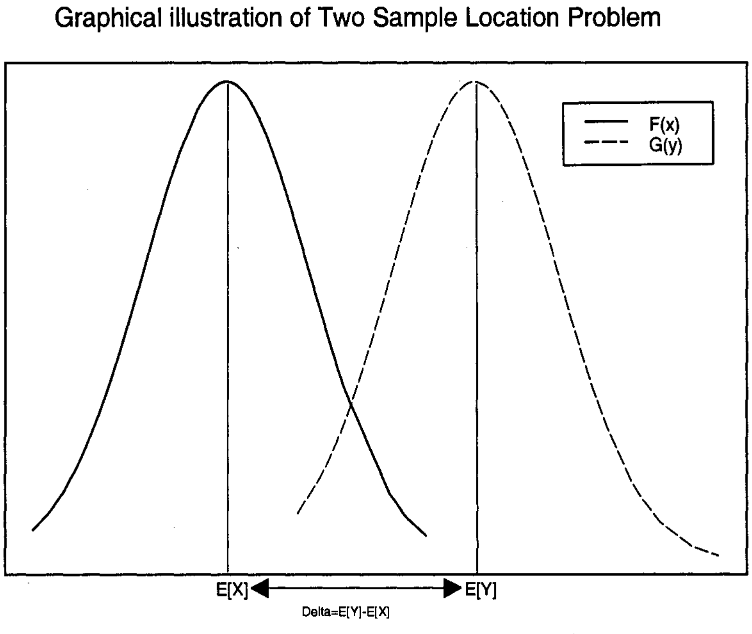

Figure 1: F(x) and G(y), shifted by Δ_0=E(Y)-E(X)

First, let $Y_j^\star = Y_j - \Delta$ for $j = 1, .., n_2$ be **arbitrary** Δ shifted Y-Sample and let

$$Z_k = \begin{cases} X_i, & \text{if } 1 \leq k \leq n_1; \\ \\ Y_j^\star, & \text{if } n_1 + 1 \leq k \leq n; \end{cases}$$

be the combined sample of $X_1, ..., X_{n_1}$; $Y_1^\star, ..., Y_{n_2}^\star$ and n $= n_1 + n_2$ is the combined sample size. We will call this the "combined shifted sample". The purpose of using the arbitrary Δ variable is to align two samples as closely as possible so that true shift parameter Δ_0 can be estimated. We should note that when we shift the sample Y with Δ parameter, the combined shifted sample Z_k can be taken as a single sample from the underlying distribution function F(x).

In this notation, for testing purposes, if we want to investigate that the difference in locations of the two samples is equal to $\Delta_0^\star$, then we should look at the null hypothesis,

$$H_0 : \Delta_0 = \Delta_0^\star \text{ vs } H_A : \Delta_0 > \Delta_0^\star$$

In addition, if we are testing that there is no difference in the locations of the two samples, then the null hypothesis becomes

$$H_0 : \Delta_0 = 0 \text{ vs } H_A : \Delta_0 > 0.$$

2.2 Likelihood Function of the Combined Sample

To estimate the true shift parameter Δ_0, we use the idea of the likelihood function which is the joint probability density function of the combined shifted sample Z_k. The maximization of the likelihood function with respect to the arbitrary shift variable Δ estimates the true shift parameter Δ_0 which aligns the two samples as closely as possible. We first define the likelihood function of random variable Z_k,

$$\begin{aligned} \ell(\Delta) &= \prod_{k=1}^{n} f(z_k) \\ &= \prod_{i=1}^{n_1} f(x_i) * \prod_{j=1}^{n_2} f(y_j^\star) \\ &= \prod_{i=1}^{n_1} f(x_i) * \prod_{j=1}^{n_2} f(y_j - \Delta) \end{aligned} \tag{2.1}$$

Then, by taking the negative log of both sides, we get the negative log likelihood function,

$$L(\Delta) = -\sum_{i=1}^{n_1} log[f(x_i)] - \sum_{j=1}^{n_2} log[f(y_j - \Delta)] \tag{2.2}$$

The negative log likelihood function defined above would be a convex function of Δ as long as we have a log concave density function f(x). Therefore, by minimizing the log likelihood function with respect to Δ, the estimate of the true shift parameter Δ_0 can be found. The problem is that we rarely know the true distribution function F(x) or density function f(x) of the populations in real life. Thus, in order to use this idea, we need to know the true density function of the combined sample or we should find one that replaces the true density function.

2.3 Kernel Density Estimation

The estimation of density functions is not a new idea. The Kernel density estimation methods first introduced by Rosenbalt(1956), Parzen(1962) and Cencov(1962). The Kernel estimator is a numerical approximation to the derivative of the empirical distribution function $F_n(x)$. The basic kernel estimator can be written simply as

$$\hat{f}(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K(\frac{x - X_i}{b_n}) = \frac{1}{n}\sum_{i=1}^{n} K_{b_n}(x - X_i) \tag{2.3}$$

There are a lot of studies and suggestions in the literature for selecting the kernel function K(x) and bandwidth parameter b_n. The choice of the kernel function K(x) is, to some extent, arbitrary, and properties of $\hat{f}(x)$ will depend mostly on the chosen value of the bandwidth parameter b_n. There is always a trade-off in the selection of the bandwidth parameter b_n. As we decrease the value of b_n, the bias of $\hat{f}(x)$ decreases but the variance increases. Conversely, as we increase the b_n, the bias of $\hat{f}(x)$ increases but the variance decreases; see Silverman (1986, section 4.3.1).

A complete solution to the problem of selecting the bandwidth parameter b_n is not currently available, although some promising work has been done in recent years. Sheather and Hettmansperger (1985) studied the selection of bandwidth in the rank regression. Silverman (1986), Hall and Marron (1987), Park and Marron (1990), and Sheather and Jones (1991) all proposed an optimum bandwidth parameter selection procedure. Sheather and Jones(1991) stressed that their bandwidth estimator has a practical performance second to none in the existing literature on the subject. In this dissertation, we will run a simulation study to investigate the kernel density estimation method as an alternative to our "smoothing by convolution" idea by using bandwidth b_n as a smoothing parameter. For the estimation of the bandwidth parameter b_n, we will be using the optimum bandwidth parameter proposed by Sheather and Jones(1991).

2.4 Smoothing by Convolution

The idea of smoothing functions by convolution has been used mostly in the mathematical sciences, engineering, physics, computer science and statistics. For example, mathematicians use Fourier transformations to approximate functions. An electronics engineer facing a noisy function will use one of the convolution filters to smooth away unwanted noisy functions. Engineers call this filter design; see Scott (1992). Similarly, a physicist would like to use convolution smoothing to smooth away unwanted high frequency components. One of the recent hot topics in computer science involves image recognition algorithms. Gaussian density function has been used in the edge smoothing of estimates of the real images. The problem in statistics is that we can have a sample set drawn from a population whose density function is unknown to us. The empirical density function of the sample can be smoothed with a weight function by using the convolution operation that makes the empirical density function smooth, less noisy and useable in the statistical inference. Cencov(1962), Kronmal and Tarter (1968), and Watson (1969) suggested smoothing the empirical density functions using orthogonal series approximation which uses Fourier transformation as a basis.

The usual nonparametric estimate of a distribution function F(x) is the empirical distribution function $F_n(x)$. It is a very common practice in statistics to use the empirical distribution function $F_n(x)$ instead of the unknown F(x) in statistical inference or testing problems. For example, the two sample Kolmogrov-Simirnov test uses the empirical distribution functions of the two samples and finds the maximum distance between them for testing purposes. There is a very strong theoretical background concerning $F_n(x)$ and F(x). The Glivenko-Cantelli Theorem states that $F_n(x)$ converges to F(x) uniformly rather than pointwise without restrictions to continuity points of F(x). Whenever we are have a data set, we can find an empirical distribution function $F_n(x)$ of the given set. These are the main reasons that we would like to use an empirical distribution function $F_n(x)$ as a main tool in the replacement of real density function. So, given that we have the combined shifted sample Z_k, we can find the empirical distribution function of the Z_k in the following fashion.

$$
\begin{aligned}
F_n^\star(x) &= \frac{1}{n}\sum_{k=1}^{n}\mathbf{I}(Z_k < x) \\
&= \frac{1}{n}\sum_{i=1}^{n_1}\mathbf{I}(X_i < x) + \frac{1}{n}\sum_{j=1}^{n_2}\mathbf{I}(Y_j - \Delta < x) \\
&= \frac{n_1}{n} * F_{n_1}(x) + \frac{n_2}{n} * G_{n_2}(x + \Delta) \qquad (2.4)
\end{aligned}
$$

where $\mathbf{I}$ is the indicator function defined as

$$
\mathbf{I(A)} = \begin{cases} 1, & \text{if event A is true} \\ 0, & \text{otherwise} \end{cases}
$$

and F_{n_1} and G_{n_2} are the empirical distribution functions for the X and Y samples respectively. Note that $F_n^\star(x)$ also depends on the arbitrary variable Δ.

We can not put $F_n^\star(x)$ in negative log likelihood function directly but one can think that we can use the theoretical relationship of $f(x) = F'(x)$ between $f_n(x)$ and $F_n(x)$. In reality, the empirical density function $f_n(x)$ is

$$
f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\delta(x - X_i)
$$

where $\delta(t)$ function is called the Dirac delta function by Scott(1992). It is always a **discrete uniform density** over the data with probability mass of 1/n for each data point. On the graph, it can be seen as one-dimensional scatter diagram. It is a useless estimate especially if the real density is continuous. So, we cannot use $f_n(x)$ in the negative log likelihood function either.

To overcome the difficulty of using empirical density function, we introduce the "smoothing by convolution" idea to find a replacement for the unknown density function f(x) in the log likelihood function. By using a convolution of $f_n(x)$ with a smooth continuous density function, the resulting empirical density function becomes a smooth and continuous density and can be used as a replacement of f(x) in the log likelihood function.

The basic notation for the convolution in one sample case is

$$\widehat{h}(t) = \int g(t-x)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} g(t-x_i) \tag{2.5}$$

For the two-sample problem, choose a continuous "smoother density" function g(t) and convolute with combined empirical distribution function $F_n^\star(x)$. The resulting convolution becomes,

$$\begin{aligned}
\widehat{h}_\Delta(t) &= \int g(t-x)dF_n^\star(x) \\
&= \int g(t-x)d[(n_1/n) * F_{n_1}(x) + (n_2/n) * G_{n_2}(x+\Delta)] \\
&= \frac{n_1}{n}\int g(t-x)dF_{n_1}(x) + \frac{n_2}{n}\int g(t-x)dG_{n_2}(x+\Delta)
\end{aligned}$$

$$= \frac{n_1}{n} \int g(t-x) dF_{n_1}(x) + \frac{n_2}{n} \int g(t-y+\Delta) dG_{n_2}(y)$$

If we apply the Reimann-Stieljes integral, we have

$$= \frac{n_1}{n}\frac{1}{n_1}\sum_{i=1}^{n_1} g(t-x_i) + \frac{n_2}{n}\frac{1}{n_2}\sum_{j=1}^{n_2} g(t-y_j+\Delta)$$

$$\widehat{h}_\Delta(t) = \frac{1}{n}\left[\sum_{i=1}^{n_1} g(t-x_i) + \sum_{j=1}^{n_2} g(t-y_j+\Delta)\right] \tag{2.6}$$

The resulting function $\widehat{h}_\Delta(t)$ is a smoothed empirical density function of the combined shifted sample according to the density function g(t) chosen in the convolution. In addition, it should be noted that $\widehat{h}_\Delta(t)$ depends on the arbitrary shift variable Δ.

Let $g(t) = \frac{1}{\sigma}g(\frac{t}{\sigma})$, where σ is a scale parameter. If we replace g(t) with $\frac{1}{\sigma}g(\frac{t}{\sigma})$ in the expression(2.6), we have,

$$\widehat{h}_\Delta(t) = \frac{1}{n\sigma}\left[\sum_{i=1}^{n_1} g(\frac{t-x_i}{\sigma}) + \sum_{j=1}^{n_2} g(\frac{t-y_j+\Delta}{\sigma})\right] \tag{2.7}$$

We should emphasize that $\widehat{h}_\Delta(t)$ is not an exact estimate of f(x) but it will have the basic shape of the density that the data comes from. If we compare kernel density expression (2.3) with smoothing by convolution expression (2.7), we see that there is no bandwidth parameter "b_n" to estimate in the convolution smoothing but instead we have a scale parameter "σ" to deal with it. b_n and σ, both do the same job, work as a smoothing parameter but estimation of b_n and σ are indeed two different problems. As we discuss in section 2.3, even though there are promising studies that have published on selection of bandwidth parameter, there is no common solution to the bandwidth parameter b_n in the literature. On the other hand, σ parameter can be estimated with classical scale estimation methods. This makes the "smoothing by convolution" procedure somewhat simpler than regular kernel density estimation procedure.

The convolution smoothing replaces the value of a function by a local weighted average of the function's values according to a weight function g(t),which will be symmetric around zero. In general, the shape of the resulting convolution function will depend on sample size n and σ. On the other hand, Kernel estimator uses a single shape for all sample sizes and width of the kernel control by smoothing parameter b_n (see, Scott 1992). The following example and plots will illustrate the convolution smoothing and the effect of σ parameter over the shape of the data.

Example 2.4.1. (*Illustration of the Convolution*) To illustrate how the smoothing by convolution works, we will generate a random sample of X from Exp(1) with $n_1 = 50$ and a random sample of Y-Δ from the same underlying distribution function Exp(1) with $n_2 = 30$. We assume that under $H_0 : \Delta_0 = 0$. Then, we will convolute the combined empirical CDF of $F_n^\star(x)$ with a smoother density function such as

$$g(t,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t)^2}{2\sigma^2}}$$

The parameter σ^2 is an important one here, we can call it the smoothing parameter, smoothness of the convolution function $\widehat{h}_\Delta(t)$ depends on the smoothing parameter σ^2. The following figures show the importance of σ^2 as a smoothing parameter in the Gaussian density function g(t). First column is the smoothed empirical density function $\widehat{h}_\Delta(t)$ and second column shows the resulting negative log likelihood function $L_n(\Delta)$. In first row, $\sigma = 0.1$ is arbitrarily chosen value of smoothing parameter, the resulting convolution does not work and the graph is rough, no smoothing is done. In the second row, σ is increased to 1 The resulting graphs are both smooth and $\widehat{h}_\Delta(t)$ resembles the true density function Exp(1). In the third row, σ is chosen to be 10, resulting $\widehat{h}_\Delta(t)$ looks over smoothed and loses the original exponential density shape. It looks very similar to the smoother density function g(t). In this example, g(t) is a normal density function and over smoothed shape is similar to the normal density function.

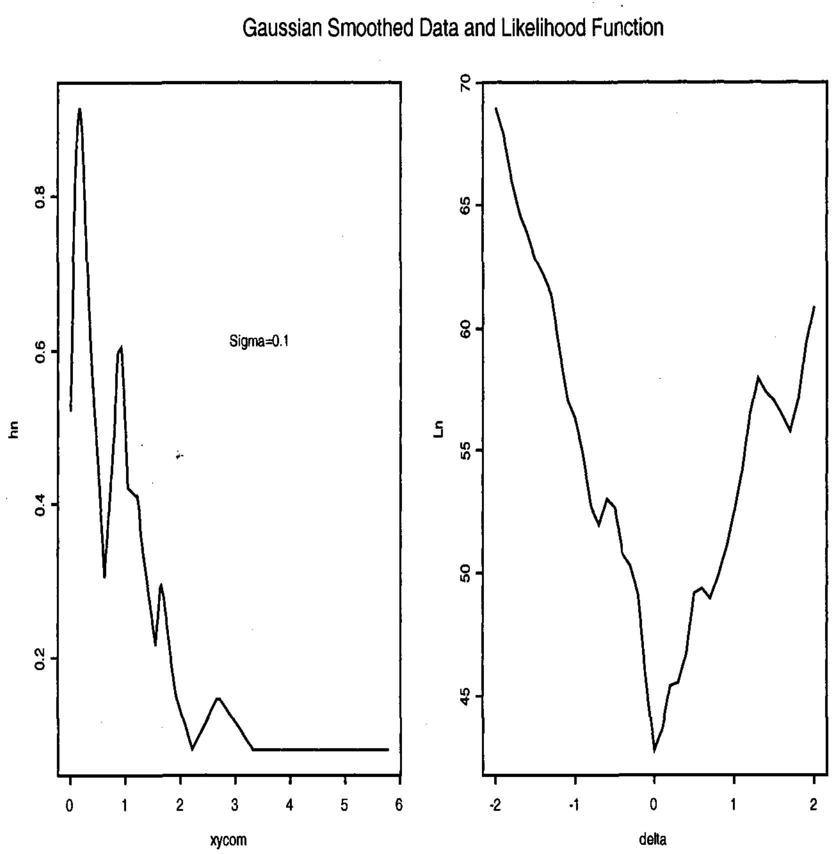

Figure 2: $\widehat{h}_{\Delta}(t)$ and $L_n(\Delta)$ with $\sigma = 0.1$

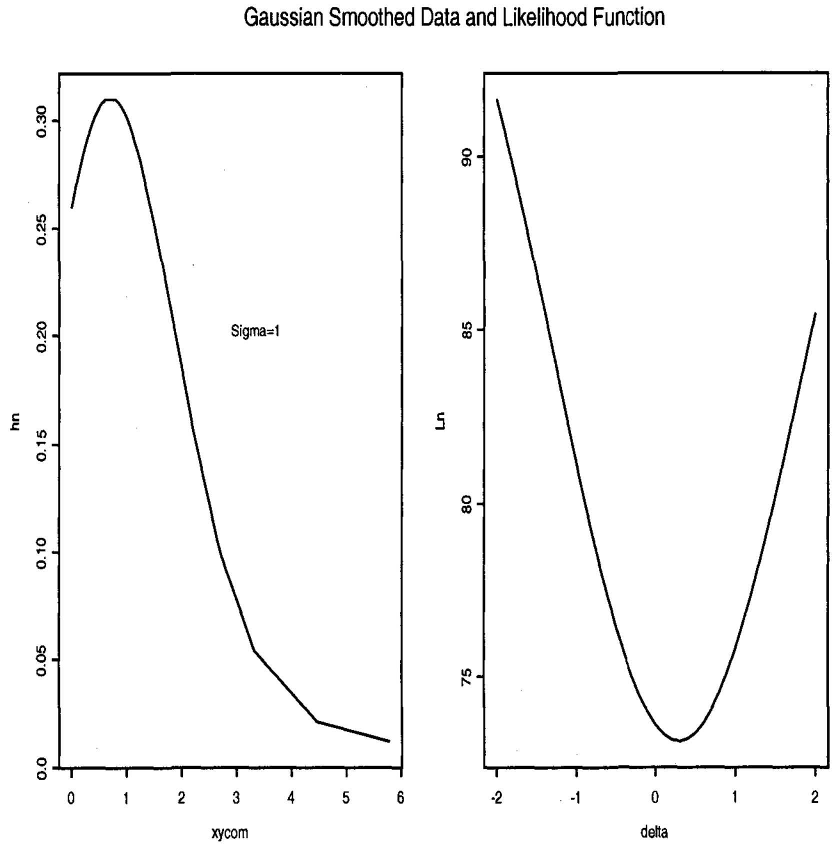

Figure 3: $\widehat{h}_{\Delta}(t)$ and $L_n(\Delta)$ with $\sigma = 1$

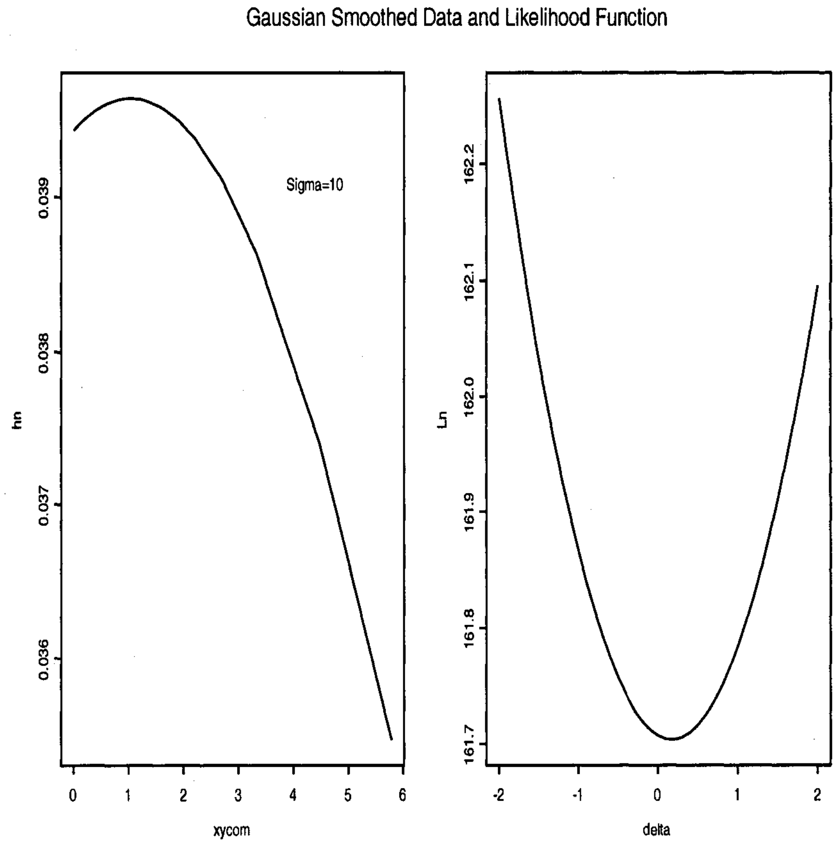

Figure 4: $\widehat{h}_{\Delta}(t)$ and $L_n(\Delta)$ with $\sigma = 10$

As we see from the example given figures above, $\widehat{h}_\Delta(t)$ is not an exact estimate of f(x) but it has the basic shape of the unknown density function f(x), in this case, an exponential density function. This shape depends on the scale parameter σ and it should be chosen carefully in order to have appropriate smoothing of the data. Even though we mentioned kernel density estimation method here, we are not going to deal with the main topics of kernel density estimation procedure such as consistency, convergence rates, selecting bandwidth parameter etc. because it is not our interest to investigate those properties in this dissertation. The purpose of this dissertation is to find a "replacement" for unknown density function f(x) which makes the log likelihood function smooth enough with respect to Δ so that we can estimate the shift parameter from the likelihood function. In the Chapter 7.2, a simulation study will be conducted to investigate the differences between using b_n and σ as a smoothing parameter.

