# TWO CONTRIBUTIONS TO ORDER RESTRICTED INFERENCES

A Dissertation by

Alexandra Kathleen Echart

Master of Science, Wichita State University, 2015

Bachelor of Science, Wichita State University, 2014

Submitted to the Department of Mathematics, Statistics, & Physics
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

December 2018

TWO CONTRIBUTIONS TO ORDER RESTRICTED INFERENCES

The following faculty members have examined the final copy of this dissertation for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy, with a major in Mathematics.

---

Hari Mukerjee, Committee Chair

---

Thomas Delillo, Committee Co-Chair

---

Dharam Chopra, Committee Member

---

Xiaomi Hu, Committee Member

---

Adam Jaegar, Committee Member

---

Rhonda Lewis, Committee Member

Accepted for the College of Liberal Arts and Sciences

---

Andrew Hippisley, Dean

Accepted for the Graduate School

---

Dennis Livesay, Dean

DEDICATION

To my parents, friends, and family

ACKNOWLEDGMENTS

I would like to thank my advisor, Hari Mukerjee, for his support throughout my undergraduate and graduate studies. I would also like to thank Hammou El Barmi for his programming guidance throughout my graduate career.

## ABSTRACT

We are proposing two separate problems from order restricted inferences. The first is a one-sided test for stochastic ordering of two distribution functions that protects against false positive conclusions because of model assumptions. In a traditional hypothesis test there is no protection against the fact that both the null and alternative hypotheses could be false. We modify the classical test to allow for any of the following multiple decisions to be made: (1) Decide the equality of distribution functions cannot be rejected, (2) Decide the distribution functions are ordered in one direction, (3) Decide the distribution functions are ordered in the opposite direction, and (4) Decide both (2) and (3) hold at the same time. Via simulations and examples we show that this procedure provides protection against false positive conclusions while reducing the power of the test minimally when the ordering is correct.

The second problem is an improved estimation of a decreasing density of a random variable. The standard Grenander (1956) nonparametric maximum likelihood estimator of a decreasing density has an $n^{-1/3}$ convergence rate with an unfamiliar asymptotic distribution whereas nonparametric kernel estimators have an $n^{-2/5}$ convergence rate with a normal asymptotic distribution. We propose a hybrid estimator that will utilize both concepts of estimation and guarantees monotonicity of the estimator. The hybrid estimator will also substantially improve upon past estimators of $f(0)$. We show this analytically through simulations.

TABLE OF CONTENTS


| Chapter | Page |
| --- | --- |
| 1. BACKGROUND OF THE MULTIPLE DECISION PROCEDURE FOR MEANS | 1 |
| 2. BACKGROUND OF ESTIMATION OF DISTRIBUTION FUNCTIONS UNDER STOCHASTIC ORDERING | 8 |
| 3. INTRODUCTION OF THE MULTIPLE DECISION PROCEDURE FOR STOCHASTIC ORDERING AMONG TWO DISTRIBUTION FUNCTIONS | 17 |
| 4. SOME SIMULATIONS IN SUPPORT OF THE MULTIPLE DECISION PROCEDURE | 25 |
| 5. AN EXAMPLE FOR THE MULTIPLE DECISION PROCEDURE | 32 |
| 6. SUMMARY AND CONCLUDING REMARKS FOR THE MULTIPLE DECISION PROCEDURE | 33 |
| 7. INTRODUCTION TO AN IMPROVED ESTIMATION OF A DECREASING DENSITY OF A RANDOM VARIABLE | 34 |
| 8. ASYMPTOTIC EQUIVALENCE WITH A KERNEL ESTIMATOR | 41 |
| 9. ESTIMATION OF $f(0)$ | 49 |
| 10. ASYMPTOTIC DISTRIBUTION OF $f_n^*(0)$ | 52 |
| 11. SIMULATIONS AND GRAPHS | 53 |
| 12. CONCLUDING REMARKS AND FUTURE WORK | 61 |
| BIBLIOGRAPHY | 62 |
| APPENDICES | 68 |

LIST OF TABLES

Table Page

2.1 CRITICAL VALUES FOR THE EMPIRICAL LIKELIHOOD RATIO TEST 15

2.2 CORRECTED TABLE 2 FROM EL-BARMI AND MCKEAGUE(2013) 16

4.1 $F_1 = U(0,1)$ VS $F_2 = x^2$ WITH $\alpha_1 = \alpha_2 = 0.01$ AND $n_1 = n_2 = 100$ 27

4.2 $F_1 = U(0,1)$ VS $F_2 = x^2$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 50$ 27

4.3 $F_1 = U(0,1)$ VS $F_2 = x^2$ WITH $\alpha_1 = \alpha_2 = 0.10$ AND $n_1 = n_2 = 50$ 27

4.4 $F_1 = 4(x-.5)^3 + .5$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.01$ AND $n_1 = n_2 = 150$ 28

4.5 $F_1 = 4(x-.5)^3 + .5$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 150$ 28

4.6 $F_1 = 4(x-.5)^3 + .5$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.10$ AND $n_1 = n_2 = 200$ 28

4.7 $F_1 = \frac{16}{9}(x-.75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 100$ 30

4.8 $F_1 = \frac{16}{9}(x-.75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 150$ 30

4.9 $F_1 = \frac{16}{9}(x-.75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 200$ 30

4.10 $F_1 = \frac{16}{9}(x-.75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 0.05$ AND $\alpha_2 = 0.10$ AND $n_1 = n_2 = 100$ 30