2.5 Estimation Procedure

From the previous section 2.4, we have the smoothed empirical density function $\widehat{h}_\Delta(t)$ of the combined shifted sample. The unknown true density f(x) in the log likelihood function (2.2) can be replaced with the $\widehat{h}_\Delta(t)$. First recall that

$$\widehat{h}_\Delta(t) = \frac{1}{n}\left[\sum_{i=1}^{n_1} g(t - x_i) + \sum_{j=1}^{n_2} g(t - y_j + \Delta)\right] \tag{2.8}$$

If we rewrite $\widehat{h}_\Delta(t)$ in two pieces, we have

$$\widehat{h}_\Delta(t) = \widehat{h}_{\mathbf{x}}(t) + \widehat{h}_{\mathbf{y}}(t+\Delta) \tag{2.9}$$

where $\widehat{h}_{\mathbf{x}}(t) = \frac{1}{n}\sum_{l=1}^{n_1} g(t-x_l)$ and $\widehat{h}_{\mathbf{y}}(t) = \frac{1}{n}\sum_{k=1}^{n_2} g(t-y_k)$.

If we replace f(x) with $\widehat{h}_\Delta(t)$, we have the new following log likelihood function,

$$\begin{aligned} L_n(\Delta) &= -\sum_{i=1}^{n_1} log\left[\widehat{h}_\Delta(x_i)\right] - \sum_{j=1}^{n_2} log\left[\widehat{h}_\Delta(y_j-\Delta)\right] \\ &= -\sum_{i=1}^{n_1} log\left[\widehat{h}_{\mathbf{x}}(x_i) + \widehat{h}_{\mathbf{y}}(x_i+\Delta)\right] - \sum_{j=1}^{n_2} log\left[\widehat{h}_{\mathbf{x}}(y_j-\Delta) + \widehat{h}_{\mathbf{y}}(y_j-\Delta+\Delta)\right] \\ L_n(\Delta) &= -\sum_{i=1}^{n_1} log\left[\widehat{h}_{\mathbf{x}}(x_i) + \widehat{h}_{\mathbf{y}}(x_i+\Delta)\right] - \sum_{j=1}^{n_2} log\left[\widehat{h}_{\mathbf{x}}(y_j-\Delta) + \widehat{h}_{\mathbf{y}}(y_j)\right] \end{aligned} \tag{2.10}$$

The minimization of the negative log likelihood function would give us the following expression

$$\widehat{\Delta}_S = \underset{\Delta}{Argmin}\{L_n(\Delta)\} \tag{2.11}$$

where $\widehat{\Delta}_S$ is the proposed estimate of the true shift parameter Δ_0.

By taking the derivative of the $L_n(\Delta)$ with respect to Δ, we can find the gradient function and solve for Δ. So the partial derivative of the first sum term in $L_n(\Delta)$ with respect to Δ is

$$\frac{\partial L_n^1}{\partial \Delta} = -\sum_{i=1}^{n_1} \frac{\hat{h}'_{\mathbf{y}}(x_i+\Delta)}{\widehat{h}_\Delta(x_i)} \tag{2.12}$$

Similarly, the partial derivative of the second sum term is

$$\frac{\partial L_n^2}{\partial \Delta} = -\sum_{j=1}^{n_2} (-1) \frac{\hat{h}'_{\mathbf{x}}(y_j-\Delta)}{\widehat{h}_\Delta(y_j-\Delta)} \tag{2.13}$$

By adding the two derivatives, we get

$$S_n(\Delta) = \frac{\partial L_n(\Delta)}{\partial \Delta} = -\sum_{i=1}^{n_1} \frac{\hat{h}'_{\mathbf{y}}(x_i+\Delta)}{\widehat{h}_\Delta(x_i)} + \sum_{j=1}^{n_2} \frac{\hat{h}'_{\mathbf{x}}(y_j-\Delta)}{\widehat{h}_\Delta(y_j-\Delta)}$$

$$S_n(\Delta) = -\sum_{i=1}^{n_1} \widehat{\psi}_1(x_i,\Delta) + \sum_{j=1}^{n_2} \widehat{\psi}_2(y_j,\Delta) \tag{2.14}$$

where

$$\widehat{\psi}_1(x,\Delta) = \frac{\hat{h}'_{\mathbf{y}}(x+\Delta)}{\widehat{h}_\Delta(x)} = \frac{\frac{1}{n}\sum_{k=1}^{n_2} g'(x-y_k+\Delta)}{\widehat{h}_\Delta(x)} \tag{2.15}$$

and

$$\widehat{\psi}_2(y,\Delta) = \frac{\hat{h}'_{\mathbf{x}}(y-\Delta)}{\widehat{h}_\Delta(y-\Delta)} = \frac{\frac{1}{n}\sum_{l=1}^{n_1} g'(y-\Delta-x_l)}{\widehat{h}_\Delta(y-\Delta)} \tag{2.16}$$

Therefore, the shift parameter can be estimated by either

$$\widehat{\Delta}_S = \underset{\Delta}{Argmin}\{L_n(\Delta)\}$$

or solving

$$S_n(\Delta) = \frac{\partial L_n(\Delta)}{\partial \Delta} = 0$$

A minimization algorithm or root finder algorithm can find $\widehat{\Delta}_S$. Details of these results will appear in Chapter III.

2.6 Translation Property

One of the important properties of the proposed estimator $\widehat{\Delta}_S$ is the translation property. This property will allow us to assume under $H_0 : \Delta_0 = 0$ without loss of generality. For convenience, let the vector $\mathbf{X}$ denote sample $X_1, ..X_{n_1}$ and the vector $\mathbf{Y}$ denote the sample $Y_1, .., Y_{n_2}$. The following definition of translation equvariance property is given by Hetmansperger and McKean(1998).

Definition 2.6.1. Let $\mathbf{X} + a\mathbf{1} = (X_1 + a, .., X_{n_1} + a)$, where a is a constant number. An estimator $\widehat{\Delta}(\mathbf{X}, \mathbf{Y})$ of Δ is said to be a location equivariant estimator of Δ if $\widehat{\Delta}(\mathbf{X} + a\mathbf{1}, \mathbf{Y}) = \widehat{\Delta}(\mathbf{X}, \mathbf{Y}) - a$ and $\widehat{\Delta}(\mathbf{X}, \mathbf{Y} + a\mathbf{1}) = \widehat{\Delta}(\mathbf{X}, \mathbf{Y}) + a$

Theorem 2.6.1. *Let* $\widehat{\Delta}_S = \underset{\Delta}{Argmin}\{L_n(\Delta|\mathbf{X}, \mathbf{Y}))\}$. *Let* $\boldsymbol{X} + a\mathbf{1} = (X_1 + a, .., X_{n_1} + a)$. *Then we have*

$$\widehat{\Delta}_a = \underset{\Delta}{Argmin}\{L_n(\Delta|\mathbf{X} + a\mathbf{1}, \mathbf{Y})\}$$

where

$$\widehat{\Delta}_a = \widehat{\Delta}_S - a$$

Proof. By the expression (2.11), we have $\widehat{\Delta}_S = \underset{\Delta}{Argmin}\{L_n(\Delta)\}$ where

$$\begin{aligned} L_n(\Delta) &= -\sum_{i=1}^{n_1} log\left[\widehat{h}_\Delta(x_i)\right] - \sum_{j=1}^{n_2} log\left[\widehat{h}_\Delta(y_j - \Delta)\right] \\ &= -\sum_{i=1}^{n_1} log\left[\widehat{h}_{\mathbf{x}}(x_i) + \widehat{h}_{\mathbf{y}}(x_i + \Delta)\right] - \sum_{j=1}^{n_2} log\left[\widehat{h}_{\mathbf{x}}(y_j - \Delta) + \widehat{h}_{\mathbf{y}}(y_j)\right] \end{aligned}$$

Note that the $L_n(\Delta)$ can be also written as $L_n(\Delta|\mathbf{X},\mathbf{Y})$ since it depends on $\mathbf{X}$ and $\mathbf{Y}$ also. Let $\mathbf{X} + a\mathbf{1} = X_1 + a, .., X_{n_2} + a$. Then, replace sample $\mathbf{X}$ with $\mathbf{X} + a\mathbf{1}$ in the $L_n(\Delta|\mathbf{X},\mathbf{Y})$. So

$$\begin{aligned} L_n(\Delta|\mathbf{X} + a\mathbf{1}, \mathbf{Y}) &= -\sum_{i=1}^{n_1} log\left[\widehat{h}_\Delta(x_i + a)\right] - \sum_{j=1}^{n_2} log\left[\widehat{h}_\Delta(y_j - \Delta)\right] \\ &= -\sum_{i=1}^{n_1} log\left[\widehat{h}_{\mathbf{x}+a}(x_i + a) + \widehat{h}_{\mathbf{y}}(x_i + a + \Delta)\right] \\ &\quad - \sum_{j=1}^{n_2} log\left[\widehat{h}_{\mathbf{x}+a}(y_j - \Delta) + \widehat{h}_{\mathbf{y}}(y_j - \Delta)\right] \\ &= -\sum_{i=1}^{n_1} log\left[\frac{1}{n}\sum_{l=1}^{n_1} g(x_i + a - x_l - a) + \frac{1}{n}\sum_{k=1}^{n_2} g(x_i + a + \Delta - y_k)\right] \\ &\quad - \sum_{j=1}^{n_2} log\left[\frac{1}{n}\sum_{l=1}^{n_1} g(y_j - \Delta - x_l - a) + \frac{1}{n}\sum_{k=1}^{n_2} g(y_j - \Delta - y_k + \Delta)\right] \end{aligned}$$

After some rearrangement inside the sums and writing $-\Delta - a = -(\Delta + a)$, we get the following expression

$$= -\sum_{i=1}^{n_1} log \left[\frac{1}{n} \sum_{l=1}^{n_1} g(x_i - x_l) + \frac{1}{n} \sum_{k=1}^{n_2} g(x_i - y_k + (\Delta + a)) \right]$$
$$- \sum_{j=1}^{n_2} log \left[\frac{1}{n} \sum_{l=1}^{n_1} g(y_j - (\Delta + a) - x_l) + \frac{1}{n} \sum_{k=1}^{n_2} g(y_j - y_k) \right] \quad (2.17)$$

Inside the brackets, the shift variable is $\Delta + a$ this time and the log likelihood function depends on $\Delta + a$. Thus we can write

$$L_n(\Delta|\mathbf{X} + a\mathbf{1}, \mathbf{Y}) = -\sum_{i=1}^{n_1} log \left[\widehat{h}_{\mathbf{x}}(x_i) + \widehat{h}_{\mathbf{y}}(x_i + (\Delta + a)) \right]$$
$$- \sum_{j=1}^{n_2} log \left[\widehat{h}_{\mathbf{x}}(y_j - (\Delta + a)) + \widehat{h}_{\mathbf{y}}(y_j) \right]$$
$$L_n(\Delta|\mathbf{X} + a\mathbf{1}, \mathbf{Y}) = L_n(\Delta + a|\mathbf{X}, \mathbf{Y}) = L_n(\Delta + a) \quad (2.18)$$

Let's define

$$\widehat{\Delta}_a = \underset{\Delta}{Argmin}\{L_n(\Delta + a)\} \quad (2.19)$$

where $\widehat{\Delta}_a$ is the argument that minimizes $L_n(\Delta + a)$. Note that $\widehat{\Delta}_S$ is the argmin of the $L_n(\Delta)$, which implies that $L_n(\widehat{\Delta}_S)$ is the minimum at $\widehat{\Delta}_S$. Then the argument that minimizes $L_n(\Delta + a)$ must be equal to $\widehat{\Delta}_S - a$. Thus, we have the following

$$\widehat{\Delta}_a = \widehat{\Delta}_S - a \quad (2.20)$$

Similarly, by using the same argument above for $Y_1 + a, .., Y_{n_2} + a$, we will have

$$\widehat{\Delta}_a = \widehat{\Delta}_S + a \tag{2.21}$$

Therefore, the proposed estimator $\widehat{\Delta}_S$ is a translation equivariant estimator. $\square$

CHAPTER III

PROPERTIES OF $\widehat{h}_\Delta(t)$ and $L_n(\Delta)$

3.1 Asymptotic Properties of $\widehat{h}_\Delta(t)$

In this section, we are shall give some important definitions and theorems, which are necessary for establishing the basic properties of the $\widehat{h}_\Delta(t)$. We investigate the asymptotic properties of the smoothed empirical density $\widehat{h}_\Delta(t)$ in this chapter. It will be proven that $\widehat{h}_\Delta(t)$ converges to a true convolution density $h_\Delta(t)$, which is the convolution of the smoother density function g(t) with CDF of $F^\star(x)$. The pointwise convergence of $\widehat{h}_\Delta(t)$ to $h_\Delta(t)$ is clear by the Strong Law of Large Numbers(S.L.L.N). The theorems of Wellner(2001) will be introduced in order to prove the uniform convergence of $\widehat{h}_\Delta(t)$ to $h_\Delta(t)$. First, we shall state a very basic theorem of empirical processes called the Glivenko-Cantelli theorem in the literature. We would like to state this theorem because we are using the empirical distribution function $F_n^\star(x)$ in our convolution transformation. The following well known theorem proves that $F_n^\star(x)$ converges to $F^\star(x)$ uniformly for i.i.d random variables. We shall state the one-sample case of the Glivenko-Cantelli Theorem first.

Theorem 3.1.1. *(Glivenko-Cantelli Theorem)*

Assume that $Z_1,..Z_n$ *are i.i.d with distribution function F(z). Then,*

$$\sup_{z\in\mathbf{R}}|F_n(z) - F(z)| \longrightarrow_{a.s} 0 \text{ as } n \to \infty$$

We can extend the one-sample Glivenko-Cantelli theorem to introduce two sample version of it. The following theorem states the result for the two-sample problem.

Theorem 3.1.2. *Assume that*

1. *The random variables* $X_1, ..., X_{n_1}$ *and* $Y_1, ..., Y_{n_2}$ *are i.i.d with distribution functions F(x) and G(y)=F(y-*Δ_0*) respectively.*

2. $n = n_1 + n_2$, $n_i/n \to \lambda_i$ *and* $0 < \lambda_i < 1$ *for i=1,2 as* $n \to \infty$.

Then,

$$\sup_{x\in\mathbf{R}}|F_n^\star(x) - F^\star(x)| \longrightarrow_{a.s} 0 \text{ as } n \to \infty$$

where

$$F_n^\star(x) = \tfrac{n_1}{n} F_{n_1}(x) + \tfrac{n_2}{n} G_{n_2}(x + \Delta)$$

and

$$\begin{aligned} F^\star(x) &= \lambda_1 F(x) + \lambda_2 G(x + \Delta) \\ &= \lambda_1 F(x) + \lambda_2 F(x + \Delta - \Delta_0) \end{aligned}$$

Corollary 3.1.1. *If we assume that the arbitrary shift variable Δ equals to true shift parameter Δ_0 (or $H_0 : \Delta_0 = 0$ and $\Delta = 0$), then*

$$\sup_{x \in \mathbf{R}} |F_n^\star(x) - F(x)| \longrightarrow_{a.s} 0 \text{ as } n \rightarrow \infty$$

Proof. If $\Delta = \Delta_0$, $F^\star(x) = \lambda_1 F(x) + \lambda_2 F(x)$. Since $\lambda_1 + \lambda_2 = 1$, $F^\star(x) = F(x)$. □

The following theorem and its proof are provided by Wellner (2001) and it is known as Glivenko-Contelli theorems in the empirical processes theory.

Theorem 3.1.3. *{Theorem 2 of Wellner 2001} Assume that*

1. *F_n is an empirical CDF of sample size n from the density f*
2. *Θ is compact set.*
3. *g(z,t)is a upper semicontinuous in t for all z.*
4. *There exists a function H(z) such that E[H(Z)] $< \infty$ and $g(z,t) \leq$H(z) for all $z \in \chi$ and $t \in \Theta$.*
5. *For all t and sufficiently small $\rho > 0$*

$$\sup_{|t'-t|<\rho} f(x,t')$$

is measurable in z. Then,

$$\limsup_{n \to \infty} \sup_{t \in \Theta} F_n g(z,t) \leq \sup_{t \in \Theta} F g(z,t) \tag{3.1}$$

The following theorem and its proof is provided by Wellner(2001). We will state the theorem and the proof for the readers.

Theorem 3.1.4. *{Theorem 1 of Wellner 2001} Assume that*

1. *$Z_1, Z_2, ...Z_n$ are i.i.d with cdf F on the measurable space(χ,A).*

2. *Θ is compact set.*

3. *$g(z,t)$is a measurable, real valued function of z and t and continuous in t for all z.*

4. *There exists a function $H(z)$ such that $E[H(Z)] < \infty$ and $|g(z,t)| \leq H(z)$ for all $z \in \chi$ and $t \in \Theta$. Then,*

$$\sup_{t\in\Theta}\left|\frac{1}{n}\sum_{k=1}^{n} g(Z_k,t) - \int g(z,t)dF(z)\right| \longrightarrow_{a.s} 0 \text{ as } n \to \infty. \tag{3.2}$$

for all fixed t.

Proof. The proof is from the Theorem 1 of Wellner(2001) and it also uses of the result of Theorem 2 of the same author.

First note that $E[g(Z,t)] = \int g(z,t)dF(z)$ exists and is finite$E[g(z,t)] < \infty$ for all $t \in \Theta$. By the Strong Law of Large Numbers(S.L.L.N), we have

$$\int g(z,t)dF_n(z) = \frac{1}{n}\sum_{k=1}^{n} g(Z_k,t) \longrightarrow_{a.s} \int g(z,t)dF(z) = E[g(Z,t)]$$

for all fixed t. We also want to show that this result holds uniformly in $t \in \Theta$ so that we have (3.2). First, define $k(z,t) \equiv g(z,t) - \int g(z,t)dF(z)$ and $-k(z,t)$. By the Theorem 2 of Wellner(2001) applied to k(z,t), we have,

$$\limsup_{n\to\infty} \sup_{t\in\Theta} F_n k(z,t) \leq \sup_{t\in\Theta} Fk(z,t)$$
$$\limsup_{n\to\infty} \sup_{t\in\Theta} \left(F_n g(z,t) - \int g(z,t)dF(z) \right) \leq \sup_{t\in\Theta} \left(Fg(z,t) - \int g(z,t)dF(z) \right)$$
$$\limsup_{n\to\infty} \sup_{t\in\Theta} \left(\frac{1}{n} \sum_{k=1}^{n} g(z_k,t) - \int g(z,t)dF(z) \right) \leq 0 \ a.s$$

Similarly, by the Theorem 2 of Wellner(2001) applied to -k(z,t), we have,

$$\limsup_{n\to\infty} \sup_{t\in\Theta} \left(\int g(z,t)dF(z) - \frac{1}{n} \sum_{k=1}^{n} g(z_k,t) \right) \leq 0 \ a.s$$

We can conclude our proof since

$$0 \leq \sup_{t\in\Theta} \left| \frac{1}{n} \sum_{k=1}^{n} g(z_k,t) - \int g(z,t)dF(z) \right|$$
$$= \sup_{t\in\Theta} \left(\frac{1}{n} \sum_{k=1}^{n} g(z_k,t) - \int g(z,t)dF(z) \right) \vee \sup_{t\in\Theta} \left(\int g(z,t)dF(z) - \frac{1}{n} \sum_{k=1}^{n} g(z_k,t) \right)$$

By taking the $\limsup_{n\to\infty}$ of both sides, gives us the desired result. Thus,

$$\sup_{t\in\Theta} \left| \frac{1}{n} \sum_{k=1}^{n} g(z_k,t) - \int g(z,t)dF(z) \right| \longrightarrow_{a.s} 0 \text{ as } n \to \infty$$

□

The following theorem can be concluded as a result of the Theorem 3.1.4 and Theorem 3.1.3 that are stated above.

Theorem 3.1.5. *Assume that*

1. $X_1, ..X_{n_1}$ *and* $Y_1, .., Y_{n_2}$ *are i.i.d with distribution functions* $F(x)$ *and* $G(y + \Delta) = F(y + \Delta - \Delta_0)$ *respectively.*

2. t *belongs to any fixed interval* $t \in [A, B]$.

3. $g(t)$ *is a bounded continuous function of* t.

4. *There exists a function* $H(z)$*such that* $E[H(Z)] < \infty$ *and* $|g(t)| \leq H(z)$ *for all* $t \in [A, B]$.