4.11 $F_1 = \frac{16}{9}(x-.75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 0.05$ AND $\alpha_2 = 0.10$ AND $n_1 = n_2 = 150$ 30

4.12 $F_1 = \frac{16}{9}(x-.75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 0.05$ AND $\alpha_2 = 0.10$ AND $n_1 = n_2 = 200$ 31

11.1 $f_n^*(0+)/f(0+)$ FOR THE STANDARD EXPONENTIAL CASE; BANDWIDTH FIXED 53

11.2 $f_n^*(0+)/f(0+)$ FOR THE STANDARD EXPONENTIAL CASE; BANDWIDTH SIMULATED 54

11.3 $f_n^*(0+)/f(0+)$ FOR THE HALF-NORMAL CASE; BANDWIDTH FIXED 54

11.4 $f_n^*(0+)/f(0+)$ FOR THE HALF-NORMAL CASE; BANDWIDTH SIMULATED 54

11.5 THE HALF-NORMAL CASE 55

LIST OF TABLES (continued)

Table Page

11.6 THE STANDARD EXPONENTIAL CASE 55

11.7 THE HALF-NORMAL $E[f_n^*(x) - f(x)]^2$ WITH A SAMPLE OF SIZE 200 AND $h_n = .35$ 59

11.8 THE STANDARD EXPONENTIAL $E[f_n^*(x) - f(x)]^2$ WITH A SAMPLE OF SIZE 200 AND $h_n = .35$ 60

## LIST OF FIGURES


| Figure | | Page |
|---|---|---|
| 4.1 | COMPARISON OF $F_1 = U(0,1)$ TO $F_2 = x^2$ | 26 |
| 4.2 | COMPARISON OF $F_1 = 4(x-.5)^3 + .5$ TO $F_2 = U(0,1)$ | 28 |
| 4.3 | COMPARISON OF $F_1 = \frac{16}{9}(x-.75)^3 + .75$ TO $F_2 = U(0,1)$ | 29 |
| 11.1 | EXPONENTIAL DISTRIBUTION, N = 200 | 55 |
| 11.2 | EXPONENTIAL DISTRIBUTION, N = 200 | 56 |
| 11.3 | HALF-NORMAL DISTRIBUTION, N = 200 | 56 |
| 11.4 | HALF-NORMAL DISTRIBUTION, N = 200 | 57 |
| 11.5 | EXPONENTIAL DISTRIBUTION, N = 1000 | 57 |
| 11.6 | EXPONENTIAL DISTRIBUTION, N = 1000 | 58 |
| 11.7 | HALF-NORMAL DISTRIBUTION, N = 1000 | 58 |
| 11.8 | HALF-NORMAL DISTRIBUTION, N = 1000 | 59 |

# CHAPTER 1

# BACKGROUND OF THE MULTIPLE DECISION PROCEDURE FOR MEANS

An important problem in order restricted inference (referred to as ORI) is to establish that a $k$-dimensional location parameter $\boldsymbol{\mu} = \langle \mu_1, \mu_2, \ldots, \mu_k \rangle$ lies within a closed convex cone (CCC) $C$. The standard process is to conduct the following hypothesis test

$$H_0 : \boldsymbol{\mu} \in L \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \in C - L, \tag{1.1}$$

where $L$ is the largest linear subspace of $C$. Perhaps the most common application arises in a one-way analysis of variance; for independent random samples from $k$ populations, with mean $\mu_i$ for the $i$th population, and letting

$$L = \{\boldsymbol{\mu} : \mu_1 = \mu_2 = \cdots \mu_k\} \quad \text{and} \quad C = \{\boldsymbol{\mu} : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k\} \tag{1.2}$$

we test as in (1.1). Some of the first to investigate and develop a nonparametric test for this nondecreasing ordered alternative, alternatively given by

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{versus} \quad H_1 : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \tag{1.3}$$

where at least one of the inequalities is strict, were Jonckheere (1954) and Terpstra (1952). We could also consider the case where the $\mu_i$'s are decreasing as an alternative to $H_1$. Chacko (1963), Cuzick (1985) and Le (1988) each proposed tests to handle the same problem, but Mahrer and Magel (1995) showed that the Jonckheere and Terpstra test (referred to in the literature at the JT test) was comparable in terms of power to both Cuzick's test and Le's test. Later Mack and Wolfe (1981) extended the JT test to test for umbrella orderings given by

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{versus} \quad H_1 : \mu_1 \leq \cdots \leq \mu_{l-1} \leq \mu_1 \geq \mu_{l+1} \geq \cdots \geq \mu_k \tag{1.4}$$

where $\mu_i \neq \mu_j$ for some $i \neq j$. They considered both the cases where the turning point is known and unknown. Many then proposed nonparametric tests for the simple tree alternative given by

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \quad \text{versus} \quad H_1 : \mu_1 \leq \mu_i \forall i \tag{1.5}$$

where at least one of these inequalities is strict. A few of those who have made contributions to the simple tree alternative include Fligner and Wolfe (1990), Magel (1988), and Desu *et al.* (1996).