Then,

$$\sup_{t \in \Theta} |\widehat{h}_\Delta(t) - h_\Delta(t)| \rightarrow_{a.s} 0 \text{ as } n \rightarrow \infty \text{ for all fixed } \Delta.$$

where,

$$\widehat{h}_\Delta(t) = \frac{1}{n}\left[\sum_{i=1}^{n_1} g(t - x_i) + \sum_{j=1}^{n_2} g(t - y_j + \Delta)\right]$$

and

$$\begin{aligned} h_\Delta(t) &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta)dG(y) \\ &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta-\Delta_0)dF(y) \end{aligned}$$

where $n_i/n \to \lambda_i$ for i=1,2 as $n \to \infty$.

Proof. By the Theorem 3.1.4, we have

$$\sup_{t\in\Theta}|\tfrac{1}{n}\sum_{i=1}^{n_1} g(t-x_i) - \lambda_1 \int g(t-x)dF(x)| \to_{a.s} 0 \text{ as } n \to \infty.$$

Similarly,

$$\sup_{t\in\Theta}|\tfrac{1}{n}\sum_{j=1}^{n_2} g(t-y_j+\Delta) - \lambda_2 \int g(t-y+\Delta)dG(y)| \to_{a.s} 0 \text{ as } n \to \infty.$$

By the Slutsky's Theorem,

$$\sup_{t\in\Theta}|\left[\tfrac{1}{n}\sum_{i=1}^{n_1} g(t-x_i) + \tfrac{1}{n}\sum_{j=1}^{n_2} g(t-y_j+\Delta)\right] -$$
$$\left[\lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta)dG(y)\right]| \to_{a.s} 0 \text{ as } n \to \infty.$$

Thus, the theorem is proved. □

It will be useful to remember the definitions of the log concavity and concavity before we start discussing local properties of $\widehat{h}_\Delta(t)$. The following definitions can be found in any calculus book. We just feel the need to state them here in order to use in our proofs and discussions in this section.

Definition 3.1.1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is concave if for all $x, y \in dom f$ and $0 \leq \theta \leq 1$, we have

$$f(\theta x + (1-\theta)y) \geq \theta f(x) + (1-\theta)f(y)$$

and we also say that -f is convex if f is concave. If we assume that f is twice differentiable then the second order condition implies that f is concave if $f''(x) \leq 0$.

Definition 3.1.2. A function f(x) is log concave if and only if log f(x) is concave for $f(x) > 0$.

Remark 3.1.1. The log concavity of a function f(x) implies that

$$f''(x)f(x) \leq [f'(x)]^2$$

It should be worth to mention that concavity implies log-concavity. Thus, a concave function is also log concave function but opposite is not true. On the other hand, sum of the concave functions are concave but sum of the log concave functions is not necessarily log concave.

We now look at the smoothed empirical density function $\widehat{h}_\Delta(t)$ which is given by expression (2.8). If we assume that the smoother density function g(t) is a log concave density, like normal density, the sum of the smoother density function g(t) is not necessarily log concave as stated in the remark above. But under some conditions, $\widehat{h}_\Delta(t)$ can be log concave function of Δ, then -log likelihood function $L_n(\Delta)$ will be a convex function of Δ under the same conditions applied to $\widehat{h}_\Delta(t)$. In the following example these conditions will be investigated.

Example 3.1.1. We want to give an example that illustrates the log concavity of $\widehat{h}_\Delta(t)$ under some conditions. Since concavity implies log concavity, the conditions that makes $\widehat{h}_\Delta(t)$ concave will be the same as the log concavity conditions. This way we will avoid working with log function. Let us assume that smoother density function is

$$g(t) = \frac{1}{\sqrt{2\pi}} \exp\{\frac{-t^2}{2}\} \tag{3.3}$$

which is the standard normal density function. We should note that univariate and multivariate normal density functions are log concave functions. Recall that

$$\widehat{h}_\Delta(t) = \frac{1}{\sigma \cdot n}\left[\sum_{l=1}^{n_1} g(\frac{t-x_l}{\sigma}) + \sum_{k=1}^{n_2} g(\frac{t-y_k+\Delta}{\sigma})\right] \tag{3.4}$$

We first replace g(t) function with standard normal density and take the first and the second derivative of $\widehat{h}_\Delta(t)$ with respect to Δ. So,

$$\frac{\partial \widehat{h}_\Delta(t)}{\partial \Delta} = \frac{1}{\sigma n}\left[\sum_{k=1}^{n_2}(-1)\frac{(t-y_k+\Delta)}{\sqrt{2\pi}\sigma^2}\exp\{\frac{-(t-y_k+\Delta)^2}{2\sigma^2}\}\right] \tag{3.5}$$

The second derivative is

$$\frac{\partial^2 \widehat{h}_\Delta(t)}{\partial^2 \Delta} = \frac{1}{\sigma^3 n}\left[\sum_{k=1}^{n_2}\left(\frac{(t-y_k+\Delta)^2}{\sigma^2} - 1\right)\frac{1}{\sqrt{2\pi}}\exp\{\frac{-(t-y_k+\Delta)^2}{2\sigma^2}\}\right]$$

Let $t = x_i$ for $i = 1, .., n_1$. Then,

$$\frac{\partial^2 \widehat{h}_\Delta(x_i)}{\partial^2 \Delta} = \frac{1}{\sigma^3 n}\left[\sum_{k=1}^{n_2}\left(\frac{(x_i-y_k+\Delta)^2}{\sigma^2} - 1\right)\frac{1}{\sqrt{2\pi}}\exp\{\frac{-(x_i-y_k+\Delta)^2}{2\sigma^2}\}\right]$$

The second product factor on the right is always positive. Therefore, in order to have concavity which implies $\frac{\partial^2 \widehat{h}_\Delta(x_i)}{\partial \Delta} \leq 0$, the first product factor inside the sum must be negative. Thus, it can be written that

$$\left[\frac{(x_i-y_k+\Delta)^2}{\sigma^2} - 1\right] < 0 \tag{3.6}$$

This implies that

$$\frac{(x_i - y_k + \Delta)^2}{\sigma^2} < 1$$

$$(x_i - y_k + \Delta)^2 < \sigma^2$$

$$|x_i - y_k + \Delta| < \sigma$$

$$-\sigma < x_i - y_k + \Delta < \sigma \tag{3.7}$$

By the last expression, we can find a σ for some x_i, y_k and for some Δ on some interval [a,b] that makes $\widehat{h}''_{\Delta}(x_i) \leq 0$.

Similarly, let $t = y_j$ for $j = 1, .., n_2$, $\widehat{h}_{\Delta}(y_j)$ can also be a concave function of Δ on some interval [a,b] for a proper σ that satisfies $\frac{\partial^2 \widehat{h}_{\Delta}(y_j)}{\partial \Delta} \leq 0$. Since concavity implies log concavity, the same condition is also valid to make $\widehat{h}_{\Delta}(y_j)$ log concave function of Δ.

We now state a theorem that can be found in Miravete (2002). The proof of the Theorem is provided by the same author.

Theorem 3.1.6. *Assume that*

1. *the smoother density function g(t) is a log concave density*

2. *the true density function f(x) is log concave density*

Then the convolution of g(t) and f(x)

$$h(t) = \int g(t-x)dF(x) = \int g(t-x)f(x)dx \tag{3.8}$$

is a log concave function of t.

Proof. Miravate(2002) proved that the convolution of the two log concave density functions is also a log concave density. □

Remark 3.1.2. If we recall the expression(2.6), we can write the smoothed empirical density $\widehat{h}_\Delta(t)$ in the following form

$$\widehat{h}_\Delta(t) = \frac{1}{\sigma \cdot n}\left[\sum_{l=1}^{n_1} g(\frac{t - x_l}{\sigma}) + \sum_{k=1}^{n_2} g(\frac{t - y_k + \Delta}{\sigma})\right] \tag{3.9}$$

For fix n, we can argue that $\widehat{h}_\Delta(t)$ is a log concave function as $\sigma \to \infty$ as long as g(t) is log concave. This is consistent with the figures 2- 4. For large σ, we will over smooth the data and lose the shape of the underlying density and resulting convolution density will look similar to smoother density function g(t). If we have a log concave smoother density function g(t), $\widehat{h}_\Delta(t)$ will be log concave for $\sigma \to \infty$. In a special case if smoother density g(t) is normal density and if $\sigma \to \infty$, the resulting estimate of Δ_0 will be same as maximum likelihood estimate. Also, the resulting negative log likelihood function $L_n(\Delta)$ will be a convex function of Δ. It is not desirable to have a large σ that can spoil the estimation or lose robustness but a reasonable σ can give us a robust estimate of the true shift parameter Δ_0. We will discuss this matter later in the simulation and examples.

The convexity of $L_n(\Delta)$ is hard to reach in practice. In a practical point of view, for finite samples, we can accomplish the log concavity of $\widehat{h}_\Delta(t)$ on some interval that $\Delta \in [a, b]$ for a reasonable smoothing parameter σ on some data points.(i.e, See Example 3.1.1). The larger σ, the wider the interval $\Delta \in [a, b]$ will be on a given data set. This is true under the normal smoothing which has been shown by the Example 3.1.1 and expression(3.7). In the following section we will introduce a topic that it belongs to a family of convex functions.

3.2 Quasi-convexity of $L_n(\Delta)$

In this section, we would like to introduce one of the important properties of the $L_n(\Delta)$. That is Quasi-convexity. The quasi-convexity is discussed by several authors in the literature. Ponstein(1967) presented "Seven Kinds of Convexity" and introduced "strict" quasi-convexity in his article. Roberts and Varberg (1973) gave a nice table that compares the properties of convex and quasi-convex functions respectively. The main difference between convexity and quasi-convexity is that the convexity ensures the existence of the global minimum but the quasi-convexity ensures the existence of the local minimum. The local minimum can be a global minimum if the quasi-convexity is "strict". These properties will be discussed in this section. The graphical presentation of the quasi-convexity will be introduced with a figure that is given after the definition of quasi-convexity.

Definition 3.2.1. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasi-convex if its domain and all its sublevel sets

$$S_\alpha = \{x \in dom f | f(x) \leq \alpha\} \tag{3.10}$$

are convex.

The definition above is illustrated in figure 5. For every α, the sublevel set S_α is convex., i.e, S_α is an interval.

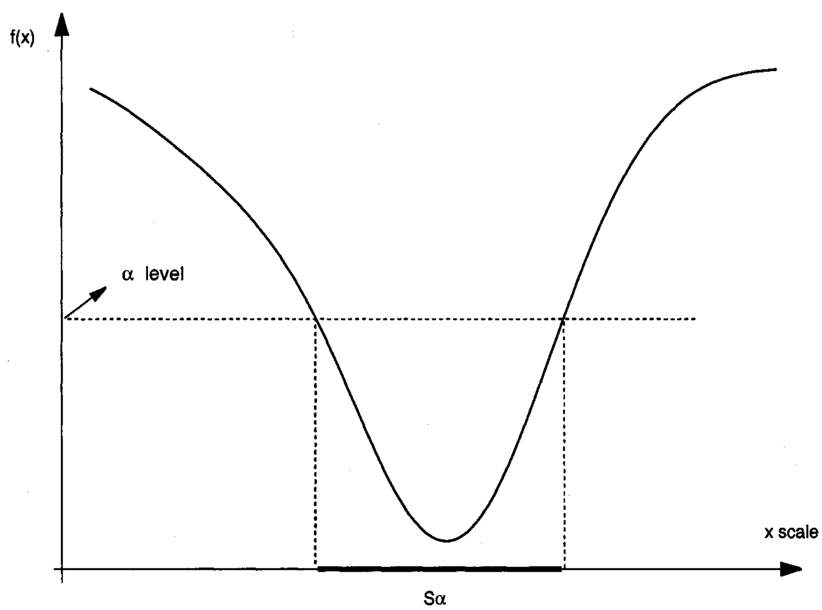

Figure 5: For each α, the α-sublevel set S_α is convex

Remark 3.2.1. The following property of the quasi-convexity is based on the Jensen's inequality that characterizes quasi-convexity of a function. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasi-convex if and only if for $x, y \in dom f$ and $0 \leq \theta \leq 1$

$$f(\theta x + (1-\theta)y) \leq max\{f(x), f(y)\} \tag{3.11}$$

Example 3.2.1. We would like to give an example to illustrate the quasi-convexity of $L_n(\Delta)$. We randomly generated two samples from Unif(1,2) and Unif(2,3) with sample sizes $n_1 = 30$ and $n_2 = 20$ respectively. So true shift parameter $\Delta_0 = 1$. The following figure shows the quasi-convexity of $L_n(\Delta)$ at $\sigma = 0.3$ level. The smoothing by convolution estimate of $\Delta_0 = 1$ is $\widehat{\Delta}_S = 1.05964$, Hodges-Lehman(H-L) estimate is $\widehat{\Delta}_{HL} = 1.05255$ and Least-Square(L-S) estimate is $\widehat{\Delta}_{LS} = 1.05913$.

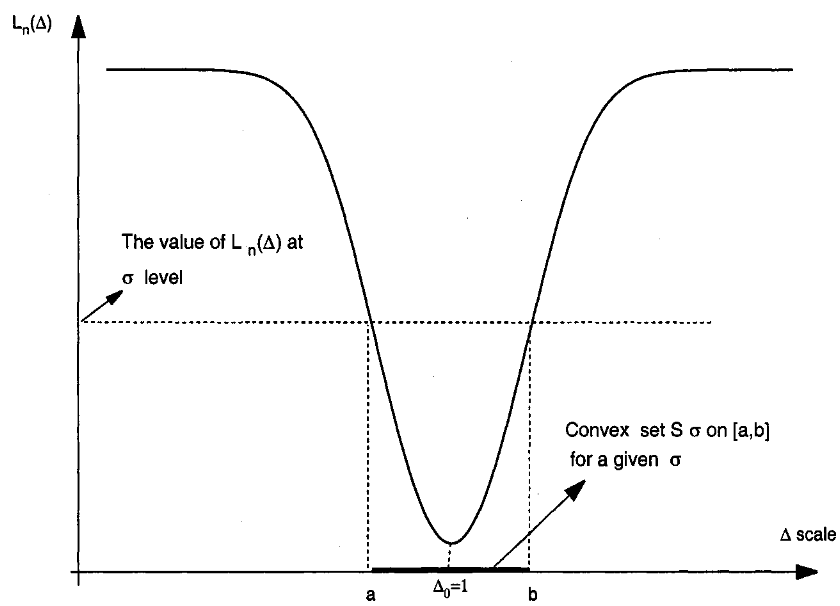

Figure 6: The quasiconvexity of $L_n(\Delta)$ with $\sigma = 0.3$ level

Remark 3.2.2. The effects of different σ levels and quasi-convexity of $L_n(\Delta)$ can be seen on the following figure. We use the same data set generated in the Example 3.2.1. Three levels of $\sigma = 0.3$, $\sigma = 1$ and $\sigma = 2$ are used in the plot. For each σ level, we have a convex subset S_σ that contains $\widehat{\Delta}_S$. S_σ is actually an interval that we defined as [a,b] in the previous examples.

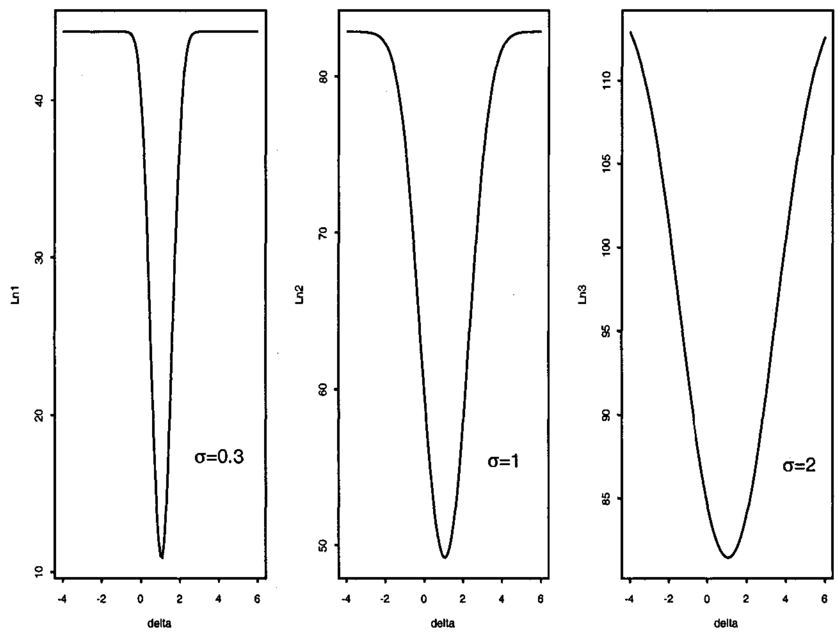

Figure 7: The quasiconvexity of $L_n(\Delta)$ with different σ levels

CHAPTER IV

ESTIMATION OF THE SHIFT PARAMETER BY USING $h_\Delta(t)$

4.1 The Likelihood function $L_n^\star(\Delta)$

In this section, we would like to introduce a new log likelihood function that is consist of $h_\Delta(t)$ instead of $\widehat{h}_\Delta(t)$. In Chapter 2, a smoother density function g(t) is convoluted with empirical distribution function $F_n^\star(x)$ of the combined sample Z_k. The resulting convolution is $\widehat{h}_\Delta(t)$. If we convolute a smoother density function g(t) with the distribution function $F^\star(x)$ of the combined sample Z_k, the resulting convolution is $h_\Delta(t)$. If we recall that by the Theorem 3.1.5, we have

$$\sup_{t\in\Theta}|\widehat{h}_\Delta(t) - h_\Delta(t)| \rightarrow_{a.s} 0 \text{ as } n \rightarrow \infty.$$

where,

$$\widehat{h}_\Delta(t) = \frac{1}{n}\left[\sum_{i=1}^{n_1} g(t - x_i) + \sum_{j=1}^{n_2} g(t - y_j + \Delta)\right]$$

and

$$\begin{aligned} h_\Delta(t) &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta)dG(y) \\ &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta-\Delta_0)dF(y) \end{aligned}$$

We define this new log likelihood function by replacing $\widehat{h}_\Delta(t)$ with $h_\Delta(t)$ in the $L_n(\Delta)$. After replacing $\widehat{h}_\Delta(t)$ with $h_\Delta(t)$, we have

$$L_n^\star(\Delta) = -\sum_{i=1}^{n_1} log[h_\Delta(x_i)] - \sum_{j=1}^{n_2} log[h_\Delta(y_j - \Delta)] \tag{4.1}$$

The purpose of this replacement is to create a new log likelihood function $L_n^\star(\Delta)$ in terms of true convolution density $h_\Delta(t)$ that can also be used to estimate the true shift parameter Δ_0. The new estimator is

$$\Delta_S^\star = \underset{\Delta}{Argmin}\,\{L_n^\star(\Delta)\}$$

We need to know underlying density function f(x) in order to estimate $\Delta_S^\star$. A question can be asked why do we need to use this estimator if we know underlying density function since the underlying density function can be used in the log likelihood function to estimate the shift parameter Δ_0. A simple answer maybe with smoothing of underlying density function efficiency of the estimator can be increased. This topic will be studied in the future because we are mainly interested in the estimator $\widehat{\Delta}_S$. We would like to investigate the log-concavity of $h_\Delta(t)$ with respect to Δ given that we have log-concave smoother density g(t) and log-concave true underlying density f(x). By the Theorem 3.1.6, it has been proven that the convolution of two log-concave density functions is a log-concave density. The result of log-concavity implies certain properties for a density function. Some important properties of log-concave functions are given in the following remark.

Remark 4.1.1. Log-concavity of a function $\psi(x)$ is equivalent to each of the following three conditions.

1. $log[\psi(x)]$ is a concave function.

2. $\psi'(x)/\psi(x)$ is monotone decreasing for all x.

3. $\psi(x)\psi''(x) - [\psi'(x)]^2 \leq 0$ for all x.

Theorem 4.1.1. *Let $A > 0$ be a constant number. Assume that ψ is twice differentiable log concave density function. Then, the function $\Phi(x) = log[A + \psi(x)]$ is a concave function on an interval (-a,a) if there is an interval (-a,a) such that*

$$\tfrac{\psi''(x)}{\psi(x)} \leq 0 \; for \; x \in (-a, a)$$

or

$$\psi''(x) \leq 0 \; for \; x \in (-a, a), \; where \; "a" \; is \; the \; inflection \; point \; of \; \psi(x).$$

Proof. By using the properties of log-concavity mentioned in the remark 4.1.1, we can find first and second derivation of function $\Phi(x)$

$$\Phi'(x) = \frac{\psi'(x)}{A + \psi(x)} \tag{4.2}$$

and

$$\Phi''(x) = \frac{[A + \psi(x)]\psi''(x) - [\psi'(x)]^2}{[A + \psi(x)]^2} \tag{4.3}$$

If we prove that $\Phi''(x) \leq 0$, then the proof is complete. The denominator of the $\Phi''(x)$ reveals that $[A + \psi(x)]^2$ is always positive. The numerator is

$$[A + \psi(x)]\psi''(x) - [\psi'(x)]^2 = A\psi''(x) + \psi(x)\psi''(x) - [\psi'(x)]^2$$

The last property in the remark 4.1.1 implies that $\psi(x)\psi''(x) - [\psi'(x)]^2 \leq 0$. Thus, by the assumption there is an interval (-a,a) such that $\psi''(x) \leq 0$ for $x \in (-a, a)$. Then, $A\psi''(x) \leq 0$ since A is a positive constant. Therefore, the numerator is

$$[A\psi''(x) + \psi(x)\psi''(x) - [\psi'(x)]^2 \leq 0 \qquad (4.4)$$

This result implies that $\Phi''(x) \leq 0$ for $x \in (-a, a)$. The point "a" in the interval (-a,a) can be called inflection point because $\psi(x)$ turns downward(concave down) between (-a,a). Therefore, the function $\Phi(x)$ will be concave on the same interval.

□

We would like to give an example to see that this condition holds for our application here.

Example 4.1.1. Let $\psi(x)$ be $N(0, \sigma^2)$ which is a log concave density. Let A be a positive constant number. First, we find the second derivative of $\psi(x)$. It is given by the following expression

$$\psi''(x) = \left[\frac{x^2}{\sigma^2} - 1\right] \left[\frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2\sigma^2}\}\right] \qquad (4.5)$$

and then we find the second derivative of $log[A + \psi(x)]$ which is

$$\frac{A\psi''(x) + \psi(x)\psi''(x) - [\psi'(x)]^2}{[A + \psi(x)]^2} \tag{4.6}$$

The first product term, $\left[\frac{x^2}{\sigma^2} - 1\right]$, in the $\psi''(x)$ implies that we must take $|x| \leq \sigma$ in order to have $\psi''(x) \leq 0$. Therefore, the numerator of the expression defined in the (4.6) will be less than or equal to zero as long as $x \in (-\sigma, \sigma)$(i.e $|x| \leq \sigma$). Thus, by Remark 4.1.1 and Theorem 4.1.1, $log[A + \psi(x)]$ is a concave on $(-\sigma, \sigma)$ and $[A + \psi(x)]$ is log concave on $(-\sigma, \sigma)$.

Corollary 4.1.1. *Suppose that*

1. *there is a positive constant A.*

2. *the smoother density function $g(t)$ is a log concave density function.*

3. *the true underlying density function f(x) is a log concave density function.*

4. *$\psi(t, \Delta) = \int g(t - y - \Delta_0 + \Delta)dF(y)$ is a log concave function of Δ for any given t.*

5. *there is an interval $(-t + \Delta_0, \Delta_0 + t)$ such that*

$$\psi''(t, \Delta) \leq 0 \text{ for } \Delta \in (-t + \Delta_0, \Delta_0 + t) \text{ with any given } t.$$

Then, by the Theorem 4.1.1,

$$\begin{aligned} h_\Delta(t) &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y-\Delta_0+\Delta)dF(y) \\ &= A + \psi(t,\Delta) \end{aligned}$$

is log concave function of Δ *on the interval* $(-t+\Delta_0, \Delta_0+t)$ *for any given t.*

Proof. We have

$$\begin{aligned} h_\Delta(t) &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta)dG(y) \\ &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta)dF(y-\Delta_0) \\ &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y-\Delta_0+\Delta)dF(y) \end{aligned} \tag{4.7}$$

The first integral, $\lambda_1 \int g(t-x)dF(x)$, is independent of Δ, therefore we can replace this integral with a positive constant A. So we have

$$h_\Delta(t) = A + \lambda_2 \int g(t-y-\Delta_0+\Delta)dF(y) \tag{4.8}$$

Let $\psi(t,\Delta) = \lambda_2 \int g(t-y-\Delta_0+\Delta)dF(y)$. Then,

$$h_\Delta(t) = A + \psi(t,\Delta) \tag{4.9}$$

By the Theorem 3.1.6, $\psi(t,\Delta)$ is a log concave function of Δ for all t since it is a convolution of two log concave density functions and log concavity is preserved under the integration. Also, log of a log concave function is concave by the Definition 3.1.2. Therefore,

$$log\,[\psi(t,\Delta)] \text{ is a concave function of } \Delta \text{ for all t.}$$

By the Theorem 4.1.1, for $A > 0$, $log\,[A + \psi(t,\Delta)]$ will be a concave function of Δ on the interval $(-t + \Delta_0, \Delta_0 + t)$ such that $\psi''(t,\Delta) \leq 0$ for any given t. Thus, $h_\Delta(t) = [A + \psi(t,\Delta)]$ is a log concave function of Δ on the interval $(-t + \Delta_0, \Delta_0 + t)$ for any given t. □

Theorem 4.1.2. *Assume that Theorem 3.1.6, Theorem 4.1.1, and Corollary 4.1.1 hold. There exists a* σ *so that*

$$L_n^\star(\Delta) = \sum_{i=1}^{n_1} -log[h_\Delta(x_i)] + \sum_{j=1}^{n_2} -log[h_\Delta(y_j - \Delta)]$$

is a convex function of Δ *on the interval* $(\Delta_0 - \sigma - c_1, \Delta_0 + \sigma + c_2)$*, where* $c_1 = max\{x_i; y_j\}$ *for* $i = 1, ..., n_1, j = 1, ..., n_2$ *and* $c_2 = min\{x_i; y_j\}$ *for* $i = 1, ..., n_1, j = 1, ..., n_2$.

Proof. By the Corollary 4.1.1 and Theorem 3.1.6, $h_\Delta(t)$ is a log concave function of Δ on the interval $(-t + \Delta_0, \Delta_0 + t)$ for each t. Therefore, the negative log of $h_\Delta(t)$ must be convex function of Δ on the same interval for each t. The -log likelihood function $L_n^\star(\Delta)$ contains observations $x_1, ...x_{n_1}, y_1, ..., y_{n_2}$ and for each observation, -log $h_\Delta(t)$ is convex function of Δ on the given interval by the

Corollary 4.1.1. In order to smooth the data, we choose a smoothing density $g(t) = \frac{1}{\sigma}\Phi(\frac{t}{\sigma})$ and use in the convolution operation to get $h_\Delta(t)$. The parameter σ is smoothing parameter and smoothness of $L_n^\star(\Delta)$ depends on σ.

Let $t = x_1$, -log $h_\Delta(x_1)$ is a convex function of Δ on $(\Delta_0-\sigma-x_1, \Delta_0+\sigma+x_1)$ by the Corollary 4.1.1. Similarly, for $t = x_2$, -log $h_\Delta(x_2)$ is a convex function of Δ on $(\Delta_0 - \sigma - x_2, \Delta_0 + \sigma + x_2)$. So, for each x_i, -log $h_\Delta(x_i)$ is a convex function of Δ on $(\Delta_0 - \sigma - x_i, \Delta_0 + \sigma + x_i)$. If we find the sum of the convex function, -log $h_\Delta(x_i)$, for each $i = 1, .., n_1$, then the sum must be a convex function of Δ on the intersection of the intervals. This intersection can be found by defining the lower limit as $\underset{i}{max}\{\Delta_0 - \sigma - x_i\}$ and the upper limit as $\underset{i}{min}\{\Delta_0 + \sigma + x_i\}$. Therefore, $\sum_{i=1}^{n_1} -log[h_\Delta(x_i)]$ is a convex function of Δ on $(\Delta_0 - \sigma - max\{x_i\}, \Delta_0 + \sigma + min\{x_i\})$.

Similarly, for $t = y_j$, $\sum_{j=1}^{n_2} -log[h_\Delta(y_j - \Delta)]$ is a convex function of Δ on $(\Delta_0 - \sigma - max\{y_j\}, \Delta_0 + \sigma + min\{y_j\})$. A basic calculus argument says that the sum of convex functions is a convex function. So,

$$L_n^\star(\Delta) = \sum_{i=1}^{n_1} -log[h_\Delta(x_i)] + \sum_{j=1}^{n_2} -log[h_\Delta(y_j - \Delta)]$$

is a convex function of Δ on the interval $(\Delta_0 - \sigma - c_1, \Delta_0 + \sigma + c_2)$, where $c1 = max\{x_i; y_j\}$ and $c_2 = min\{x_i; y_j\}$ for $i = 1, ..., n_1, j = 1, ..., n_2$ and σ is the smoothing scale parameter that can be estimated by observations x_i and y_j. □

Since $L_n^\star(\Delta)$ is a convex function of Δ on the interval $(\Delta_0 - \sigma - c_1, \Delta_0 + \sigma + c_2)$, then the minimum value of $L_n^\star(\Delta)$ exists and we can write it as

$$\Delta_S^\star = \underset{\Delta}{Argmin}\{L_n^\star(\Delta)\} \tag{4.10}$$

The following theorem shows the approximation of $L_n^\star(\Delta)$ to $L^\star(\Delta)$.

Theorem 4.1.3. *Assume that* $\frac{n_i}{n} \to \lambda_i$ *as* $n \to \infty$. *Then,*

$$P\left(\lim_{n\to\infty} \frac{1}{n} L_n^\star(\Delta) = L^\star(\Delta)\right) = 1 \tag{4.11}$$

where

$$L^\star(\Delta) = \lambda_1 \int -log[h_\Delta(x)]dF(x) + \lambda_2 \int -log[h_\Delta(y-\Delta)]dG(y) \tag{4.12}$$

Proof. Recall that

$$\begin{aligned}
L_n^\star(\Delta) &= -\sum_{i=1}^{n_1} log[h_\Delta(x_i)] - \sum_{j=1}^{n_2} log[h_\Delta(y_j - \Delta)] \\
\frac{1}{n} L_n^\star(\Delta) &= \frac{1}{n}\left[-\sum_{i=1}^{n_1} log[h_\Delta(x_i)] - \sum_{j=1}^{n_2} log[h_\Delta(y_j - \Delta)]\right] \\
&= -\frac{n_1}{n}\frac{1}{n_1}\left[\sum_{i=1}^{n_1} log[h_\Delta(x_i)]\right] - \frac{n_2}{n}\frac{1}{n_2}\left[\sum_{j=1}^{n_2} log[h_\Delta(y_j - \Delta)]\right]
\end{aligned}$$

If we take the limit as $n \to \infty$, we have

$$\lim_{n\to\infty}\frac{1}{n}L_n^\star(\Delta) \to \lambda_1 \int -log[h_\Delta(x)]dF(x) + \lambda_2 \int -log[h_\Delta(y-\Delta)]dG(y) = L^\star(\Delta)$$

where $n_i/n \to \lambda_i$ as $n \to \infty$ for $i = 1, 2$. Therefore, $\frac{1}{n}L_n^\star(\Delta) \to^{a.s} L^\star(\Delta)$ which completes the proof. □

In the next theorem, the identifiability condition will be discussed. Before that we investigate the convexity of the log likelihood function $L^\star(\Delta)$. It is known from the calculus that the convexity is preserved under the infinite sums and integrals. By the corollary 4.1.1, $h_\Delta(t)$ is a log concave function of Δ on the given interval by the same corollary. Thus, $-log[h_\Delta(t)]$ is a convex function of Δ for any t on the same interval given by the corollary 4.1.1. Therefore, $\int -log[h_\Delta(x)]dF(x)$ is a convex function of Δ for each x since convexity is preserved under the integration. Similarly, $\int -log[h_\Delta(y-\Delta)]dG(y)$ is a convex function of Δ for each y. Therefore, the log likelihood function $L^\star(\Delta)$ is a convex function of Δ since it is the sum of two convex functions. In fact, $L^\star(\Delta)$ is strictly convex since $-log$ implies strict convexity.

In the following theorem, it will be shown that $L^\star(\Delta)$ attaints its unique minimum at $\Delta = \Delta_0$. The result of the theorem indicates that $\Delta_S^\star = \underset{\Delta}{Argmin} L^\star(\Delta)$ would be consistent for true shift parameter Δ_0.

Theorem 4.1.4. *The function $L^\star(\Delta)$ attains its unique minimum at $\Delta = \Delta_0$, where Δ_0 is the true shift parameter.*

Proof. If we prove that $\frac{\partial L^\star(\Delta)}{\partial \Delta} = 0$ when $\Delta = \Delta_0$, then the proof of the theorem is complete.

Recall that

$$L^{\star}(\Delta) = \lambda_1 \int -log[h_{\Delta}(x)]dF(x) + \lambda_2 \int -log[h_{\Delta}(y-\Delta)]dG(y)$$
$$= \lambda_1 \int -log[h_{\Delta}(x)]dF(x) + \lambda_2 \int -logh_{\Delta}(y-\Delta)]dF(y-\Delta_0) \quad (4.13)$$

Let $y = y + \Delta_0$ in the second integral. Then,

$$L^{\star}(\Delta) = \lambda_1 \int -log[h_{\Delta}(x)]dF(x) + \lambda_2 \int -log[h_{\Delta}(y+\Delta_0-\Delta)]dF(y) \quad (4.14)$$

Take the first partial derivative with respect to Δ.

$$\frac{\partial L^{\star}(\Delta)}{\partial \Delta} = -\lambda_1 \int \frac{h'_{\Delta}(x)}{h_{\Delta}(x)}dF(x) - \lambda_2 \int \frac{(-1)h'_{\Delta}(y+\Delta_0-\Delta)}{h_{\Delta}(y+\Delta_0-\Delta)}dF(y)$$
$$= -\lambda_1 \int \frac{\lambda_2 \int g'(x+\Delta-y-\Delta_0)dF(y)}{\lambda_1 \int g(x-x^{\star})dF(x^{\star}) + \lambda_2 \int g(x+\Delta-y-\Delta_0)dF(y)}dF(x)$$
$$+ \lambda_2 \int \frac{\lambda_1 \int g'(y+\Delta_0-\Delta-x)dF(x)}{\lambda_1 \int g(y+\Delta_0-\Delta-x)dF(x) + \lambda_2 \int g(y-y^{\star})dF(y^{\star})}dF(y)$$

If we put $\Delta = \Delta_0$,

$$= -\lambda_1\lambda_2 \int \frac{\int g'(x-y)dF(y)}{\lambda_1 \int g(x-x^{\star})dF(x^{\star}) + \lambda_2 \int g(x-y)dF(y)}dF(x)$$
$$+ \lambda_1\lambda_2 \int \frac{\int g'(y-x)dF(x)}{\lambda_1 \int g(y-x)dF(x) + \lambda_2 \int g(y-y^{\star})dF(y^{\star})}dF(y)$$
$$= -\lambda_1\lambda_2 \int \frac{\int g'(x-y)dF(y)}{h(x)}dF(x)$$
$$+ \lambda_1\lambda_2 \int \frac{\int g'(y-x)dF(x)}{h(y)}dF(y)$$

Let y=x in the second integral, then

$$
\begin{aligned}
&= -\lambda_1\lambda_2 \int \frac{\int g'(x-y)dF(y)}{h(x)} dF(x) \\
&+ \lambda_1\lambda_2 \int \frac{\int g'(x-y)dF(y)}{h(x)} dF(x) \\
&= 0 \qquad (4.15)
\end{aligned}
$$

So $\frac{\partial L^\star(\Delta)}{\partial \Delta} = 0$ when $\Delta = \Delta_0$. Since $L^\star(\Delta)$ is strictly convex function, it attains its unique minimum at $\Delta = \Delta_0$. Thus, the proof is complete. □

4.2 Asymptotic Normality of $\Delta_S^\star$

In this section, we want to find asymptotic distribution and variance of $\Delta_S^\star = \underset{\Delta}{Argmin}\{L_n^\star(\Delta)\}$. We will satisfy the Pitman regularity conditions for the gradient function $S_n^\star(\Delta) = -\frac{\partial L_n^\star(\Delta)}{\partial \Delta}$ and find the efficacy of it. The Pitman regularity conditions are stated in the Hettmansperger-McKean (1998). We will also use the Theorem 1.5.8 of Hettmansperger-McKean (1998) in order to prove asymptotic normality result:

$$\sqrt{n}(\Delta_S^\star - \Delta_0) \rightarrow \mathrm{N}(0,\tfrac{1}{c^2})$$

where c is the efficacy of the $S_n^\star(\Delta)$. Without loss of generality, we will also assume that

Assumption (A2): Under $H_0 : \Delta_0 = 0$ and $G(y) = F(y - \Delta_0) = F(y)$

First recall that

$$L_n^\star(\Delta) = \sum_{i=1}^{n_1} -log[h_\Delta(x_i)] + \sum_{j=1}^{n_2} -log[h_\Delta(y_j - \Delta)] \tag{4.16}$$

where

$$\begin{aligned} h_\Delta(t) &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta)dG(y) \\ &= \lambda_1 \int g(t-x)dF(x) + \lambda_2 \int g(t-y+\Delta)dF(y) \end{aligned}$$

Define that $h_x(t) = \lambda_1 \int g(t-x)dF(x)$ and

$$h_y(t) = \lambda_2 \int g(t-y)dG(y) = \lambda_2 \int g(t-y)dF(y) \tag{4.17}$$

Then, $h_\Delta(t)$ can be written as

$$h_\Delta(t) = h_x(t) + h_y(t+\Delta)$$

So we can rewrite $L_n^\star(\Delta)$ in the following way,

$$\begin{aligned} L_n^\star(\Delta) = &-\sum_{i=1}^{n_1} log\left[h_x(x_i) + h_y(x_i+\Delta)\right] \\ &-\sum_{j=1}^{n_2} log\left[h_x(y_j-\Delta) + h_y(y_j - \Delta + \Delta)\right] \end{aligned} \tag{4.18}$$

Let $S_n^\star(\Delta) = -\frac{\partial L_n^\star(\Delta)}{\partial \Delta}$. Note that we can call $L_n^\star(\Delta)$ as the dispersion function and $S_n^\star(\Delta)$ as the gradient function.

The partial derivative of the first sum term in the $L_n^\star(\Delta)$ is

$$\frac{\partial L_n^\star(\Delta)}{\partial \Delta} = -\sum_{i=1}^{n_1} \frac{h_y'(x_i + \Delta)}{h_\Delta(x_i)} \tag{4.19}$$

Similarly, the partial derivative of the second sum term in the $L_n^\star(\Delta)$ is

$$\frac{\partial L_n^\star(\Delta)}{\partial \Delta} = -\sum_{j=1}^{n_2} \frac{(-1)h_x'(y_j - \Delta)}{h_\Delta(y_j - \Delta)} \tag{4.20}$$

If we add the two partial derivatives, we get

$$\begin{aligned} S_n^\star(\Delta) &= -\left[-\sum_{i=1}^{n_1} \frac{h_y'(x_i + \Delta)}{h_\Delta(x_i)} - \sum_{j=1}^{n_2} \frac{(-1)h_x'(y_j - \Delta)}{h_\Delta(y_j - \Delta)} \right] \\ &= \sum_{i=1}^{n_1} \frac{h_y'(x_i + \Delta)}{h_\Delta(x_i)} - \sum_{j=1}^{n_2} \frac{h_x'(y_j - \Delta)}{h_\Delta(y_j - \Delta)} \\ S_n^\star(\Delta) &= \sum_{i=1}^{n_1} \psi_1(x_i, \Delta) - \sum_{j=1}^{n_2} \psi_2(y_j, \Delta) \end{aligned} \tag{4.21}$$

where

$$\psi_1(t, \Delta) = \lambda_2 \frac{\int g'(t - y + \Delta) dG(y)}{h_\Delta(t)} = \frac{h_y'(t + \Delta)}{h_\Delta(t)} \tag{4.22}$$

and

$$\psi_2(t, \Delta) = \lambda_1 \frac{\int g'(t - \Delta - x) dF(x)}{h_\Delta(t - \Delta)} = \frac{h_x'(t - \Delta)}{h_\Delta(t - \Delta)} \tag{4.23}$$

The following definition of Pitman regularity conditions is given by Hettmansperger-McKean (1998).

Definition 4.2.1. We say that the function $S_n^\star(\Delta)$ is Pitman regular if the following four conditions hold;

1. The gradient function $S_n^\star(\Delta)$ is nonincreasing function of Δ.

2. Let $\bar{S}(\Delta) = (1/n^\gamma)S_n^\star(\Delta)$ for some $\gamma > 0$. There exists a function $\mu(\Delta) = E_\Delta[\bar{S}(0)] = E_0[\bar{S}(-\Delta)]$ such that $\mu(0) = 0$ and $\mu'(0) > 0$.

3. There is a constant $\sigma^2(0) = lim_{n\to\infty} nVar[\bar{S}(0)]$ such that

$$\sqrt{n}\{\tfrac{\bar{S}(0)}{\sigma(0)}\} \xrightarrow{D} N(0,1)$$

4. The asymptotic linearity of the process $S_n^\star(\Delta)$.

$$\sup_{\sqrt{n}|\delta|\le B} |S_n^\star(\frac{\delta}{\sqrt{n}}) - S_n^\star(0) + \delta\mu'(0)| \to^P 0 \qquad as \quad n \to \infty \tag{4.24}$$

for all $B > 0$.

Then, by the Definition 1.5.3 of Hettmansperger-McKean (1998), the quantity

$$c = \mu'(0)/\sigma(0)$$

is called the efficacy of $S_n^\star(\Delta)$.

Theorem 4.2.1. *The function $S_n^\star(\Delta)$ is Pitman regular if it satisfies the conditions of the definition 4.2.1. Then, we have*

$$\sqrt{n}\{\frac{\bar{S}(0)}{\sigma(0)}\} \xrightarrow{D} N(0,1) \tag{4.25}$$

where $\bar{S}(0) = \frac{1}{n}S_n^\star(0)$ *and*

$$\sigma(0) = \sqrt{\lim_{n\to\infty} nVar[\bar{S}(0)]} \tag{4.26}$$

Proof. We start proving each conditions of Pitman regularity from 1 to 4, respectively.

1. In our problem, $L_n^\star(\Delta)$ is a convex function of Δ on the given interval by Theorem 4.1.2. Then there is $\epsilon > 0$ such that $\frac{\partial L_n^\star(\Delta)}{\partial \Delta}$ is a nondecreasing function of Δ around $[\Delta_0 - \epsilon, \Delta_0 + \epsilon]$. Thus, $S_n^\star(\Delta) = -\frac{\partial L^\star(\Delta)}{\partial \Delta}$ is nonincreasing function of Δ within the interval $[\Delta_0 - \epsilon, \epsilon + \Delta_0]$.