Under the usual assumption of independent identically distributed (IID) $N(0, \sigma^2)$ errors, Bartholomew (1961 a,b) developed the likelihood ratio test (LRT) for the nondecreasing ordering in (1.1). The text by Robertson, Wright, and Dykstra (1988), referred to as RWD (1988), contains an account of it and many other developments. The procedure is outlined here. Let $n_i$ denote the sample size for the $i$th population with $N = \sum_i n_i$ the total sample size. Also let $\{Y_{ij} : 1 \leq i \leq k, 1 \leq j \leq n_i\}$ be the collection of observations for the $k$ populations where the population mean vector is $\widehat{\mu} = \{\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_k\}$, the maximum likelihood estimator, or MLE, of $\mu$ when no restrictions are placed on $\mu$. Under the restriction $\mu \in C$, the MLE $\mu^*$ can be shown to be the least squares projection of $\widehat{\mu}$ onto $C$, denoted by $\Pi_{\boldsymbol{w}}(\widehat{\mu}|C)$. that minimizes

$$\sum_{i=1}^{k} w_i(\widehat{\mu}_i - f_i)^2 \quad \text{in the class of } f = \{f_i\} \in C, \tag{1.6}$$

where $\boldsymbol{w} = \langle w_1, \ldots, w_k \rangle$ is a weight vector such that $w_i = n_i/\sigma^2$ for $1 \leq i \leq k$.

The popular pool adjacent violator algorithm (PAVA) could be used to compute $\mu^*$; this algorithm is described in RWD (1988). The $\widehat{\mu}_i$'s have $k$ distinct values with probability one and the restricted estimator will average a number of adjacent $\widehat{\mu}_i$'s giving rise to $M$ level sets; the number of level sets $M$ is random with $1 \leq M \leq k$ since $C$ is a polyhedral cone, and the least squares projection of $\widehat{\mu}$ could be in any one of the $l$-dimensional faces of $C$ where $l \leq l \leq k$. Let $\overline{\mu}_0 = \sum_{i=1}^{k} w_i \overline{Y}_i / \sum_{i=1}^{k} w_i$ be the estimate of the common mean under $H_0$. When the variance $\sigma^2$ is known, the likelihood ratio test rejects $H_0$ for large values of the test statistic $\overline{\chi}_{01}^2$ defined as

$$\overline{\chi}_{01}^2 = \sum_{i=1}^{k} w_i(\mu_i^* - \overline{\mu}_0)^2. \tag{1.7}$$

The null distribution of the test statistic is given by

$$P(\overline{\chi}_{01}^2 \geq c) = \sum_{l=1}^{k} P(l, k, w) P(\chi_{l-1}^2 \geq c), \tag{1.8}$$

where $P(l, k, w) = P(\text{number of level sets } M = l)$, $\chi_{l-1}^2$ is a central $\chi^2$ random variable with $l-1$ degrees of freedom, and $\chi_0^2 \equiv 0$. The $P(l, k, w)$'s are probabilities of certain wedges (the pre-images of sectors of $\mathbb{R}^k$ that project onto the $l$-dimensional faces of $C$) of $\mathbb{R}^k$ under the $N_k(0, W^{-1})$ distribution, where $W^{-1} = \text{diag}\{w_1^{-1}, w_2^{-1}, ..., w_k^{-1}\}$. It should be noted that the $P(l, k, w)$'s depend on $w$ only through direction and not magnitude. For some simple cases these can be computed exactly, and they have been tabulated in RWD (1988). Many approximation schemes have been developed and these probabilities are generally found via simulations.

When $\sigma^2$ is unknown, we estimate it as in ANOVA via the equation:

$$\widehat{\sigma}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / (N - k). \tag{1.9}$$

One rejects $H_0$ for large values of the test statistic

$$\bar{E}_{01}^2 = \frac{\sum_{i=1}^{k} w_i(\mu_i^* - \overline{\mu}_0)^2}{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{\mu}_0)^2}. \tag{1.10}$$

Its null distribution is given by

$$P(\bar{E}_{01}^2 \geq c) = \sum_{l=1}^{k} P(l,k,w) P(B_{(l-1)/2,(N-l)/2} \geq c), \tag{1.11}$$

where $B_{a,b}$ is a beta variable with parameters $a$ and $b$ ($B_{0,b} = 0$). It can be shown that a transformed version is asymptotically $\bar{\chi}_{01}^2$:

$$S_{01} \equiv \frac{(N-k)\bar{E}_{01}^2}{1-\bar{E}_{01}^2} \xrightarrow{d} \bar{\chi}_{01}^2 \quad \text{as} \quad N-k \to \infty. \tag{1.12}$$

The discussion above is an abbreviated description of the test in (1.1): $\mu \in H_0$ versus $\mu \in H_1 - H_0$ in the usual ANOVA setting under the normality assumption. Notice that we will use $H_{()}$ both as a symbol for a hypothesis as well as a symbol for the set $\mu$ is in under $H_{()}$; the possible confusion is minimal.

For all of the tests mentioned above, an *a priori* knowledge about the ordering is assumed. What if when testing for the nondecreasing ordering in (1.1), the location parameters differ in a manner not under the alternative? Terpstra and Magel (2002) consider the case when $k = 4$ populations and the location parameters are $\mu_1 = 0$, $\mu_2 = 1.5$, $\mu_3 = 0.5$, and $\mu_4 = 1.5$. A simulation with normally distributed data shows that the power of a JT test is 0.932 ($n = 20$ for all populations). These authors point out that this could lead to devastating consequences in many applications. They stated that a good test should have the following three properties:

P1. The power of the test should be approximately equal to the stated significance level when $H_0$ is true,

P2. The test should have a higher power than the general alternative test when $H_1$ is true, and

P3. The test should have low power for any alternative that does not fit the profile given in $H_1$.