2. Let $\bar{S}(\Delta) = \frac{1}{n}S_n^\star(\Delta)$. Since $E_\Delta[\bar{S}(0)] = E_0[\bar{S}(-\Delta)]$, we write

$$\begin{aligned} E_0[\bar{S}(-\Delta)] &= E_0\left[\frac{1}{n}\sum_{i=1}^{n_1}\psi_1(x_i,-\Delta) - \frac{1}{n}\sum_{j=1}^{n_2}\psi_2(y_j,-\Delta)\right] \\ &= E_0\left[\frac{1}{n}\sum_{i=1}^{n_1}\psi_1(x_i,-\Delta)\right] - E_0\left[\frac{1}{n}\sum_{j=1}^{n_2}\psi_2(y_j,-\Delta)\right] \\ &= E_0\left[\frac{n_1}{n}\sum_{i=1}^{n_1}\frac{\psi_1(x_i,-\Delta)}{n_1}\right] - E_0\left[\frac{n_2}{n}\sum_{j=1}^{n_2}\frac{\psi_2(y_j,-\Delta)}{n_2}\right] \end{aligned}$$

After moving the signs of Δ inside the ψ_1 and ψ_2 functions and taking the expectation of the each terms, we can write

$$\begin{aligned}
&= \lambda_1 \int \psi_1(x,\Delta)dF(x) - \lambda_2 \int \psi_2(y,\Delta)dG(y) \\
&= \lambda_1 \int \psi_1(x,\Delta)dF(x) - \lambda_2 \int \psi_2(y-\Delta,\Delta)dF(y) \\
&= \mu(\Delta) \qquad (4.27)
\end{aligned}$$

where $\frac{n_1}{n} \to \lambda_1$ and $\frac{n_2}{n} \to \lambda_2$ as n$\to \infty$.

Thus, we write

$$\begin{aligned}
\mu(\Delta) &= \lambda_1 \int \psi_1(x,\Delta)dF(x) - \lambda_2 \int \psi_2(y-\Delta,\Delta)dF(y) \\
&= \lambda_1 \int \lambda_2 \frac{\int g'(x-y+\Delta)dF(y)}{h_\Delta(x)} dF(x) \\
&- \lambda_2 \int \lambda_1 \frac{\int g'(y-\Delta-x)dF(x)}{h_\Delta(y-\Delta)} dF(y) \qquad (4.28)
\end{aligned}$$

Let's put $\Delta = 0$ in the expression of $\mu(\Delta)$. Therefore, under the assumption (A2), we have

$$\begin{aligned}
\mu(0) &= \lambda_1 * \lambda_2 \int \frac{\int g'(x-y)dF(y)}{h(x)} dF(x) \\
&- \lambda_2 * \lambda_1 \int \frac{\int g'(y-x)dF(x)}{h(y)} dF(y)
\end{aligned}$$

Again we let y=x in the second integral term of $\mu(0)$, then we have

$$\mu(0) = 0$$

Second part of the condition is to show that $\mu'(0) > 0$.

First note that we have the expression (4.27) which is

$$\mu(\Delta) = \lambda_1 \int \psi_1(x, \Delta) dF(x) - \lambda_2 \int \psi_2(y, \Delta) dG(y)$$

The partial derivatives of ψ_1 and ψ_2 functions with respect to Δ is

$$\psi_1'(t, \Delta) = \frac{\partial \psi_1(t, \Delta)}{\partial \Delta} = \frac{[h_y'(t+\Delta)]^2 - h_y''(t+\Delta) * h_\Delta(t)}{[h_\Delta(t)]^2} \tag{4.29}$$

and

$$\psi_2'(t, \Delta) = \frac{\partial \psi_2(t, \Delta)}{\partial \Delta} = (-1)\frac{[h_x'(t-\Delta)]^2 - h_x''(t-\Delta) * h_\Delta(t-\Delta)}{[h_\Delta(t-\Delta)]^2} \tag{4.30}$$

Let's take the derivative of $\mu(\Delta)$ with respect to Δ. So we have

$$\mu'(\Delta) = \frac{\partial \mu(\Delta)}{\partial \Delta}$$

$$= \lambda_1 \int \psi_1'(x,\Delta)dF(x) + \lambda_2 \int \psi_2'(y,\Delta)dG(y)$$
$$= \lambda_1 \int \frac{[h_y'(x+\Delta)]^2 - h_y''(x+\Delta) * h_\Delta(x)}{[h_\Delta(x)]^2} dF(x)$$
$$+ \lambda_2 \int \frac{[h_x'(y-\Delta)]^2 - h_x''(y-\Delta) * h_\Delta(y-\Delta)}{[h_\Delta(y-\Delta)]^2} dG(y)$$

By the assumption (A2) and taking $\Delta = 0$, we have

$$\mu'(0) = \lambda_1 \int \frac{[h_y'(x)]^2 - h_y''(x) * h(x)}{[h(x)]^2} dF(x)$$
$$+ \lambda_2 \int \frac{[h_x'(y)]^2 - h_x''(y) * h(y)}{[h(y)]^2} dF(y) \qquad (4.31)$$

The numerator in the first integral is positive since the function $h(x) = \int g(x - y)dF(y) > 0$ and the log concavity of h(x) implies that $[h_y'(x)]^2 - h_y''(x) * h(x) \geq 0$. Similarly, the numerator in the second integral is also positive with similar reason. Thus, the first and second integrals are both positive which implies that $\mu'(0) > 0$. Therefore, second condition of the Pitman regularity condition satisfied.

3. To prove the third condition of the Pitman regularity, we follow the path of Hettmansperger and McKean (1998)

$$\sigma^2(0) = lim_{n\to\infty} nVar[\bar{S}(0)]$$
$$= lim_{n\to\infty} nVar[\frac{1}{n}\sum_{i=1}^{n_1}\psi_1(x_i) - \frac{1}{n}\sum_{j=1}^{n_2}\psi_2(y_j)]$$

Since two samples are i.i.d and mutually independent, we can write

$$
\begin{aligned}
&= lim_{n\to\infty} n[\frac{1}{n^2}\sum_{i=1}^{n_1} Var[\psi_1(x_i)] + \frac{1}{n^2}\sum_{j=1}^{n_2} Var[\psi_2(y_j)]] \\
&= lim_{n\to\infty}[n * \frac{1}{n^2} n_1 * Var[\psi_1(x_1)] + n * \frac{1}{n^2} n_2 * Var[\psi_2(y_1)]] \\
&= lim_{n\to\infty}[\frac{n_1}{n} * Var[\psi_1(x_1)] + \frac{n_2}{n} * Var[\psi_2(y_1)]] \\
\sigma^2(0) &= \lambda_1 * Var[\psi_1(x)] + \lambda_2 * Var[\psi_2(y)] \qquad (4.32)
\end{aligned}
$$

Therefore, by the Central Limit Theorem, we have

$$\sqrt{n}\{\frac{\bar{S}(0) - E_0[\bar{S}(0)]}{\sigma(0)}\} \longrightarrow^D N(0,1).$$

□

4. In order to prove the fourth condition of the Pitman Regularity, we will give the following theorem. The theorem is from Hetmansperger-McKean(1998) and it proves the asymptotic linearity of $S_n^\star(\Delta)$.

Theorem 4.2.2. *Let $\bar{S}(\Delta) = (1/n^\gamma)S_n^\star(\Delta)$ for some $\gamma > 0$ such that $S_n^\star(\Delta)$ satisfies 1,2 and 3 of the Pitman Regularity conditions. If for any $\delta \in \mathbb{R}$*

$$nVar[\bar{S}(\frac{\delta}{\sqrt{n}}) - \bar{S}(0)] \longrightarrow 0 \qquad as \quad n \to \infty.$$

Then,

$$\sup_{|\delta|\le B} |\sqrt{n}\bar{S}(\frac{\delta}{\sqrt{n}}) - \sqrt{n}\bar{S}(0) + \mu'(0)\delta| \longrightarrow^P 0$$

for any $B > 0$

Proof. As we know from the previous section, $S_n^\star(\Delta)$ satisfies the conditions 1, 2 and 3 of Pitman regularity. By the Theorem 1.5.6 of Hettmansperger-McKean (1998), we only need to show

$$nVar[\bar{S}(\frac{\delta}{\sqrt{n}}) - \bar{S}(0)] \longrightarrow 0 \qquad as \quad n \to \infty$$

Let $\bar{S}(\Delta) = \frac{1}{n}S_n^\star(\Delta)$ for $\gamma = 1$. Replace Δ with $\frac{\delta}{\sqrt{n}}$ in $\bar{S}(\Delta)$.

By the definition, we have

$$= lim_{n\to\infty}\{nVar[\bar{S}(\frac{\delta}{\sqrt{n}}) - \bar{S}(0)]\}$$

$$= lim_{n\to\infty}\{nVar[\frac{1}{n}\sum_{i=1}^{n_1}\psi_1(x_i, \frac{\delta}{\sqrt{n}}) - \frac{1}{n}\sum_{j=1}^{n_2}\psi_2(y_j, \frac{\delta}{\sqrt{n}})$$

$$-\frac{1}{n}\sum_{i=1}^{n_1}\psi_1(x_i, 0) + \frac{1}{n}\sum_{j=1}^{n_2}\psi_2(y_j, 0)]\}$$

$$= lim_{n\to\infty}\{n * \frac{1}{n^2}\sum_{i=1}^{n_1}Var[\psi_1(x_i, \frac{\delta}{\sqrt{n}}) - \psi_1(x_i, 0)]$$

$$+ n * \frac{1}{n^2}\sum_{j=1}^{n_2}Var[\psi_2(y_j, \frac{\delta}{\sqrt{n}}) - \psi_2(y_j, 0)]\}$$

$$= lim_{n\to\infty}\{n * \frac{1}{n^2} * n_1 * Var[\psi_1(x_1, \frac{\delta}{\sqrt{n}}) - \psi_1(x_1, 0)]$$

$$+ n * \frac{1}{n^2} * n_2 * Var[\psi_2(y_1, \frac{\delta}{\sqrt{n}}) - \psi_2(y_1, 0)]\}$$

Now if we move the limit inside, we get $\frac{\delta}{\sqrt{n}} \rightarrow 0$ and $n_i/n \rightarrow \lambda_i$ for i=1,2 as $n \rightarrow \infty$

$$
\begin{aligned}
&= \lambda_1 Var[\psi_1(x,0) - \psi_1(x,0)] \\
&+ \lambda_2 Var[\psi_2(y,0) - \psi_2(y,0)] \\
&= 0 \qquad (4.33)
\end{aligned}
$$

Thus, $lim_{n\rightarrow\infty}\{nVar[\bar{S}(\frac{\delta}{\sqrt{n}}) - \bar{S}(0)]\} = 0$. Therefore, the fourth condition of Pitman regularity also holds. □

The following theorem proves the asymptotic normality of $\Delta_S^\star$. The Pitman regularity conditions and the Theorem 1.5.8 from Hettmansperger-McKean(1998) proves the result.

Theorem 4.2.3. *Suppose that $S_n^\star(\Delta)$ is Pitman regular with efficacy c. Then, by the Theorem 1.5.8 of Hettmansperger-McKean(1998),*

$$\sqrt{n}(\Delta_S^\star - \Delta_0) \rightarrow N(0, \frac{1}{c^2}) \qquad (4.34)$$

and the efficacy c is

$$c = \frac{\mu'(0)}{\sigma(0)}$$

where $\mu(0)$ is given by the expression(4.31) and $\sigma(0)$ is given by the expres-

sion(4.32).

Proof. Since $S_n^\star(\Delta)$ satisfies the Pitman regularity conditions and by the Theorem 1.5.8 of Hettmansperger-McKean (1998), we can conclude that

$$\sqrt{n}(\Delta_S^\star - \Delta_0) \rightarrow \mathrm{N}(0, \tfrac{1}{c^2})$$

Thus, the proof is complete. □

CHAPTER V

ASYMPTOTIC PROPERTIES OF THE PROPOSED ESTIMATOR

5.1 Asymptotic Linearity of the $S_n(\Delta)$

In Chapter 4, we discussed the asymptotic properties of the estimator $\Delta_S^{\star}$. It is clear that it is not practical to use $\Delta_S^{\star}$ as an estimator of Δ_0 since we need to know true density function f(x). This was the reason we propose $\widehat{\Delta}_S$ as an estimator of the true shift parameter since it does not depend on the true density function f(x) in the estimation process. In this chapter we will develop asymptotic properties of the $\widehat{\Delta}_s$ by using the gradient function $S_n(\Delta)$.

Now let's switch our attention to $S_n(\Delta)$ which is the gradient function of $L_n(\Delta)$. Recall that we have

$$L_n(\Delta) = \sum_{i=1}^{n_1} -log[\widehat{h}_{\Delta}(x_i)] + \sum_{j=1}^{n_2} -log[\widehat{h}_{\Delta}(y_j - \Delta)] \tag{5.1}$$

where

$$\begin{aligned}\widehat{h}_{\Delta}(t) &= \frac{1}{n}[\sum_{l=1}^{n_1} g(t - x_l) + \sum_{k=1}^{n_2} g(t - y_k + \Delta)] \\ &= \widehat{h}_{\mathbf{x}}(t) + \widehat{h}_{\mathbf{y}}(t + \Delta)\end{aligned}$$

where $\widehat{h}_{\mathbf{x}}(t) = \frac{1}{n}\sum_{l=1}^{n_1} g(t - x_l)$ and $\widehat{h}_{\mathbf{y}}(t) = \frac{1}{n}\sum_{k=1}^{n_2} g(t - y_k)$. Let

$$S_n(\Delta) = \frac{\partial L_n(\Delta)}{\partial \Delta} = -\sum_{i=1}^{n_1} \frac{\hat{h}'_{\mathbf{y}}(x_i + \Delta)}{\widehat{h}_{\Delta}(x_i)}. + \sum_{j=1}^{n_2} \frac{\hat{h}'_{\mathbf{x}}(y_j - \Delta)}{\widehat{h}_{\Delta}(y_j - \Delta)}$$

$$S_n(\Delta) = -\sum_{i=1}^{n_1} \widehat{\psi}_1(x_i, \Delta) + \sum_{j=1}^{n_2} \widehat{\psi}_2(y_j, \Delta) \tag{5.2}$$

where

$$\widehat{\psi}_1(x, \Delta) = \frac{\hat{h}'_{\mathbf{y}}(x + \Delta)}{\widehat{h}_{\Delta}(x)} = \frac{\frac{1}{n}\sum_{k=1}^{n_2} g'(x - y_k + \Delta)}{\widehat{h}_{\Delta}(x)} \tag{5.3}$$

and

$$\widehat{\psi}_2(y, \Delta) = \frac{\hat{h}'_{\mathbf{x}}(y - \Delta)}{\widehat{h}_{\Delta}(y - \Delta)} = \frac{\frac{1}{n}\sum_{l=1}^{n_1} g'(y - \Delta - x_l)}{\widehat{h}_{\Delta}(y - \Delta)} \tag{5.4}$$

First,we replace Δ with $\frac{\delta}{\sqrt{n}}$ in the $S_n(\Delta)$. Then, we have

$$S_n(\frac{\delta}{\sqrt{n}}) = -\sum_{i=1}^{n_1} \widehat{\psi}_1(x_i, \frac{\delta}{\sqrt{n}}) + \sum_{j=1}^{n_2} \widehat{\psi}_2(y_j, \frac{\delta}{\sqrt{n}})$$

Second, let's take $\delta = 0$. Then,

$$S_n(0) = -\sum_{i=1}^{n_1} \widehat{\psi}_1(x_i, 0) + \sum_{j=1}^{n_2} \widehat{\psi}_2(y_j, 0)$$

Now we recall $S_n^\star(\Delta) = \frac{\partial L_n^\star(\Delta)}{\partial \Delta}$ from Chapter 4

$$S_n^\star(\Delta) = -\sum_{i=1}^{n_1} \psi_1(x_i, \Delta) + \sum_{j=1}^{n_2} \psi_2(y_j, \Delta) \tag{5.5}$$

where

$$\psi_1(t,\Delta) = \lambda_2 \frac{\int g'(t-y+\Delta)dG(y)}{h_\Delta(t)} = \frac{h'_y(t+\Delta)}{h_\Delta(t)}$$

and

$$\psi_2(t,\Delta) = \lambda_1 \frac{\int g'(t-\Delta-x)dF(x)}{h_\Delta(t-\Delta)} = \frac{h'_x(t-\Delta)}{h_\Delta(t-\Delta)}$$

Again replace Δ with $\frac{\delta}{\sqrt{n}}$ in the $S_n^\star(\Delta)$.

$$S_n^\star(\frac{\delta}{\sqrt{n}}) = -\sum_{i=1}^{n_1} \psi_1(x_i, \frac{\delta}{\sqrt{n}}) + \sum_{j=1}^{n_2} \psi_2(y_j, \frac{\delta}{\sqrt{n}}) \tag{5.6}$$

We now subtract $S_n^\star(\frac{\delta}{\sqrt{n}})$ from the $S_n(\frac{\delta}{\sqrt{n}})$.

$$\begin{aligned} S_n(\frac{\delta}{\sqrt{n}}) - S_n^\star(\frac{\delta}{\sqrt{n}}) &= \left[-\sum_{i=1}^{n_1} \widehat{\psi}_1(x_i, \frac{\delta}{\sqrt{n}}) + \sum_{j=1}^{n_2} \widehat{\psi}_2(y_j, \frac{\delta}{\sqrt{n}})\right] \\ &- \left[-\sum_{i=1}^{n_1} \psi_1(x_i, \frac{\delta}{\sqrt{n}}) + \sum_{j=1}^{n_2} \psi_2(y_j, \frac{\delta}{\sqrt{n}})\right] \\ &= \sum_{j=1}^{n_2} [\widehat{\psi}_2(y_j, \frac{\delta}{\sqrt{n}}) - \psi_2(y_j, \frac{\delta}{\sqrt{n}})] - \sum_{i=1}^{n_1} [\widehat{\psi}_1(x_i, \frac{\delta}{\sqrt{n}}) - \psi_1(x_i, \frac{\delta}{\sqrt{n}})] \end{aligned} \tag{5.7}$$

Similarly,

$$\begin{aligned} S_n(0) - S_n^\star(0) &= \left[-\sum_{i=1}^{n_1} \widehat{\psi}_1(x_i, 0) + \sum_{j=1}^{n_2} \widehat{\psi}_2(y_j, 0)\right] + \left[\sum_{i=1}^{n_1} \psi_1(x_i, 0) - \sum_{j=1}^{n_2} \psi_2(y_j, 0)\right] \\ &= \sum_{j=1}^{n_2} [\widehat{\psi}_2(y_j, 0) - \psi_2(y_j, 0)] - \sum_{i=1}^{n_1} [\widehat{\psi}_1(x_i, 0) - \psi_1(x_i, 0)] \end{aligned} \tag{5.8}$$

After this step, we can state the following theorem which is necessary to show the asymptotic linearity of $S_n(\Delta)$.

Theorem 5.1.1. *Assume that the assumptions of Theorem 4.2.2 hold. Then,*

$$\sup_{\sqrt{n}|\delta|\leq B} \left| \left(\frac{1}{\sqrt{n}}S_n(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S_n(0)\right) - \left(\frac{1}{\sqrt{n}}S_n^\star(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S_n^\star(0)\right) \right| \rightarrow^P 0$$

as $n \rightarrow \infty$ *for any* $B > 0$

Proof. First we start our proof by noting that by the Theorem 4.2.2, we have

$$\sup_{\sqrt{n}|\delta|\leq B} \left|\frac{1}{\sqrt{n}}S_n^\star(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S_n^\star(0) + \delta\mu'(0)\right| \rightarrow^P 0 \qquad as \quad n \rightarrow \infty \tag{5.9}$$

for any $B > 0$.

Now define a function

$$S(\frac{\delta}{\sqrt{n}}) = S_n(\frac{\delta}{\sqrt{n}}) - S_n^\star(\frac{\delta}{\sqrt{n}}) \tag{5.10}$$

Then if we put $\delta = 0$, we get

$$S(0) = S_n(0) - S_n^\star(0) \tag{5.11}$$

By the Mean Value Theorem, we can write

$$S(\frac{\delta}{\sqrt{n}}) - S(0) = S'(\tilde{\delta})(\frac{\delta}{\sqrt{n}} - 0) \tag{5.12}$$

where $|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|$. Thus, we have

$$S'(\tilde{\delta}) = \sum_{j=1}^{n_2}[\widehat{\psi}'_2(y_j, \tilde{\delta}) - \psi'_2(y_j, \tilde{\delta}] - \sum_{i=1}^{n_1}[\widehat{\psi}'_1(x_i, \tilde{\delta}) - \psi'_1(x_i, \tilde{\delta})] \tag{5.13}$$

If we multiply the both side of the expression(5.12) with $\frac{1}{\sqrt{n}}$, we get

$$\begin{aligned}\frac{1}{\sqrt{n}}S(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S(0) &= \frac{1}{\sqrt{n}}S'(\tilde{\delta})(\frac{\delta}{\sqrt{n}} - 0) \\ &= \frac{1}{n}S'(\tilde{\delta}) \qquad (5.14)\end{aligned}$$

where $|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|$.

Therefore, by the expression(5.14), we can write

$$\begin{aligned}&\left[\frac{1}{\sqrt{n}}S(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S(0)\right] \\ &= \frac{\delta}{n}\sum_{j=1}^{n_2}[\widehat{\psi}'_2(y_j, \tilde{\delta}) - \psi'_2(y_j, \tilde{\delta}] - \frac{\delta}{n}\sum_{i=1}^{n_1}[\widehat{\psi}'_1(x_i, \tilde{\delta}) - \psi'_1(x_i, \tilde{\delta})] \\ &= \frac{\delta}{n}\frac{n_2}{n_2}\sum_{j=1}^{n_2}[\widehat{\psi}'_2(y_j, \tilde{\delta}) - \psi'_2(y_j, \tilde{\delta})] - \frac{\delta}{n}\frac{n_1}{n_1}\sum_{i=1}^{n_1}[\widehat{\psi}'_1(x_i, \tilde{\delta}) - \psi'_1(x_i, \tilde{\delta})] \\ &= \frac{\delta * n_2}{n}\frac{\sum_{j=1}^{n_2}[\widehat{\psi}'_2(y_j, \tilde{\delta}) - \psi'_2(y_j, \tilde{\delta})]}{n_2} - \frac{\delta * n_1}{n}\frac{\sum_{i=1}^{n_1}[\widehat{\psi}'_1(x_i, \tilde{\delta}) - \psi'_1(x_i, \tilde{\delta})]}{n_1} \qquad (5.15)\end{aligned}$$

If we take the absolute value of both sides, we get

$$\leq \frac{\delta * n_2}{n}\frac{\sum_{j=1}^{n_2}|\widehat{\psi}'_2(y_j, \tilde{\delta}) - \psi'_2(y_j, \tilde{\delta})|}{n_2} + \frac{\delta * n_1}{n}\frac{\sum_{i=1}^{n_1}|\widehat{\psi}'_1(x_i, \tilde{\delta}) - \psi'_1(x_i, \tilde{\delta})|}{n_1} \tag{5.16}$$

Since each sum terms in the above inequality is an average, we can put max as an upper bound

$$|\frac{1}{\sqrt{n}}S(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S(0))| \leq \frac{\delta * n_2}{n} \max_j\{|\widehat{\psi}_2'(y_j, \tilde{\delta}) - \psi_2'(y_j, \tilde{\delta}|\}$$
$$+ \frac{\delta * n_1}{n} \max_i\{|\widehat{\psi}_1'(x_i, \tilde{\delta}) - \psi_1'(x_i, \tilde{\delta})|\}$$

Now take the sup of both sides, left and right side, the result we get

$$\sup_{|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|} |\frac{1}{\sqrt{n}}S(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S(0))| \leq \sup_{|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|} \{\frac{\delta * n_2}{n} \max_j\{|\widehat{\psi}_2'(y_j, \tilde{\delta}) - \psi_2'(y_j, \tilde{\delta}|\}$$
$$+ \frac{\delta * n_1}{n} \max_i\{|\widehat{\psi}_1'(x_i, \tilde{\delta}) - \psi_1'(x_i, \tilde{\delta})|\}\}$$

If we distribute the sup inside, we have

$$\leq \frac{\delta * n_2}{n} \sup_{|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|} \max_j\{|\widehat{\psi}_2'(y_j, \tilde{\delta}) - \psi_2'(y_j, \tilde{\delta}|\}$$
$$+ \frac{\delta * n_1}{n} \sup_{|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|} \max_i\{|\widehat{\psi}_1'(x_i, \tilde{\delta}) - \psi_1'(x_i, \tilde{\delta})|\}\}$$

By taking the $limsup_{n \to \infty}$ of both sides, we get

$$\leq \delta * \lambda_2 \limsup_{n\to\infty} \sup_{|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|} \max_j\{|\widehat{\psi}_2'(y_j, \tilde{\delta}) - \psi_2'(y_j, \tilde{\delta}|\}$$
$$+ \delta * \lambda_1 \limsup_{n\to\infty} \sup_{|\tilde{\delta}| \leq |\frac{\delta}{\sqrt{n}}|} \max_i\{|\widehat{\psi}_1'(x_i, \tilde{\delta}) - \psi_1'(x_i, \tilde{\delta})|\} \tag{5.17}$$

The term inside $\max_j$ is

$$\widehat{\psi}_2'(y_j,\tilde{\delta}) - \psi_2'(y_j,\tilde{\delta}) = \frac{[\frac{n_1}{n}\frac{\sum_{i=1}^{n_1} g'(y_j-\tilde{\delta}-x_i)}{n_1}]^2 - \frac{n_1}{n}\frac{\sum_{i=1}^{n_1} g''(y_j-\tilde{\delta}-x_i)}{n_1}\widehat{h}_{\tilde{\delta}}(y_j-\tilde{\delta})}{[\widehat{h}_{\tilde{\delta}}(y_j-\tilde{\delta})]^2}$$

$$-\frac{[\lambda_1 \int g'(y_j-\tilde{\delta}-x)dF(x)]^2 - \lambda_1 \int g''(y_j-\tilde{\delta}-x)dF(x)h_{\tilde{\delta}}(y_j-\tilde{\delta})}{[h_{\tilde{\delta}}(y_j-\tilde{\delta})]^2}$$

In a more compact form,

$$\begin{aligned}\widehat{\psi}_2'(y_j,\tilde{\delta}) - \psi_2'(y_j,\tilde{\delta}) &= \frac{n_1}{n}\frac{[\widehat{h}_{\tilde{\delta}}'(y_j-\tilde{\delta})]^2 - \widehat{h}_{\tilde{\delta}}''(y_j-\tilde{\delta})\widehat{h}_{\tilde{\delta}}(y_j-\tilde{\delta})}{[\widehat{h}_{\tilde{\delta}}(y_j-\tilde{\delta})]^2} \\ &\quad - \lambda_1\frac{[h_{\tilde{\delta}}'(y_j-\tilde{\delta})]^2 - h_{\tilde{\delta}}''(y_j-\tilde{\delta})h_{\tilde{\delta}}(y_j-\tilde{\delta})}{[h_{\tilde{\delta}}(y_j-\tilde{\delta})]^2} \end{aligned} \tag{5.18}$$

We know that x_i's and y_j's are i.i.d, g is measurable function, $\tilde{\delta}$ belong to compact set Θ. $E_x[g(y_j,\tilde{\delta}] < \infty$. Then, by the Theorem 2 of Wellner(2001), for fixed y_j, we have

$$\sup_{\tilde{\delta}\in\Theta} |\widehat{h}_{\tilde{\delta}}(y_j-\tilde{\delta}) - h_{\tilde{\delta}}(y_j-\tilde{\delta})| \to 0 \text{ as } n\to\infty$$

By assuming $E_x[g'(y_j,\tilde{\delta})] < \infty$, we also have,

$$\sup_{\tilde{\delta}\in\Theta} |\widehat{h}_{\tilde{\delta}}'(y_j-\tilde{\delta}) - h_{\tilde{\delta}}'(y_j-\tilde{\delta})| \to 0 \text{ as } n\to\infty$$

Similarly, by assuming $E_x[g''(y_j,\tilde{\delta})] < \infty$, we have,

$$\sup_{\tilde{\delta}\in\Theta} |\widehat{h}_{\tilde{\delta}}''(y_j-\tilde{\delta}) - h_{\tilde{\delta}}''(y_j-\tilde{\delta})| \to 0 \text{ as } n\to\infty$$

By the Slutsky's Theorem, the following statements are also true,

$$\sup_{\tilde{\delta}\in\Theta} |[\widehat{h}'_{\tilde{\delta}}(y_j - \tilde{\delta})]^2 - [h'_{\tilde{\delta}}(y_j - \tilde{\delta})]^2| \to 0 \text{ as } n \to \infty$$

and

$$\sup_{\tilde{\delta}\in\Theta} |\widehat{h}''_{\tilde{\delta}}(y_j - \tilde{\delta})\widehat{h}_{\tilde{\delta}}(y_j - \tilde{\delta}) - h''_{\tilde{\delta}}(y_j - \tilde{\delta})h_{\tilde{\delta}}(y_j - \tilde{\delta})| \to 0 \text{ as } n \to \infty$$

and

$$\sup_{\tilde{\delta}\in\Theta} | \left\{ [\widehat{h}'_{\tilde{\delta}}(y_j - \tilde{\delta})]^2 - \widehat{h}''_{\tilde{\delta}}(y_j - \tilde{\delta})\widehat{h}_{\tilde{\delta}}(y_j - \tilde{\delta}) \right\} - \left\{ [h'_{\tilde{\delta}}(y_j - \tilde{\delta})]^2 - h''_{\tilde{\delta}}(y_j - \tilde{\delta})h_{\tilde{\delta}}(y_j - \tilde{\delta}) \right\} | \to 0 \; as \; n \to \infty \tag{5.19}$$

Therefore, by the Slutsky's Theorem, we have

$$\begin{aligned} &\sup_{\tilde{\delta}\in\Theta} |\widehat{\psi}'_2(y_j, \tilde{\delta}) - \psi'_2(y_j, \tilde{\delta})| \\ &= \sup_{\tilde{\delta}\in\Theta} | \frac{n_1}{n} \frac{\{[\widehat{h}'_{\tilde{\delta}}(y_j - \tilde{\delta})]^2 - \widehat{h}''_{\tilde{\delta}}(y_j - \tilde{\delta})\widehat{h}_{\tilde{\delta}}(y_j - \tilde{\delta})\}}{[\widehat{h}_{\tilde{\delta}}(y_j - \tilde{\delta})]^2} \\ &- \lambda_1 \frac{\{[h'_{\tilde{\delta}}(y_j - \tilde{\delta})]^2 - h''_{\tilde{\delta}}(y_j - \tilde{\delta})h_{\tilde{\delta}}(y_j - \tilde{\delta})\}}{[h_{\tilde{\delta}}(y_j - \tilde{\delta})]^2} | \longrightarrow 0 \end{aligned} \tag{5.20}$$

where $\frac{n_1}{n} \to \lambda_1$ as $n \to \infty$.

Similarly, the term inside $\max_i$ is

$$\widehat{\psi}'_1(x_i, \tilde{\delta}) - \psi'_1(x_i, \tilde{\delta}) = \frac{[\frac{n_2}{n} \frac{\sum_{k=1}^{n_2} g'(x_i - y_k + \tilde{\delta})}{n_2}]^2 + \frac{n_2}{n} \sum_{k=1}^{n_2} g''(x_i - y_k + \tilde{\delta})\widehat{h}_{\tilde{\delta}}(x_i)}{[\widehat{h}_{\tilde{\delta}}(x_i)]^2}$$
$$- \frac{[\lambda_2 \int g'(x_i - y + \tilde{\delta})dG(y)]^2 + \lambda_2 \int g''(x_i - y + \tilde{\delta})dG(y)h_{\tilde{\delta}}(x_i)}{[h_{\tilde{\delta}}(x_i)]^2}$$

In a more compact form,

$$\widehat{\psi}_1'(x_i,\tilde{\delta}) - \psi_1'(x_i,\tilde{\delta})$$
$$= \frac{n_2}{n}\frac{[\widehat{h}_{\tilde{\delta}}'(x_i)]^2 - \widehat{h}_{\tilde{\delta}}''(x_i)\widehat{h}_{\tilde{\delta}}(x_i)}{[\widehat{h}_{\tilde{\delta}}(x_i)]^2} - \lambda_2\frac{[h_{\tilde{\delta}}'(x_i)]^2 - h_{\tilde{\delta}}''(x_i)h_{\tilde{\delta}}(x_i)}{[h_{\tilde{\delta}}(x_i)]^2}$$

By using the similar arguments as above and by the Slutsky's Theorem,

$$\sup_{\tilde{\delta}\in\Theta}|\widehat{\psi}_1'(x_i,\tilde{\delta}) - \psi_1'(x_i,\tilde{\delta})|$$
$$= \sup_{\tilde{\delta}\in\Theta}|\frac{n_2}{n}\frac{[\widehat{h}_{\tilde{\delta}}'(x_i)]^2 - \widehat{h}_{\tilde{\delta}}''(x_i)\widehat{h}_{\tilde{\delta}}(x_i)}{[\widehat{h}_{\tilde{\delta}}(x_i)]^2} - \lambda_2\frac{[h_{\tilde{\delta}}'(x_i)]^2 - h_{\tilde{\delta}}''(x_i)h_{\tilde{\delta}}(x_i)}{[h_{\tilde{\delta}}(x_i)]^2}| \longrightarrow 0 \quad (5.21)$$

where $\frac{n_2}{n} \to \lambda_2$ as $n \to \infty$.

Therefore, combining results of the (5.20) and (5.21), we have that

$$\sup_{|\delta|\le|\frac{\delta}{\sqrt{n}}|}|\frac{1}{\sqrt{n}}S(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S(0)| \to^P 0 \; as \; n \to \infty$$

Since

$$\frac{1}{\sqrt{n}}S(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S(0) = \left(\frac{1}{\sqrt{n}}S_n(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S_n(0)\right) - \left(\frac{1}{\sqrt{n}}S_n^\star(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S_n^\star(0)\right)$$

The result concludes the proof. □

Theorem 5.1.2. *Assume that Theorem 4.2.2 and Theorem 5.1.1 hold. Then, we* *have*

$$\sup_{\sqrt{n}|\delta|\leq B} |\frac{1}{\sqrt{n}} S_n(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}} S_n(0) + \delta\mu'(0)| \to^P 0 \qquad as \quad n \to \infty \tag{5.22}$$

for any $B > 0$

Proof. By the Theorem 4.2.2 and Theorem 5.1.1, the proof is complete. □

By the result of this theorem, it can be written that

$$\frac{1}{\sqrt{n}} S_n(\frac{\delta}{\sqrt{n}}) = \frac{1}{\sqrt{n}} S_n(0) - \delta\mu'(0) + o_P(1) \tag{5.23}$$

which is going to be the foundation of our asymptotic normality proof.

5.2 Influence Function of $S_n(\Delta)$

In section 5.1, we proved the asymptotic linearity of $S_n(\Delta)$. In this section, we will find the influence function of $S_n(\Delta)$ that will determine the asymptotic distribution of $S_n(\Delta)$. We will use the influence function and asymptotic linearity results of $S_n(\Delta)$ to prove asymptotic normality of the estimator $\widehat{\Delta}_S$. Asymptotic normality is also needed for developing hypothesis testing and confidence interval estimation of the Δ_0. Note that our estimator $\widehat{\Delta}_S$ is translation equivariant estimator by the Theorem 2.6 and we can assume that G(y)=F(y) by the assumption A2.

Definition 5.2.1. Let T be a statistical functional defined on a space of distribution functions and let F denote a distribution function in the domain of T. Let $\gamma_t(x)$ be a point mass distribution function at t. Then, the Gateux derivative of T(F)

$$lim_{s\to 0}\frac{T[(1-s)F(x)+s\gamma_t(x)]-T[F(x)]}{s}=\psi_F(t) \tag{5.24}$$

is called the influence function of T(F).

Note that the influence function $\psi_F(t)$ is the derivative of the functional $T[(1-s)F(x)+s\gamma_t(x)]$ at s=0. It measures the influence of a point in the estimator. A functional said to be robust if influence function is bounded (Hettmansperger-McKean 1998).

In our investigation of the influence function, we will use the idea of Huber (1981) who suggested using the empirical distribution functions in the functional that is used to estimate the true parameter. Therefore, before we find the influence function of $S_n(0)$, we need to write $S_n(0)$ as a function of its empirical distribution functions F_{n_1} and G_{n_2}. Recall that

$$S_n(\Delta)=-\sum_{i=1}^{n_1}\frac{\frac{1}{n}\sum_{k=1}^{n_2}g'(x_i-y_k+\Delta)}{\widehat{h}_\Delta(x_i)}-\sum_{j=1}^{n_2}\frac{\frac{1}{n}(-1)\sum_{l=1}^{n_1}g'(y_j-\Delta-x_l)}{\widehat{h}_\Delta(y_j-\Delta)}$$

where $\widehat{h}_\Delta(t)=\frac{1}{n}[\sum_{l=1}^{n_1}g(t-x_l)+\sum_{k=1}^{n_2}g(t-y_k+\Delta)]$.

Let $\Delta=0$ in $S_n(\Delta)$. We can rewrite a functional $S_n(0)=T(F_{n_1},G_{n_2})$ in terms of empirical distribution functions F_{n_1} and G_{n_2}. So we have

$$T(F_{n_1}, G_{n_2}) = -\frac{n_1}{n_1}\sum_{i=1}^{n_1} \frac{\frac{n_2}{nn_2}\sum_{k=1}^{n_2} g'(x_i - y_k)}{\frac{n_1}{nn_1}\sum_{l=1}^{n_1} g(x_i - x_l) + \frac{n_2}{nn_2}\sum_{k=1}^{n_2} g(x_i - y_k)} + \frac{n_2}{n_2}\sum_{j=1}^{n_2} \frac{\frac{n_1}{nn_1}\sum_{l=1}^{n_1} g'(y_j - x_l)}{\frac{n_1}{nn_1}\sum_{l=1}^{n_1} g(y_j - x_l) + \frac{n_2}{nn_2}\sum_{k=1}^{n_2} g(y_j - y_k)}$$

After appliying Reimann-Stieljes aproximation, we can write

$$T(F_{n_1}, G_{n_2}) = -n_1 \int \frac{\frac{n_2}{n}\int g'(x-y)dG_{n_2}(y)}{\frac{n_1}{n}\int g(x-x^\star)dF_{n_1}(x^\star) + \frac{n_2}{n}\int g(x-y)dG_{n_2}(y)} dF_{n_1}(x) + n_2 \int \frac{\frac{n_1}{n}\int g'(y-x)dF_{n_1}(x)}{\frac{n_1}{n}\int g(y-x)dF_{n_1}(x) + \frac{n_2}{n}\int g(y-y^\star)dG_{n_2}(y^\star)} dG_{n_2}(y) \tag{5.25}$$

Thus, $S_n(0) = T(F_{n_1}, G_{n_2})$ can be written in terms of empirical distribution functions F_{n_1} and G_{n_2}.

Theorem 5.2.1. *Let $S_n(0) = T(F_{n_1}, G_{n_2})$ where F_{n_1} and G_{n_2} are empirical distribution functions of X and Y respectively. Let*

$$F_1(x, \epsilon) = (1-\epsilon)F(x) + \epsilon\gamma_{x_1}(x) \tag{5.26}$$

and

$$F_2(y, \epsilon) = (1-\epsilon)G(y) + \epsilon\gamma_{y_1}(y) \tag{5.27}$$

where $0 < \epsilon < 1$ is the proportion of the contamination and γ_t is the distribution *function for a point mass at t. Then, the influence function of $S_n(0)$ is*

$$IF(x_1, y_1) = \frac{dT(F_1(x,\epsilon), F_2(y,\epsilon))}{d\epsilon} |_{\epsilon=0}$$
$$= \frac{n_1 n_2}{n}[\Omega(x_1) - \Omega(y_1)]$$

where

$$\Omega(t) = \int \frac{g'(y-t)}{\int g(y-x)dF(x)} dG(y) - \frac{\int g'(t-y)dG(y)}{\int g(t-y)dG(y)}$$

Proof. We start our proof by replacing $F_{n_1}(x)$ and $G_{n_2}(y)$ with $F_1(x,\epsilon)$ and $F_2(y,\epsilon)$ in the $S_n(0)$ given by expression(5.25). After the replacement, we have

$$T(F_2(x), F_2(y)) = -n_1 \int \frac{\frac{n_2}{n}\int g'(x-y)dF_2(y)}{\frac{n_1}{n}\int g(x-x^\star)dF_1(x^\star) + \frac{n_2}{n}\int g(x-y)dF_2(y)} dF_1(x)$$
$$+ n_2 \int \frac{\frac{n_1}{n}\int g'(y-x)dF_1(x)}{\frac{n_1}{n}\int g(y-x)dF_1(x) + \frac{n_2}{n}\int g(y-y^\star)dF_2(y^\star)} dF_2(y)$$

After putting $F_1(x) = (1-\epsilon)F(x) + \epsilon\gamma_{x_1}(x)$ and $F_2(y) = (1-\epsilon)G(y) + \epsilon\gamma_{y_1}(y)$ in the $S_n(0)$, and applying change of variable technique, we get

$$T(F_1(x), F_2(y))$$
$$= -\frac{n_1 n_2}{n}[(1-\epsilon) \int \frac{[(1-\epsilon)\int g'(x-y)dG(y) + \epsilon\int g'(x-y)d\gamma_{y_1}(y)]}{[(1-\epsilon)\int g(x-y)dF(y)] + \epsilon[\int g(x-y)d\gamma_{x_1}(y)} dF(x)$$

$$+\epsilon\int\frac{[(1-\epsilon)\int g'(x-y)dG(y)+\epsilon\int g'(x-y)d\gamma_{y_1}(y)]}{[(1-\epsilon)\int g(x-y)dG(y)+\epsilon\int g(x-y)d\gamma_{x_1}(y)]}d\gamma_{x_1}(x)]$$

$$+\frac{n_2n_1}{n}[(1-\epsilon)\int\frac{[(1-\epsilon)\int g'(y-x)dF(x)+\epsilon\int g'(y-x)d\gamma_{x_1}(x)]}{[(1-\epsilon)\int g(y-x)dF(x)+\epsilon\int g(y-x)d\gamma_{x_1}(x)]}dG(y)$$

$$+\epsilon\int\frac{[(1-\epsilon)\int g'(y-x)dF(x)+\epsilon\int g'(y-x)d\gamma_{x_1}(x)]}{[(1-\epsilon)\int g(y-x)dF(x)+\epsilon\int g(y-x)d\gamma_{x_1}(x)]}d\gamma_{y_1}(y)] \tag{5.28}$$

Let $\frac{dT(F_1(x,\epsilon),F_2(y,\epsilon))}{d\epsilon}|_{\epsilon=0}$. Apply the chain rule and replace $\epsilon=0$, we get

$$=-\frac{n_1n_2}{n}[(-1)\int\frac{\int g'(x-y)dG(y)}{\int g(x-y)dG(y)}dF(x)$$

$$+\int\frac{[(-1)\int g'(x-y)dG(y)+\int g'(x-y)d\gamma_{y_1}(y)]\int g(x-y)dG(y)}{[\int g(x-y)dG(y)]^2}dF(x)$$

$$-\int\frac{[(-1)\int g(x-y)dG(y)+\int g(x-y)d\gamma_{x_1}(y)]\int g'(x-y)dG(y)}{[\int g(x-y)dG(y)]^2}dF(x)$$

$$+\int\frac{\int g'(x-y)dG(y)}{\int g(x-y)dG(y)}d\gamma_{x_1}(x)]$$

$$+\frac{n_2n_1}{n}[(-1)\int\frac{\int g'(y-x)dF(x)}{\int g(y-x)dF(x)}dG(y)$$

$$+\int\frac{[(-1)\int g'(y-x)dF(x)+\int g'(y-x)d\gamma_{x_1}(x)]\int g(y-x)dF(x)}{[\int g(y-x)dF(x)]^2}dG(y)$$

$$-\int\frac{[(-1)\int g(y-x)dF(x)+\int g(y-x)d\gamma_{x_1}(x)]\int g'(y-x)dF(x)}{[\int g(y-x)dF(x)]^2}dG(y)$$

$$+\int\frac{\int g'(y-x)dF(x)}{\int g(y-x)dF(x)}d\gamma_{y_1}(y)] \tag{5.29}$$

By applying change of variable in the first term, we can further simplify the equation. Note that γ is a point mess and $\int g(x-y)d\gamma_{y_1}(y) = g(x-y_1)$ and $\int g(y-x)d\gamma_{x_1}(x) = g(y-x_1)$

$$\begin{aligned} T(F_1(x), F_2(y)) = &-\frac{n_1 n_2}{n}[\int \frac{g'(x-y_1)\int g(x-y)dG(y)}{[\int g(x-y)dG(y)]^2}dF(x) \\ &+ \frac{\int g'(x_1-y)dG(y)}{\int g(x_1-y)dG(y)}] \\ &+ \frac{n_2 n_1}{n}[\int \frac{g'(y-x_1)\int g(y-x)dF(x)}{[\int g(y-x)dF(x)]^2}dG(y) \\ &+ \frac{\int g'(y_1-x)dF(x)}{\int g(y_1-x)dF(x)}] \end{aligned} \quad (5.30)$$

Finally, in the last equation, the square term$[\int g(x-y)dG(y)]^2$ in the first and third denominators cancel out the integral term $\int g(x-y)dG(y)$ in the numerators. After this cancellation, we have

$$\begin{aligned} = &-\frac{n_1 n_2}{n}\left[\int \frac{g'(x-y_1)}{\int g(x-y)dG(y)}dF(x) + \frac{\int g'(x_1-y)dG(y)}{\int g(x_1-y)dG(y)}\right] \\ &+ \frac{n_2 n_1}{n}\left[\int \frac{g'(y-x_1)}{\int g(y-x)dF(x)}dG(y) + \frac{\int g'(y_1-x)dF(x)}{\int g(y_1-x)dF(x)}\right] \end{aligned} \quad (5.31)$$

The first and last terms in the expression depend on y_1, we can put these terms together. Similarly, the second and third terms depend on x_1, we can put them together.