Terpstra and Magel (2003) then propose their own test statistic

$$T = \sum_{i_1=1}^{n_1} \cdots \sum_{i_k=1}^{n_k} [I(Y_{1i_1} \leq \cdots \leq Y_{ki_k}) - I[(Y_{1i_1} = \cdots = Y_{ki_k})]. \tag{1.13}$$

This test statistic measures the number of times the population-permuted observations follow the order restriction of $\mu \in H_1 - H_0$. There are $N!/\prod_{i=1}^{k} n_i!$ partitions of (1,...N) of this form, and each is equally likely under the null hypothesis. Although the exact distribution of the Terpstra-Magel test statistic can be found in principle, the computation is not realistically possible even for small numbers. For instance, the number of partitions is approximately $1.7 \times 10^{10}$ for four populations and $n_i \equiv 5$. However, they were able to derive the asymptotic distribution. Additionally, it is possible to perform a Monte Carlo approximation of the exact distribution by taking a random sample of the partitions, but the computations will be computationally intensive. They called their test the new nonparametric test (NNT). Let $\text{NNT}_a$ denote the test based on the asymptotic results and $\text{NNT}_s$ based on simulations (but only for very small numbers). Their main result was that both NNT's lose little power compared to the JT test when the alternatives are in $C - L$, but lose a large percentage of the power when the alternatives are in $C^C$. In particular, for the extreme example, $\boldsymbol{\mu} = \langle 0, 1.5, 0.5, 1.5 \rangle$, that Terpstra and Magel considered, the JT test rejects $H_0$ 93.2% of the time while the $\text{NNT}_a$ rejects only 53.9% of the time. Using many alternatives, both in $C - L$ and in $C^C$, they show that their new test is clearly favorable over the JT test if one wants some protection against "false positives." However, this was done only for $k = 3$ and $k = 4$ because of computational complexity.

We have introduced an entirely new approach to protection against the false conclusion of $H_1 - H_0$ in the normal error model. Many times it is possible to perform a goodness of fit likelihood ratio test of $H_1$ versus $H_2 - H_1$, where $H_2$ puts no restriction on $\mu$, i.e., $\mu \in \mathbb{R}^k$. Suppose we reject $H_0$ in favor of $H_1 - H_0$ if the likelihood ratio test statistic $T_{01}$ is large, and we reject $H_1$ in favor of $H_2 - H_1$ if the likelihood ratio test statistic $T_{12}$ is large. We suggested a multiple decision procedure whereby we could make three possible decisions:

$D_0$: Conclude that there is no strong evidence against $H_0$ if $T_{01}$ is small and $T_{12}$ is small.

$D_1$: Conclude against $H_0$ and in favor of $H_1 - H_0$ if $T_{01}$ is large and $T_{12}$ is small.

$D_2$: Conclude against $H_1$ in favor of $H_2 - H_1$ if $T_{12}$ is large, suggesting possible further multiple pairwise comparisons.

Although the multiple decision procedure uses two standard tests of the ordinary type, it differs from a usual test in several ways.

1. If $\alpha_1$ and $\alpha_2$ are the levels of significance for the $T_{01}$ and $T_{12}$ tests, respectively, they are chosen independently, e.g., a large $\alpha_2$ will be called for if avoidance of a false $D_1$ is very important.

2. The level of significance $= \sup_{H_0} P(\text{Reject } H_0 | H_0 \text{ true})$ is of no interest, although one could define an entity, $\sup_{H_0} P(\text{Conclude } D_1 | H_0 \text{ true})$.

3. The power of the multiple decision procedure under various alternatives does not seem to make any sense.

4. As a related matter, the $p$-value does not make any sense either, although, given the observed $T_{01} = c_1$ and $T_{12} = c_2$, one could define a bivariate $p$-value as

$$(p_1, p_2) = (\sup_{H_0}(P(T_{01} \geq c_1 | H_0), \sup_{H_1} P(T_{12} > c_2 | H_1)). \tag{1.14}$$

Our ability to conclude in favor of $H_1 - H_0$ gets stronger as $p_1$ decreases and $p_2$ increases (easier to reject $H_1$ in favor of $H_2 - H_1$) and the maximum value of $p_2$ occurs when $c_2 = 0$, i.e., $\hat{\mu} = \mu^*$. Our ability to conclude against $H_1 - H_0$ gets stronger as $p_1$ increases and/or

$p_2$ decreases. Some simulation results showed how the probability of concluding $H_1 - H_0$ when $\mu \in H_1 - H_0$ is reduced very little from the likelihood ratio test (as described in (1.7)-(1.12)) and how the probability of concluding $H_1 - H_0$ is reduced very substantially when $\mu \in H_2 - H_1$. In particular, for the example $\boldsymbol{\mu} = \langle 0, 1.5, 0.5, 1.5 \rangle$, the likelihood ratio test concludes $D_1$ 99.8% of the time, while the multiple decision procedure concludes $D_1$ only (47.0%,24.6%,15.7%) of the time when $\alpha_2 = (0.01, 0.05, 0.10)$ with $\alpha_1$ held fixed at 0.05.