Thus, we have

$$\begin{aligned}
&\frac{dT(F_1(x,\epsilon),F_2(y,\epsilon))}{d\epsilon}\Big|_{\epsilon=0} \\
&= \frac{n_2n_1}{n}\left\{\int \frac{g'(y-x_1)}{\int g(y-x)dF(x)}dG(y) - \frac{\int g'(x_1-y)dG(y)}{\int g(x_1-y)dG(y)}\right\} \\
&- \left\{\int \frac{g'(x-y_1)}{\int g(x-y)dG(y)}dF(x) - \frac{\int g'(y_1-x)dF(x)}{\int g(y_1-x)dF(x)}\right\}
\end{aligned} \tag{5.32}$$

Let

$$\Omega(t) = \left\{\int \frac{g'(y-t)}{\int g(y-x)dF(x)}dG(y) - \frac{\int g'(t-y)dG(y)}{\int g(t-y)dG(y)}\right\} \tag{5.33}$$

Therefore, the influence function is

$$IF(x_1,y_1) = \frac{n_2n_1}{n}\left[\Omega(x_1) - \Omega(y_1)\right] \tag{5.34}$$

which completes the proof. □

One of the properties of influence function is that $E[IF(x_1,y_1)] = 0$. Note that X and Y are independent by the assumption A1.

$$\begin{aligned}
E[IF(x_1,y_1)] &= E\left[\frac{n_1n_2}{n}(\Omega(x_1)-\Omega(y_1))\right] \\
&= \frac{n_1n_2}{n}\left[E(\Omega(x_1)) - E(\Omega(y_1))\right] \\
&= \frac{n_1n_2}{n}\left[\int \Omega(x_1)dF(x_1) - \int \Omega(y_1)dG(y_1)\right]
\end{aligned} \tag{5.35}$$

Let $y_1 = x_1$ in the second integration and G(y)=F(y) under H_0

$$= \frac{n_1 n_2}{n}\left[\int \Omega(x_1)dF(x_1) - \int \Omega(x_1)dF(x_1)\right]$$

$$E[IF(x_1, y_1)] = 0$$

Now let's look at the individual terms in the $IF(x_1, y_1)$. First, consider the function $\Omega(x_1)$,

$$\begin{aligned}
E[\Omega(x_1)] &= \int \Omega(x_1)dF(x_1) \\
&= \int \left[\int \frac{g'(y-x_1)}{\int g(y-x)dF(x)}dG(y) - \frac{\int g'(x_1-y)dG(y)}{\int g(x_1-y)dG(y)}\right] dF(x_1) \\
&= \int\int \frac{g'(y-x_1)}{\int g(y-x)dF(x)}dG(y)dF(x_1) - \int \frac{\int g'(x_1-y)dG(y)}{\int g(x_1-y)dG(y)}dF(x_1)
\end{aligned}$$

By the independence of x and y, we can apply change of order of integration in the first term

$$= \int \left[\int \frac{g'(y-x_1)}{\int g(y-x)dF(x)}dF(x_1)\right] dG(y) - \int \frac{\int g'(x_1-y)dG(y)}{\int g(x_1-y)dG(y)}dF(x_1)$$

The denominator term $\int g(y-x)dF(x)$ is independent of $dF(x_1)$, therefore, moving $dF(x_1)$ to numerator keeps the integratiy of the double integral

$$= \int \left[\frac{\int g'(y-x_1)dF(x_1)}{\int g(y-x)dF(x)}\right] dG(y) - \int \frac{\int g'(x_1-y)dG(y)}{\int g(x_1-y)dG(y)}dF(x_1)$$

Let $y = x_1$ and $dG(y) = dF(x_1)$ in the first integral, then

$$= \int \frac{\int g'(x_1 - y)dG(y)}{\int g(x_1 - x)dF(x)} dF(x_1) - \int \frac{\int g'(x_1 - y)dG(y)}{\int g(x_1 - y)dG(y)} dF(x_1)$$

Let x=y and $dF(x) = dG(y)$ in the denominator of first term, then

$$= \int \frac{\int g'(x_1 - y)dG(y)}{\int g(x_1 - y)dG(y)} dF(x_1) - \int \frac{\int g'(x_1 - y)dG(y)}{\int g(x_1 - y)dG(y)} dF(x_1)$$

Therefore, difference of the two same integral indicates

$$E[\Omega(x_1)] = 0$$

Similarly, using the same arguments, we can show that $E[\Omega(y_1)] = 0$.

5.3 Asymptotic Normality of $S_n(0)$

The influence function (IF) often provides a representation suggesting the asymptotic distribution of the estimator. Based on the above influence function result, we will prove that gradient function $S_n(0)$ is asymptotically normally distributed. For one-sample case Huber (1981, Section 2.5) point out that, under regularity conditions,

$$\sqrt{n}(T(F_n) - T(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(x_i) + o_P(1) \tag{5.36}$$

where $o_P(1)$ tends to 0 in probability. By applying the Central Limit Theorem to the first term on the right side, and using the fact that E[IF(x)]=0, then we have

$$\sqrt{n}(T(F_n) - T(F)) \to N(0, E[IF^2(x)]) \tag{5.37}$$

For two-sample case, we will extent Huber's idea to our case. We will state following theorem in order to prove asymptotic normality of $S_n(0)$. Here, recall that $S_n(0) = T(F_{n_1}, G_{n_2})$ from section 5.25 and

Assumption (A3): Assume that $o_P(1)$ tends to 0 in probability.

Remark 5.3.1. Since we assumed that the remainder term $o_P(1)$ goes to "zero", it must be proven in order to be theoretically true. The proof is not easy to show but it is often true so that asymptotically negligible. In the literature, several authors, for example, Huber (1981) and Serfling (1980) mentioned that this proof is very difficult task to be accomplished. Therefore, we will assume that $o_P(1)$ goes to "zero" but the proof will be shown in the future studies related this topic.

Theorem 5.3.1. *Assume that the Theorem 5.2.1 and (A3) holds. Then we have*

$$\sqrt{n}[T(F_{n_1}, G_{n_2}) - T(F, G)] \longrightarrow N(0, v^2(0)) \tag{5.38}$$

where $v^2(0) = (\frac{n_1 n_2}{\sqrt{n}n})^2 \, [n_1 E[\Omega^2(x)] + n_2 E[\Omega^2(y)]]$ *and* $n = n_1 + n_2$.

Proof. By the Theorem 5.2.1 and Huber (1981, section 2.5), we can write

$$\sqrt{n}\,[T(F_{n_1}, G_{n_2}) - T(F,G)]$$
$$= \frac{1}{\sqrt{n}}\left\{\frac{n_1 n_2}{n}\left[\sum_{i=1}^{n_1}\Omega(x_i) - \sum_{j=1}^{n_2}\Omega(y_j)\right]\right\} + o_P(1) \tag{5.39}$$

where $\Omega(x_i)$ and $\Omega(y_j)$ are independent since sample X and sample Y independent by the assumption and both samples are i.i.d. Then, by the Central Limit Theorem, the right side of the expression(5.39) is Normally distributed with mean 0 and variance

$$v^2(0) = (\frac{n_1 n_2}{\sqrt{nn}})^2\left[n_1 E[\Omega^2(x)] - n_2 E[\Omega^2(y)]\right] \tag{5.40}$$

Note that expected value of the influence function is zero(i.e $E[IF(x,y)] = 0$) and by the assumption (A3) the remainder term $o_P(1)$ goes to 0 in probability. □

5.4 Asymptotic Normality of $\widehat{\Delta}_S$

In the final part of the chapter V, we will show the asymptotic normality of estimator $\widehat{\Delta}_S$. The result of this section will also help us to develop an asymptotic α level hypothesis testing and $(1-\alpha)100\%$ confidence interval estimation.. We will use asymptotic linearity results for $S_n(\Delta)$ from chapter IV in proving the asymptotic normality of $\widehat{\Delta}_S$.

Recall that by the Theorem 5.1.2, we have

$$\sup_{\sqrt{n}|\delta|\leq c} |\frac{1}{\sqrt{n}}S_n(\frac{\delta}{\sqrt{n}}) - \frac{1}{\sqrt{n}}S_n(0) + \delta\mu'(0)| \rightarrow^P 0 \qquad as \quad n \rightarrow \infty$$

This result is an important one that will be used to prove Asymptotic Normality of the estimator $\widehat{\Delta}_S$. We will give following theorem in order to prove asymptotic normality.

Theorem 5.4.1. *Assume that Theorem 5.1.2, Theorem 5.3.1 and the assumption (A3) hold. Then, we have*

$$\sqrt{n}(\widehat{\Delta}_S - \Delta_0) \xrightarrow{D} N(0, V^2(0))$$

where $V(0) = v(0)/n\mu'(0)$.

Proof. By the Theorem 5.1.2, we have the asymptotic linearity result of

$$\frac{1}{\sqrt{n}}S_n(\frac{\delta}{\sqrt{n}}) = \frac{1}{\sqrt{n}}S_n(0) - \delta\mu'(0) + o_P(1)$$

If we substitute $\sqrt{n}\widehat{\Delta}_S = \delta$ and by the assumption (A2), we can write,

$$\sqrt{n}\widehat{\Delta}_S = \frac{S_n(0)}{\sqrt{n}\mu'(0)} + o_P(1)$$

If we know the asymptotic distribution of $\frac{S_n(0)}{\sqrt{n}\mu'(0)}$, we can find the asymptotic distribution of $\sqrt{n}\widehat{\Delta}_S$.

By the Theorem 5.3.1, we have $S_n(0) = T(F_{n_1}, G_{n_2})$ and

$$\sqrt{n}S_n(0) \to^D N(0, v^2(0))$$

where $v^2(0) = (\frac{n_1 n_2}{\sqrt{n}n})^2\,[n_1 E[\Omega^2(x)] + n_2 E[\Omega^2(y)]]$.

Therefore, we also have

$$S_n(0) \to^D N(0, v^2(0)/n)$$

and

$$\frac{S_n(0)}{\sqrt{n}\mu'(0)} \to^D N(0, \frac{v^2(0)}{n^2(\mu'(0))^2})$$

By the asymptotic linearity of $S_n(0)$

$$\sqrt{n}\widehat{\Delta}_S = \frac{S_n(0)}{\sqrt{n}\mu'(0)} + o_P(1)$$

The remainder term $o_P(1) \to 0$ in probability as $n \to \infty$ by the assumption (A3). Therefore, we have

$$\sqrt{n}(\widehat{\Delta}_S - \Delta_0) \xrightarrow{D} N(0, V^2(0))$$

where $V(0) = \frac{v(0)}{n\mu'(0)}$. By the assumption A2, $\Delta_0 = 0$. Thus, we can also write that

$$\sqrt{n}\widehat{\Delta}_S \xrightarrow{D} N(0, V(0))$$

where $V(0) = \frac{v(0)}{n\mu'(0)}$. □

CHAPTER VI

HYPOTHESIS TESTING AND CONFIDENCE INTERVAL ESTIMATION

6.1 Hypothesis Testing

In this section, we will develop a hypothesis testing that will use the result of chapter V. By the asymptotic normality of the $S_n(0)$(Theorem 5.3.1), we have

$$\frac{\sqrt{n}(S_n(0) - E[S_n(0)])}{\sqrt{v^2(0)}} \to Normal(0,1)$$

where $v^2(0) = (\frac{n_1 n_2}{\sqrt{n}n})^2 \left[n_1 E[\Omega^2(x)] + n_2 E[\Omega^2(y)]\right]$. Next, we will develop an asymptotic level α test of hypothesis $H_0 : \Delta_0 = 0$ vs $H_A : \Delta_0 > 0$ which is

$$\text{Reject } H_0 \text{ in favor of } H_A \text{ if } S_n(0) > z_\alpha v(0)/\sqrt{n}$$

Here $v(0)$ has to be estimated from data since it will not be practical to use it directly because underlying distribution functions of the samples are unknown to us. We approximate $v(0)$ in the following way, first recall that we have

$$\Omega(t) = \left\{ \int \frac{g'(y-t)}{\int g(y-x)dF(x)} dG(y) - \frac{\int g'(t-y)dG(y)}{\int g(t-y)dG(y)} \right\} \tag{6.1}$$

Then $\Omega(t)$ can be approximated by the following expression

$$\widehat{\Omega}(t) = \left\{ \frac{1}{n_2} \sum_{j=1}^{n_2} \frac{g'(y_j - t)}{\frac{1}{n_1}\sum_{i=1}^{n_1} g(t - x_i)} - \frac{\frac{1}{n_2}\sum_{j=1}^{n_2} g'(t - y_j)}{\frac{1}{n_2}\sum_{j=1}^{n_2} g(t - y_j)} \right\} \tag{6.2}$$

Then $v(0)$ can be approximated by

$$\widehat{v}^2(0) = (\frac{n_1 n_2}{\sqrt{n}n})^2 \left[n_1 \frac{1}{n_1} \sum_{i=1}^{n_1} \widehat{\Omega}(x_i) + n_2 \frac{1}{n_2} \sum_{j=1}^{n_2} \widehat{\Omega}(y_j) \right] \tag{6.3}$$

6.2 Confidence Interval Estimation

As we discussed in chapter II, the estimate $\widehat{\Delta}_S$ is a solution to the equation $S_n(\widehat{\Delta}_S) \doteq 0$. Based on this equation and asymptotic distribution result of $S_n(0)$, $(1-\alpha)100\%$ level confidence interval for Δ_0 can be found by solving

$$S_n(\widehat{\Delta}_U) \doteq -z_{\alpha/2} v(0)/\sqrt{n} \text{ and } S_n(\widehat{\Delta}_L) \doteq z_{\alpha/2} v(0)/\sqrt{n}$$

which yield a $(1-\alpha)100\%$ level confidence interval $(\widehat{\Delta}_L, \widehat{\Delta}_U)$ for the shift parameter Δ_0. Similar to the estimator $\widehat{\Delta}_S$, $\widehat{\Delta}_L$ and $\widehat{\Delta}_U$ can be found by using an iterative algorithm. Examples are given in Chapter 7 to illustrate finding a $(1-\alpha)100\%$ level confidence intervals for the true shift parameter Δ_0.

The result of asymptotic normality of $\widehat{\Delta}_S$ indicates that

$$\sqrt{n}\widehat{\Delta}_S \xrightarrow{D} N(0, V^2(0))$$

where $V(0) = \frac{v(0)}{n\mu'(0)}$. Therefore, based on the asymptotic distribution of $\widehat{\Delta}_S$, a $(1-\alpha)100\%$ level confidence interval can be written as

$$\widehat{\Delta}_S \pm z_{\alpha/2} V(0)/\sqrt{n}$$

The estimation of V(0) and properties of this asymptotic level confidence interval will be investigated in a future study.

CHAPTER VII

NUMERICAL EXAMPLES AND A SIMULATION STUDY

7.1 Numerical Examples

Example 7.1.1. (*Random Generated Data Example*) We mentioned that $L_n(\Delta)$ is a quasiconvex function of Δ. To see this property clearly, the plot of dispersion function $L_n(\Delta)$ is generated by using two arbitrary samples, see the Figure 8. Here, for simplicity reasons, smoother density function g(t) chosen to be g(t)=Normal$(0, \sigma^2)$ in the convolution and σ^2 will be estimated from the data such that $\widehat{\sigma}^2 = \frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}$. Sample X with $n_1 = 20$ is generated from Normal(0,1) and sample Y with $n_2 = 10$ is generated from Normal(1,1). So, the true shift is $\Delta_0 = 1$.

In Table 1, the estimate of Δ_0 is found by using LS, H-L and proposed Smoothing method. As we see from the Table 1, the proposed Smoothing method performed better or equal compare to LS and H-L. A 95% confidence interval is also indicates that the proposed method performed equal to the other two. Actually, it is close to the least square in estimating the shift and the confidence interval. This can be a good indication since the data comes from Normal Distribution in which least square works the best.

Table 1: Estimate of Δ_0 and 95% Confidence Limits

Method	$\widehat{\Delta}$	95% Lower	95% Upper
Smoothing	1.0257	0.1466	1.8723
Hodges Lehman	1.0485	0.1127	1.8422
Least Square	1.0135	0.1575	1.8694

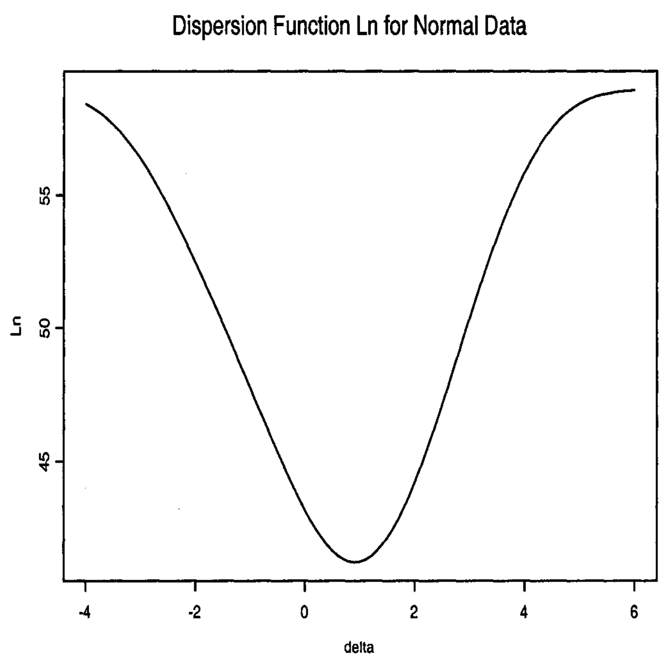

Figure 8: Log-likelihood function $L_n(\Delta)$ of the random data

Example 7.1.2. (*Quail Data Example*) This example comes from a drug screening program that finds compounds which reduce low-density lipoprotein(LDL) in quails. See McKean, Vidmar and Sievers (1989) for the discussion of this screen. The Main purpose of the study was to examine the effects of a drug which is designed to lower the cholesterol levels. Two groups of quails have been randomly selected, the first group were fed with a special diet and given the drug, the second group were fed with same diet but didn't get the drug over the same time period. The first group is referred to the treatment group and the second group refered to the control group in the study. The data is displayed in Table 2. we can observe

Table 2: Quail Data

Control	64 49 54 64 97 66 76 44 71 89 70 72 71 55 60 62 46 77 86 71
Treatment	40 31 50 48 152 44 74 38 81 64

that 5th observation in the treatment group is an outlier. This type of outliers are common in this drug study program, see McKean et al(1989). We can also note that treatment group has lower values than the control group. This can be a rough indication of the effectiveness of the drug but we should also consider that sample size in treatment group is 10 against 20 in the control group.