# CHAPTER 2

# BACKGROUND OF ESTIMATION OF DISTRIBUTION FUNCTIONS UNDER STOCHASTIC ORDERING

In 1955, Lehmann developed the concept of stochastic ordering of distribution functions. If $F_1$ is the distribution function of a random variable $X_1$ and $F_2$ is the distribution function of a random variable $X_2$, then he defined $X_1$ to be stochastically larger than $X_2$ when $F_1(x) \leq F_2(x)$ for all $x$. We denote this stochastic ordering as $F_1 \succ F_2$. The definition can be extended to $k$ distributions, so that $F_1 \succ F_2 \succ \cdots \succ F_k$ if $F_1(x) \leq F_2(x) \leq \cdots \leq F_k(x)$ for all $x$. Orderings of these types arise in health sciences and reliability engineering. For example in the health professions, if $F_1$ models the treatment effect and $F_2$ models the control group response, then determining whether a treatment has been beneficial is a matter of testing for a stochastic ordering between the two distribution functions. Consider the null hypothesis showing the equality of two distribution functions

$$H_0 : F_1 = F_2 \tag{2.1}$$

and the two hypotheses of ordering the two distribution functions

$$H_1 : F_1 \geq F_2 \tag{2.2}$$

$$H_2 : F_1 \leq F_2. \tag{2.3}$$

First, let us discuss asymptotic tests of $H_0$ versus $H_1 - H_0$. For the case of two distribution functions, two of the most widely used tests are the Kolmogorov-Smirnov test and the Mann-Whitney-Wilcoxon tests.

Let $X_{i1}, X_{i2}, \ldots, X_{in_i}$, for $i = 1, 2$ be independent random samples from distributions $F_1$ with sample size $n_1$ and $F_2$ with sample size $n_2$. Let $\widehat{F}_i$ denote the empirical cumulative distribution function based off of the random sample of size $n_i$, that is

$$\widehat{F}_i(x) = \frac{\#(X_{ij} \leq x)}{n_i}, \text{ for } 1 \leq j \leq n_i,\ i = 1, 2. \tag{2.4}$$

Then for each $x$, define the vector of empirical estimates $\widehat{\mathbf{F}}(x) = \left\langle \widehat{F}_1(x), \widehat{F}_2(x) \right\rangle$ and define the vector of weights $\mathbf{n} = \langle n_1, n_2 \rangle$. Under the Kolmogorov-Smirnov test, the test statistics for the two-sided and one-sided hypothesis tests can be determined as follows:

$$D = \max_x |\widehat{F}_1(x) - \widehat{F}_2(x)|, \tag{2.5}$$

$$D^+ = \max_x \left[\widehat{F}_1(x) - \widehat{F}_2(x)\right], \tag{2.6}$$

$$D^- = \max_x \left[\widehat{F}_2(x) - \widehat{F}_1(x)\right], \tag{2.7}$$

where $D^+$ corresponds to the one-sided test with the alternative hypothesis that $F_1$ is greater than $F_2$ for at least one value of $x$ and $D^-$ is the test statistic for the one-sided test with the alternative hypothesis that $F_2$ exceeds $F_1$ for at least one value of $x$. If the two samples were drawn from identical populations, then the empirical distributions should be nearly equal for all values of $x$ and the values of the test statistics $D, D^+$, and $D^-$ would be close to zero. The test statistics are measurements of the extent to which the empirical distribution functions fail to agree. Should the test statistic values be sufficiently small, there is not enough evidence to suggest the distributions are not from identical distributions. Should the test statistic values be sufficiently large, there is evidence to suggest that the distributions are different; depending on the test conducted possibly one is greater than the other. For equal sample sizes, quantiles for the one-sided Kolmogorov-Smirnov test have been calculated and tabulated for small choices of $n$. For $\alpha = (.10, .05, .01)$ and large choices of $n$, we will use the quantile approximations

$$t = \left(\frac{1.52}{\sqrt{n}}, \frac{1.73}{\sqrt{n}}, \frac{2.15}{\sqrt{n}}\right). \tag{2.8}$$

While the Kolmogorov-Smirnov test has long been used because it is sensitive to changes in location and shape of the two distributions, many authors have still sought to make improvements to the two distribution case. Lee and Wolfe (1976) introduced a variant of the Mann-Whitney-Wilcoxon test based on the non-parametric maximum likelihood estimators (NPMLEs) of the two distribution functions. El Barmi and Mukerjee (2005) proposed an improvement upon a Kolmogorov-Smirnov type test. El Barmi and McKeague (2013) developed an empirical likelihood based test for the stochastic ordering problem in which the asymptotic null distribution of their test statistic was found to have a simple distribution free representation in terms of standard Brownian bridge processes. Other tests were developed by Robertson and Wright (1981), Dykstra, Madsen and Fairbanks (1983), Franck (1984), Mau (1988), and Bohn and Wolfe (1992).

El Barmi and Mukerjee (2005) recommend the following procedure for testing $H_0$ versus $H_1 - H_0$ for stochastic ordering among $k$ populations. Define

$$N_{rs} = \sum_{j=r}^{s} n_j \tag{2.9}$$

so that

$$Av_{\mathbf{n}}\left[\widehat{\mathbf{F}}; r, s\right] = \sum_{j=r}^{s} \frac{n_j \widehat{F}_j}{N_{rs}} \tag{2.10}$$

is defined, where $\mathbf{n}$ is the vector comprised of the sample sizes and $\widehat{\mathbf{F}}$ is the vector of empirical distribution functions. As we are only interested in the case of two distribution functions we can simplify the notation substantially; $N_{11} = n_1$ and $N_{22} = n_2$ so we will reference them simply by their respective sample sizes, $n_i$, and $N_{12} = n_1 + n_2$ which we define to be $n$ the total sample size. Then $Av_{\mathbf{n}}\left[\widehat{\mathbf{F}}; 1, 1\right] = \widehat{F}_1$, the empirical distribution function of the first sample, $Av_{\mathbf{n}}\left[\widehat{\mathbf{F}};2,2\right]=\widehat{F}_2$, the empirical distribution function of the second sample, and

$$Av_{\mathbf{n}}\left[\widehat{\mathbf{F}};1,2\right]=\frac{n_1\widehat{F}_1+n_2\widehat{F}_2}{n}, \tag{2.11}$$