Let θ_C and θ_T denote the true median levels of the control and treatment populations, respectively. Then, one can say that the parameter of interest should be $\Delta = \theta_C - \theta_T$. Since we are interested in the alternative hypothesis that effectiveness of the drug in the study, the hypothesis set up should be

$$H_0 : \Delta_0 = 0 \text{ vs } H_A : \Delta_0 > 0$$

By using the data given in Table 2, we run our code to find proposed Smoothing, Hodges-Lehmann(H-L) and Least Squares(L-S) estimations of the shift parameter Δ_0. We also find a 90% level confidence intervals for the shift Δ_0. The results are given in Table 3. Here, for simplicity reasons, smoother density function g(t) chosen to be

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} \sim N[0, \sigma^2]$$

and σ^2 estimated by $\widehat{\sigma}^2 = \frac{(n_1-1)Mad_x^2 + (n_2-1)Mad_y^2}{n_1+n_2-2}$ from the data.

As we see from the table, the proposed estimation method is rejecting H_0 since

Table 3: Estimate of Δ_0 and 90% Confidence Limits

Method	$\widehat{\Delta}$	90% Lower	90% Upper
Smoothing	16.02	5.10	26.20
Hodges Lehman	14	-2	24
Least Square	5	-10.25	20.25

"Zero" is not included in the interval, on the other hand, other two have included "Zero" in the intervals and they don't reject H_0. Thus, the proposed method finds that the treatment effect is significant. We can also note that proposed method gives us the narrowest confidence interval compared to the other two confidence intervals.

In the following Table 4, we illustrated same problem without the outlier which is the observation #5 in the sample Y. As we see that, LS jump 10 units

Table 4: Quail Data Results without the Outlier

Method	$\widehat{\Delta}$	90% Lower	90% Upper
Smoothing	16.22	8.58	23.30
Hodges Lehman	16	5	26
Least Square	14.97	4.75	25.20

from 5 to almost 15, HL increased 2 units from 14 to 16, Smoothing only increased by 0.2 from 16.02 to 16.22. %90 confidence intervals are now very similar to each other, almost estimated the same interval. "Zero" is included in all three intervals and all of them rejects $H_0 : \Delta_0 = 0$, therefore, there is significant treatment effect.

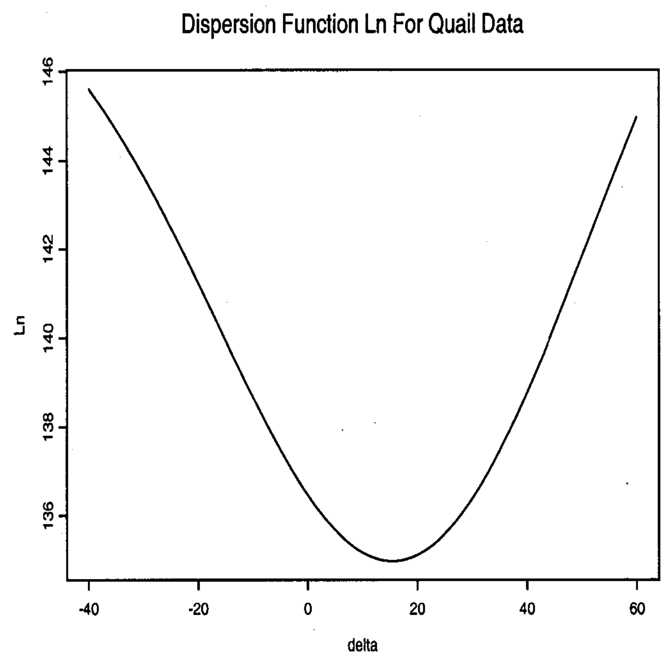

Figure 9: Log-likelihood function $L_n(\Delta)$ of the quail data

7.2 A Simulation Study

In this section, we are going to illustrate a bootstrap simulation study to compare Asymptotic Relative Efficiencies(ARE) of Smoothing, Hodges-Lehmann and Least Square. It is very well known that (see Hetmansperger and McKean 1998), ARE of Hodges-Lehmann with respect to Least Square is 0.955 when errors are normally distributed with mean 0 and variance 1. It is natural to ask what is the ARE of Smoothing with respect to Least Square when we have the same underlying normal distribution. Next question, maybe, what is the ARE of Smoothing versus Least Square and Hodges-Lehmann when we contaminated underlying distributions. We will answer these questions with comparing finite sample relative efficiencies(RE) of the methods. ARE is the ratio of bootstrap variances of Smoothing estimation of the shift with respect to Hodges-Lehmann estimation and with respect to Least Square estimation. For this simulation, the smoother function g(t) will be Gaussian Density(Normal Density) function,

$$g(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{t^2}{2\sigma^2}} \sim Normal[0, \sigma^2]$$

and σ^2 will be estimated by $\widehat{\sigma}^2$. Different choices of $\widehat{\sigma}^2$ will be investigated. We will generate two random samples from Normal Distributions such as,

$$\text{Let } X_1 \ldots X_{n_1} \sim Normal(\mu_x, \sigma_0^2)$$

and

$$\text{Let } Y_1 \ldots Y_{n_2} \sim Normal(\mu_y, \sigma_0^2)$$

Therefore, the true shift parameter will be $\Delta_0 = \mu_y - \mu_x$. If we shift the sample Y with true shift parameter Δ_0, the combine shifted sample becomes,

$$X_1, ...X_{n_1}, Y_1 - \Delta_0, ...Y_{n_2} - \Delta_0 \sim Normal(0, \sigma_0^2)$$

We will also define that

$$C.N(\epsilon, \mu_1) \equiv (1-\epsilon)Normal(0, \sigma_0^2) + \epsilon\ Normal(\mu_1, \sigma_1^2) \qquad 0 < \epsilon < 1.$$

which is a contaminated normal with contamination level ϵ.

We run 1000 replications in the simulation by generating $n_1 = 20$ observations from Normal(2,4) and $n_2 = 10$ observations from Normal(4,4). Therefore, true shift parameter is $\Delta_0 = 2$ and $\sigma_0^2 = 4$. Furthermore, the contaminated part is generated from Normal(10,16) where $\mu_1 = 10$ and $\sigma_1^2 = 16$. Estimated Asymptotic Relative Efficiency(ARE) values of the Smoothing relative to Hodges-Lehmann and Least Square is given in the tables with a different ϵ contamination levels.

The smoothing parameter σ^2 in the smoother density function g(t) must be estimated from the data. There are several ways to estimate this parameter. We can use the pooled variance if we assume that underlying distribution is normal or approximately normal.

1. σ^2 can be estimated by $\hat{\sigma}_M^2 = \frac{(n_1-1)mad_x^2 + (n_2-1)mad_y^2}{n_1+n_2-2}$, where mad is the median absolute deviation, a robust scale parameter estimate from the data.

2. σ^2 can be estimated by $\widehat{\sigma}_\tau^2 = \frac{(n_1-1)\tau_x^2+(n_2-1)\tau_y^2}{n_1+n_2-2}$, where τ is a robust scale parameter estimate proposed by Huber(1981).

3. σ^2 can be estimated by $\widehat{\sigma}_S^2 = \frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}$, where $S^2 = \frac{\sum_{i=1}^n (x_i-\bar{x})^2}{n-1}$ is a non-robust classical estimate of the variance from the data.

4. If we use Kernel Density Estimation method instead of "smoothing by convolution" idea in the replacement of true density function f(x), the bandwidth parameter(b_n) must be estimated from the data. We will use optimal bandwidth parameter selection procedure proposed by Sheather and Jones(1991).

First, we will run all three smoothing parameter σ^2 replacement options and see which one gives us the highest relative efficiency compared to Hodges-Lehman(H-L) and Least Square(L-S) procedures under the contamination. Later, we will run a simulation to investigate the Kernel Density Estimation which is alternative to the smoothing by convolution.

We start the simulation with using Median Absolute Deviation (MAD) as a smoothing parameter σ estimate. We compare relative efficiencies of Smoothing(SM), H-L and L-S. The Following Table 5 shows the results. The relative efficiencies of Smoothing(SM) versus H-L and L-S compared by using Huber's τ as a smoothing parameter estimate. At $\epsilon = 0$, ARE of SM with respect to L-S is 0.9649 which is almost equal to H-L under no contamination case. L-S is the best at 0% contamination compared to H-L and SM. At the 10% contamination of the data, ARE of SM versus H-L is 1.0103 which indicates Smoothing is 1%

more efficient than H-L and almost 18% more efficient than L-S. At the extreme case, 30% contamination, ARE of Smoothing versus H-L is 1.1099 which indicates Smoothing is almost 11% more efficient than H-L, and almost 56% more efficient than L-S.

Table 5: SM relative to HL and LS, using $\widehat{\sigma}^2_M$ as a smoothing parameter

ϵ	ARE(SM,LS)	ARE(SM,HL)	ARE(HL,LS)
ϵ=0	0.9649	1.0013	0.9637
ϵ=0.10	1.1802	1.0103	1.1681
ϵ=0.20	1.4715	1.0693	1.3760
ϵ=0.30	1.5615	1.1099	1.4068

In Table 6, the relative efficiencies of Smoothing(SM) versus H-L and L-S compared by using Huber's τ as a smoothing parameter estimate. At $\epsilon = 0$, ARE of SM with respect to L-S is 0.9672 which is almost equal to H-L under no contamination case. L-S is the best at 0% contamination compared to H-L and SM. At the 10% contamination of the data, ARE of SM versus H-L is 1.0126 which indicates Smoothing is 1% more efficient than H-L and almost 18% more efficient than L-S. At the extreme case, 30% contamination, ARE of Smoothing versus H-L is 1.1160 which indicates Smoothing is 11% more efficient than H-L, and almost 57% more efficient than L-S.

Table 7 shows relative efficiencies of Smoothing(SM) versus H-L and L-S, by using $\widehat{\sigma}^2_S$ as a smoothing parameter estimate. At $\epsilon = 0$, ARE of SM with respect to L-S is almost 98% which is 2% greater than H-L under no contamination

Table 6: SM relative to HL and LS, using $\widehat{\sigma}_{T}^{2}$ as a smoothing parameter

ϵ	ARE(SM,LS)	ARE(SM,HL)	ARE(HL,LS)
ϵ=0	0.9672	1.0036	0.9637
ϵ=0.10	1.1829	1.0126	1.1681
ϵ=0.20	1.4806	1.0761	1.3760
ϵ=0.30	1.5701	1.1160	1.4068

case. But still L-S is the best at 0% contamination. At the 10% contamination of the data, ARE of SM versus H-L is 1.0211 which indicates Smoothing is 2% more efficient than H-L and almost 20% more efficient than L-S. At the extreme case, 30% contamination, ARE of Smoothing versus H-L is 1.0608 which indicates Smoothing is 6% more efficient than H-L, and almost 50% more efficient than L-S.

Table 7: SM relative to HL and LS, using $\widehat{\sigma}_{S}^{2}$ as a smoothing parameter

ϵ	ARE(SM,LS)	ARE(SM,HL)	ARE(HL,LS)
ϵ=0	0.9791	1.0160	0.9637
ϵ=0.10	1.1927	1.0211	1.1681
ϵ=0.20	1.4460	1.0508	1.3760
ϵ=0.30	1.4924	1.0608	1.4068

Table 8 shows relative efficiency of Kernel Density Estimation Method with respect to H-L and L-S by using optimal bandwidth procedure from Sheather and Jones(1991). The algorithm used in this simulation is available upon request from the authors.

Table 8: SM relative to HL and LS, using bandwidth b_n as a smoothing parameter

ϵ	ARE(SM,LS)	ARE(SM,HL)	ARE(HL,LS)
ϵ=0	0.8140	0.8414	0.9586
ϵ=0.10	1.0086	0.8664	1.1641
ϵ=0.20	1.4242	1.0259	1.3882
ϵ=0.30	1.6236	1.1274	1.4401

CHAPTER VIII

CONCLUSION

8.1 Concluding Remarks

In this study, we proposed a new estimation method for the shift parameter Δ_0 in the two-sample location problem. Our main purpose was to develop an alternative procedure to current parametric and nonparametric procedures that are widely used in the literature.

The parametric shift parameter estimation method, Least Squares and non-parametric estimation methods Hodges-Lehman and General Rank Scores have been presented in the Chapter I. Their asymptotic properties have been referred to Hettmasperger and McKean(1998). Their advantages and disadvantages have been pointed out. In Chapter II, the proposed method has been described and notation introduced. The sample Y is shifted by an arbitrary shift variable Δ and then sample X and Δ shifted Y sample are combined in a one sample. The purpose of our arbitrary Δ shift is to align two sample as closely as possible and find that value which aligns two sample. We can not use true shift parameter Δ_0 in place of Δ because it is unknown. Later the empirical distribution function $F_n^\star(x)$ of the combined shifted sample is convoluted with a smoother density function g(t) to find a smooth replacement for the true underlying density f(x). This new density is called $\widehat{h}_\Delta(t)$ which carries over the overall(approximate) shape of the true population density f(x) that the combined shifted sample comes from. It should be noted that even though proposed smoothing by convolution idea is analogous to Kernel Density estimation, we don't exactly estimate the density but we smooth the empirical distribution function with convolution idea to get the overall shape of the data. The smoothing parameter σ plays an important role in the smoothing by convolution idea. The smaller values of the σ can under smooth the data, the larger values of the σ can over smooth the data. The importance of σ is illustrated by figures in Chapter 2. Later $\widehat{h}_\Delta(t)$ replaces the unknown true density function f(x) in the log likelihood function $L(\Delta)$. After the replacement, $L(\Delta)$ is called $L_n(\Delta)$ which can be minimized with respect to Δ. The resulting solution $\widehat{\Delta}_S$ is estimate of the true shift parameter Δ_0. This estimator is a translation equivariant estimator and this is proved in the last part of the Chapter 2.

In Chapter III, some of the theoretical properties of the smoothed density function $\widehat{h}_\Delta(t)$ is presented. We discussed that $L_n(\Delta)$ is a quasiconvex function of Δ and minimum exists by Roberts and Varberg(1973). $L_n(\Delta)$ and its approximation $L_n^\star(\Delta)$ have been developed to support the asymptotic results. In Chapter IV, the Pitman regularity conditions for $S_n^\star(\Delta) = \frac{\partial L_(^\star \Delta)}{\partial \Delta}$ have been satisfied and the asymptotic linearity results of gradient function $S_n^\star(\Delta)$ has been shown.

In Chapter V, The asymptotic linearity of $S_n(\Delta)$ is proven by the Theorem 5.1.2. The influence function IF(x,y) derived by using the idea proposed by Huber(1982) in the one sample case. Based on the influence function and the asymptotic linearity results of $S_n(\Delta)$, the asymptotic normality result of the estimator $\widehat{\Delta}_S$ has been derived. In Chapter VI, we developed the asymptotic α level testing and confidence interval estimation procedure for the proposed method.

The examples and a bootstrap simulation study presented in the Chapter VII demonstrated that proposed method is as competitive as current methods, in most cases even better. By the simulation study, different choices of smoothing parameter σ have been investigated. The regular non-robust sample variance estimate $\widehat{\sigma}_S^2$, robust median absolute deviation(MAD) estimate $\widehat{\sigma}_M^2$ and robust Huber's Tau(τ) estimate $\widehat{\sigma}_\tau^2$ are compared and found that the proposed estimator works similarly in the different choices of σ^2 estimations. Surprisingly, non-robust sample variance estimation of σ^2 is very competitive with other two robust estimations of σ^2. Kernel Density estimation method is also used to estimate the unknown true density function of the combined sample. As an alternative to estimate smoothing parameter σ from regular scale estimation procedures, the bandwidth parameter proposed by Sheather and Jones(1991) is estimated and used in the smoothing by convolution idea. The result of the simulation is presented in Chapter VII.

By the finite sample asymptotic efficiency result, the proposed solution is approximately 98% efficient against classical method, the Least Square under the normality assumptions and far better than the Least Squares if underlying distribution is contaminated. Under the same conditions proposed method has almost 1%-10% more efficient than the Hodges-Lehman method. As a result, we can confirm that proposed method works as good as currently used methods and can be included in the literature of two-sample location problem.

8.2 Future Research

In the future, we will investigate some of properties of the smoothed density function $\widehat{h}_\Delta(t)$ and log-likelihood function $L_n(\Delta)$. We will try to develop conditions on the log concavity of $\widehat{h}_\Delta(t)$ with respect to smoothing parameter σ. The quasi-convexity of the $L_n(\Delta)$ will be investigated further and mathematical proofs will be sought. The different choices of the smoothing parameter σ will be considered and optimum smoothing parameter σ will be proposed.

From Chapter IV, the efficiency of the estimator $\Delta_S^\star$ will be investigated against the maximum likelihood estimator given that underlying distribution functions are known.

We would like to investigate the Behrens-Fisher problem by using the same smoothing by convolution idea. In the Behrens-Fisher problem, we have a case of different scale parameters for each population distribution functions. If we recall that in regular two-sample location problem, we assumed that scale parameters are equal. By this assumption, we guaranteed that shapes of the two distributions are identical but locations are different. In fact, the scale parameters are different in Behrens-Fisher problem. Therefore, the shapes of the underlying distribution functions are not identical any more. In this case, we propose to smooth each empirical density functions of the samples separately without combining the two samples. We will have two different smoothing parameter σ_x and σ_y for each convolution smoothing. It will be a very complicated problem to work on.

REFERENCES

Antony, M. (2002) Uniform Glivenko-Cantelli Theorems and Concentration of Measure in Mathematical Modelling of Learning, *CDAM Research Report LSE-CDAM-2002-07*, London School of Economics.

Bohn, L. L. and Wolfe, D.A.(1992) Nonparametric two-sample procedures for ranked-set samples data, *Journal of American Statistical Association*, 87, 552-561.

Cencov, N.N.(1962) Evaluation of an Unknown Density from Observatioons, *Soviet Mathematics*, 3, 1559-1562.

Doksum, K.(1974). Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case, *The Annals of Statistics*, 2, 2767-277.

Hall, P. and Marron, J.S. (1987) Estimation of integrated squared density derivatives. *Statist. Probability Letters*, 6, 109-115.

Hettmansperger, T.P.(1984) *Statistical Inference Based on Ranks*, New York: John Wiley and Sons.

Hettmansperger, T.P. and McKean, J.W (1998) *Robust Nonparametric Statistical Methods*, New York: John Wiley and Sons.

Hodges, J.L.,and Lehmann, E.L. (1963) Estimates of location based on rank tests,*Annals of Mathematical Statistics*, 34, 598-611.

Hollander, M. and Wolfe, D.A. (1999) *Nonparametric Statistical Methods*, John Wiley and Sons.

Hoyland, A.(1965) Robustness of the Hodges-Lehmann estimates for shift,*Annals of Mathematical Statistics*, 36, 174-197.

Huber, P.J.(1964) Robust Estimation of a Location Parameter, *Annals of Mathematical Statistics*, 35, 73-101.

Huber, P.J.(1972) The 1972 Wald Lecture Robust Statistics: A Review,*Annals of Mathematical Statistics*, 43, 1041-1067.

Huber, P.J.(1981) *Robust Statistics*, New York: John Wiley and Sons.

Jaeckel, L. A. (1972) Estimating regression coefficients by minimizing the dispersion of the residuals, *Annals of Mathemetical Statistics*, 43, 1449-1458.

Kronmal, R.A and Tarter, M.E.(1968) The Estimation of Probability Densities and Cumulatives by Fourier Series Methods, *Journal of American Statistical Association*, 63, 925-952.

Lehmann,E.L (1985)*Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden-Day.

Mann, H. B. and Whitney, D. R. (1947) On a test of whether one of the two random variables is stocastically larger than the other, *Annals of Mathematical Statistics*, 18, 50-60.

McKean, J. W., Vidmar, T.J., and Sievers, G.L. (1989) A robust two stage multiple comparison procedure with application to a random drug screen, *Biometrics*, 45, 1281-1297.

Miravete, E. J. (2001) On preservation of Increasing Hazard Rate Under Convolution, *CARESS Working Paper 01-05*, University of Pennsylvania.

Park, B.U. and Marron, J.S. (1990) Comparision of data driven bandwidth selectors. *Journal of American Statistical Association*, 85, 66-72.

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

Ponstein, J. (1967) Seven Kinds of Convexity. *Society for Industrial and Applied Mathematics*, Volume 9, Issue 1, 115-119.

Roberts, A. W and Varberg D.E.(1973) *Convex Functions*, New York: Academik Press.

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

Scott W.D. (1992) *Multivariate Density Estimation*, New York: John Wiley and Sons

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, New york: John Wiley and Sons.

Sheather, S.J. and Jones, M.C. (1991) A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation, *Journal of the Royal Statistical Society. Seriess B(Methodological)*, Volume 53, Issue 3, 683-690.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Student. (1927) Errors of routine analysis, *Biometrika*, 19, 151-164.

Tukey, J.W. (1960) A survey of sampling from contaminated distributions, *Contributions to Probability and Statistics in Honor of Harold Hotelling*, Standford University Press, CA.

Watson, G.S.(1969) Density Estimation by Orthogonal Series, *Annals of Mathemetical Statistics*, 40, 1496-1498.

Wellner, J (2002) Stat-583 Advance Theory of Statistical Inference Lecture Notes, *http://stat.washington.edu/jaw/COURSES/580s/581/lectnotes.html*

Wilcoxon, F. (1945) Indivudual comparisons by ranking methods, *Biometrics*, 1, 80-83.