the weighted average of the empirical distribution functions. Define $\lambda_n=\frac{n_1}{n}$ and assume that $\lambda_n\to\lambda>0$, that is the proportion of the observations in the first sample remains fixed as the total sample size grows without bound. Also $1-\lambda_n\to1-\lambda$ where $1-\lambda_n=\frac{n_2}{n}$. For each $i=1,2$, let us define

$$Z_{in_i}=\sqrt{n_i}\left[\widehat{F}_i-F_i\right]. \tag{2.12}$$

From Billingsley (1968), we know that

$$\langle Z_{1n_1},Z_{2n_2}\rangle \overset{w}{\Longrightarrow} \langle Z_1,Z_2\rangle, \tag{2.13}$$

that is converges weakly to a bi-variate Gaussian process with independent components where each $Z_i\overset{d}{=}B_i\left(F_i\right)$ (that is equal in distribution to) where $B_1$ and $B_2$ are independent standard Brownian bridges. Now we may define

$$\widetilde{Z}_{1n_1}=\sqrt{n}\left[\widehat{F}_1-F_1\right]=\frac{Z_{1n_1}}{\sqrt{\lambda_n}}, \tag{2.14}$$

$$\widetilde{Z}_{2n_2}=\sqrt{n}\left[\widehat{F}_2-F_2\right]=\frac{Z_{2n_2}}{\sqrt{1-\lambda_n}}, \tag{2.15}$$

$$\widetilde{Z}_1=\frac{Z_1}{\sqrt{\lambda}}, \tag{2.16}$$

and

$$\widetilde{Z}_2=\frac{Z_2}{\sqrt{1-\lambda}}. \tag{2.17}$$

Then $\left\langle\widetilde{Z}_{1n_1},\widetilde{Z}_{2n_2}\right\rangle$ must converge weakly to $\left\langle\widetilde{Z}_1,\widetilde{Z}_2\right\rangle$ by applying (2.13).

For the $k$ sample case, El Barmi and Mukerjee (2005) define the following notations. Let $a_{jn} = n_j / \sum_{i=1}^{k} n_i$ and $a_j = \lim_n a_{jn} > 0$. Then let $A_{jn} = \sum_{i=1}^{j} a_{in}$ for $1 \leq j \leq k$. Notice that the $a_{jn}$ correspond with our notions of $\lambda_n$ and $1 - \lambda_n$ for the two sample case. Now let $\Lambda_{1n} = \lambda_n$ and $\Lambda_{2n} = \lambda_n + (1 - \lambda_n) = 1$ to correspond with the $A_{jn}$. Then

$$c_{jn} = \frac{a_{jn} A_{j-1,n}}{A_{jn}} \tag{2.18}$$

for all $j \geq 2$. In the case of two distributions, we only have one such $c_{jn}$ so we will suppress the dependence on $j$ and define

$$c_n = \frac{(1 - \lambda_n)\,\Lambda_{1n}}{\Lambda_{2n}} = (1 - \lambda_n) \cdot \lambda_n = \frac{n_2}{n} \cdot \frac{n_1}{n} = \frac{n_1 n_2}{n^2}. \tag{2.19}$$

The test statistic for the $k$ sample case is then defined as follows:

$$T_n = \max_{2 \leq j \leq k} \sup_x T_{jn}(x), \tag{2.20}$$

where

$$T_{jn} = \sqrt{n}\sqrt{c_{jn}} \left[ \widehat{F}_j - Av_{\mathbf{n}} \left[ \widehat{\mathbf{F}}; 1, j-1 \right] \right] \tag{2.21}$$

and $n$ is the total sample size of the $k$ populations. For the test of two distribution functions, the test statistic $T_n$ becomes

$$S_n = \sup_x S_{2n}(x) \tag{2.22}$$

where

$$S_{2n} = \sqrt{n}\sqrt{c_n} \left[ \widehat{F}_2 - Av_{\mathbf{n}} \left[ \widehat{\mathbf{F}}; 1, 1 \right] \right] = \sqrt{\frac{n_1 n_2}{n}} \left[ \hat{F}_2 - \hat{F}_1 \right]. \tag{2.23}$$

We reject the null hypothesis when the value of $S_n$ is sufficiently large. Values for which we can reject the null hypothesis were determined as follows.

First, let $F$ denote the cumulative distribution function of the pooled sample under the assertion that the null hypothesis is true. Then,

$$S_{2n} = \sqrt{c_n}\left[\widetilde{Z}_{2n_2} - Av_{\mathbf{n}}\left[\widetilde{\mathbf{Z}}_n; 1, 1\right]\right], \tag{2.24}$$

where $\widetilde{\mathbf{Z}}_n = \langle \widetilde{Z}_{1n_1}, \widetilde{Z}_{2n_2} \rangle$ and $\widetilde{Z}_{in_i}$ are defined as in (2.14) and (2.15) for $i = 1, 2$ respectively. By (2.13) and the arguments that follow, $S_{2n}$ converges weakly to $S_2$ where

$$S_2 = \sqrt{c}\left[\widetilde{Z}_2 - Av_{\boldsymbol{\lambda}}\left[\widetilde{\boldsymbol{Z}}; 1, 1\right]\right], \tag{2.25}$$

where $\boldsymbol{\lambda} = \langle \lambda_n, 1 - \lambda_n \rangle$, $\widetilde{Z} = \langle \widetilde{Z}_1, \widetilde{Z}_2 \rangle$, and $c = \lim_n c_n$, that is equal to $\lambda\Lambda_1/\Lambda_2$ where $\Lambda_i$ are defined as before. Now $S_2 \overset{d}{=} B_2(F)$ where $B_2$ is a Brownian Bridge. Then the test statistic $S_n$ converges in distribution to $S$, where $S = \sup\limits_x S_2(x)$. From Billingsley (1968), the distribution of $S$ is given by

$$\begin{aligned} P(S \geq s) &= 1 - P\left(\sup_x S_2(x) < s\right) \\ &= 1 - P\left(\sup_x B_2(F(x)) < s\right) \\ &= 1 - \left(1 - e^{-2s^2}\right) \\ &\approx e^{-2s^2}. \end{aligned} \tag{2.26}$$

To test $H_0$ versus $H_1 - H_0$, El Barmi and McKeague (2013) propose an alternative test statistic. Assume still that the proportion of observations in each sample remains fixed as the total sample size grows, with the proportions bounded between 0 and 1. Consider the following localized empirical likelihood function:

$$\mathcal{R}(x) = \frac{\sup\left\{\prod_{j=1}^{2} L(F_j) : F_1(x) = F_2(x)\right\}}{\sup\left\{\prod_{j=1}^{2} L(F_j) : F_1(x) \leq F_2(x)\right\}}, \tag{2.27}$$

where $L(F_i)$ is the nonparametric likelihood function and by convention $\sup \emptyset = 0$ and $0/0 = 1$. They have shown that it suffices to maximize

$$\prod_{j=1}^{2} \phi_j^{n_j \widehat{F}_j(x)} \left[1 - \phi_j\right]^{n_j \left(1 - \widehat{F}_j(x)\right)}, \tag{2.28}$$

subject to either $0 < \phi_1 = \phi_2 < 1$ or $0 < \phi_1 \leq \phi_2 < 1$. Under the first constraint, $0 < \phi_1 = \phi_2 < 1$, (2.28) is maximized by $\phi_j = \widehat{F}(x)$, where $\widehat{F}$ is the empirical cumulative distribution function of the pooled sample. Under the second constraint, $0 < \phi_1 \leq \phi_2 < 1$, (2.28) is maximized by the weighted least squares projection of $\widehat{\boldsymbol{\phi}} = \left\langle \widehat{F}_1(x), \widehat{F}_2(x) \right\rangle$ onto $\mathcal{I} = \{\mathbf{x} \in \mathbb{R}^2 : x_1 \leq x_2\}$ with weights $\boldsymbol{\lambda}$. Denote this projection as $E_{\boldsymbol{\lambda}}\left(\widehat{\boldsymbol{\phi}} | \mathcal{I}\right)_j = \widetilde{F}_j(x)$ for $j = 1, 2$. One algorithm that has been developed to compute this projection is the pool-adjacent-violators algorithm (Robertson *et al*). We can now express the localized empirical likelihood function as follows:

$$\mathcal{R}(x) = \prod_{j=1}^{2} \left[\frac{\widehat{F}(x)}{\widetilde{F}_j(x)}\right]^{n_j \widehat{F}_j(x)} \left[\frac{1 - \widehat{F}(x)}{1 - \widetilde{F}_j(x)}\right]^{n_j \left(1 - \widehat{F}_j(x)\right)} \tag{2.29}$$

where we will follow the usual convention that any term raised to a zero power is defined to be one. In order to test the null hypothesis $H_0$ against the alternative $H_1 - H_0$, they proposed the test statistic

$$T_n = -2 \int_{-\infty}^{\infty} \log \mathcal{R}(x) \; d\widehat{F}(x). \tag{2.30}$$

whose asymptotic null distribution is given by the following theorem (notice that $w_j = n_j/n$ which under our notation for the two sample case is $w_1 = \lambda_n$ and $w_2 = 1 - \lambda_n$):

Theorem 2 (El Barmi-McKeague) *Under $H_0$ and assuming that the common distribution* *function $F$ is continuous,*

$$T_n \xrightarrow{d} \sum_{j=1}^{k} w_j \int_0^1 \frac{\left(E_{\mathbf{w}}\left[\mathbf{B}(t)|\mathcal{I}\right]_j - \overline{B}(t)\right)^2}{t(1-t)}\, dt, \tag{2.31}$$

*where $\boldsymbol{B} = \left(B_1/\sqrt{w_1}, B_2/\sqrt{w_2}, \ldots, B_k/\sqrt{w_k}\right)^T$, the processes $B_1, B_2, \ldots, B_k$ are independent Brownian bridges, and $\overline{B} = \sum_{j=1}^{k} \sqrt{w_j} B_j$.*

For the two-sample case, the limiting distribution in Theorem 2 of El Barmi and McKeague's 2013 paper coincides with that in the one-sample case given in Theorem 1.

Theorem 1 (El Barmi-McKeague) *If $F_0$ is continuous, then under $H_0$,*

$$T_n \xrightarrow{d} \int_0^1 \frac{B^2(t)}{t(1-t)} I\left(B(t) \geq 0\right)\, dt, \tag{2.32}$$

*where $B$ is a standard Brownian bridge.*

We used simulations to obtain the critical values for the empirical likelihood ratio test. The critical values given in Table 1 are based on 100,000 data sets distributed as $N(0,1)$ with samples of size 10,000 for each case. The program used to obtain these results is available in Appendix A.

TABLE 2.1

CRITICAL VALUES FOR THE EMPIRICAL LIKELIHOOD RATIO TEST

| $k$ | 0.10 | 0.05 | 0.01 |
|---|---|---|---|
| 2 | 1.3318741 | 1.872133 | 3.243249 |

After studying the behavior of the two test statistics $S_n$ and $T_n$, we noticed that the values in Table 2 of El Barmi and McKeague (2013) seemed inaccurate. In each case, we use 10,000 data sets to simulate the power (percentage of time the test rejected the null hypothesis in favor of the alternative) at $\alpha = 0.05$. El Barmi and McKeague (2013) point out that in all cases, $T_n$ has greater power than $S_n$ and better agreement with $\alpha$. However, we also notice that with the corrected values, the power is even greater than they originally noticed.

TABLE 2.2

CORRECTED TABLE 2 FROM EL-BARMI AND MCKEAGUE (2013)

| Distributions | | $n_1 = 50,$ | $n_2 = 30$ | $n_1 = 30,$ | $n_2 = 50$ | $n_1 = 50,$ | $n_2 = 50$ |
|---|---|---|---|---|---|---|---|
| $F_1$ | $F_2$ | $T_n$ | $S_n$ | $T_n$ | $S_n$ | $T_n$ | $S_n$ |
| $U(0,1)$ | $U(0,1)$ | 0.0543 | 0.0373 | 0.0541 | 0.0404 | 0.0523 | 0.0369 |
| $U(0,1.1)$ | $U(0,1)$ | 0.2044 | 0.1086 | 0.2222 | 0.1194 | 0.2575 | 0.1150 |
| $U(0,2)$ | $U(0,1)$ | 1.0000 | 0.9994 | 0.9992 | 0.9962 | 1.0000 | 1.0000 |
| $U(0.1,1.1)$ | $U(0,1)$ | 0.5172 | 0.2354 | 0.5008 | 0.2400 | 0.6301 | 0.2830 |
| | | | | | | | |
| $Exp(1)$ | $Exp(1)$ | 0.0532 | 0.0397 | 0.0537 | 0.0407 | 0.0522 | 0.0355 |
| $Exp(1)$ | $Exp(1.1)$ | 0.1070 | 0.0704 | 0.1039 | 0.0720 | 0.1130 | 0.0680 |
| $Exp(1)$ | $Exp(2)$ | 0.8598 | 0.7181 | 0.8380 | 0.7138 | 0.9312 | 0.8185 |
| $0.1 + Exp(1)$ | $Exp(1)$ | 0.2361 | 0.1178 | 0.2238 | 0.1068 | 0.2703 | 0.1211 |
| | | | | | | | |
| $N(0,1)$ | $N(0,1)$ | 0.0550 | 0.0400 | 0.0555 | 0.0366 | 0.0526 | 0.0341 |
| $N(0.1,1)$ | $N(0,1)$ | 0.1109 | 0.0795 | 0.1157 | 0.0837 | 0.1214 | 0.0781 |
| $N(0.5,1)$ | $N(0,1)$ | 0.6692 | 0.5293 | 0.6747 | 0.5331 | 0.7791 | 0.6254 |
| $N(1,1)$ | $N(0,1)$ | 0.9940 | 0.9764 | 0.9948 | 0.9767 | 0.9990 | 0.9948 |

# CHAPTER 3

# INTRODUCTION OF THE MULTIPLE DECISION PROCEDURE FOR STOCHASTIC ORDERING AMONG TWO DISTRIBUTION FUNCTIONS

In 2012, Ledwina and Wylupek introduced a variation of Terpstra and Magel's (2002) approach to non-parametric location parameters to solve the problem of stochastic dominance. When conducting the traditional hypothesis test of $H_0$ in (2.1) versus $H_1 - H_0$ where $H_1$ is given in (2.2), an *a priori* knowledge about the distribution functions is assumed. Similar to the problem of non-parametric location parameters, this test precludes all other possible orderings or non-orderings of the distribution functions. Ledwina and Wylupek (2012) formulate their test as follows: test that there is no stochastic dominance against the alternative that the stochastic dominance takes place. In other words, let $H$ : lack of stochastic dominance and let $H_+$ : presence of stochastic dominance. They begin their process by reparameterizing the hypotheses $H$ and $H_+$ in terms of the Fourier coefficients of the two comparison density functions $s_1$ and $s_2$ and the function $\overline{s}$. To begin defining these functions, let $n_1, n_2$, and $n$ be defined as before and assume that the proportion of the observations in each sample remains fixed as the total sample size grows, with the proportions $\lambda_n$ and $1-\lambda_n$ still bounded between 0 and 1. Let $\mu$ denote Lebesgue measure on (0, 1). Define

$$H_0 = \lambda_n F_1 + (1-\lambda_n)F_2, \tag{3.1}$$

and define the comparison densities $s_1$ and $s_2$ as follows:

$$s_1 = \frac{d\left(F_1 \circ H_0^{-1}\right)}{d\mu} \tag{3.2}$$

and

$$s_2 = \frac{d\left(F_2 \circ H_0^{-1}\right)}{d\mu}. \tag{3.3}$$

Then let

$$\overline{s}(z) = s_2(z) - s_1(z), \tag{3.4}$$

which the authors propose calling the contrast function. This function plays a crucial role in the two-sample case and is investigated by many authors including Behnen and Neuhaus (1983), Handcock and Morris (1999), and Wylupek (2010). In order to express $H$ and $H_+$ in terms of the Fourier coefficients of the two comparison density functions $s_1$ and $s_2$ and the contrast function $\overline{s}$ in a system in $L_2\left((0,1),\mu\right)$, we begin with the standard Haar basis. Define $a_1$, $a_2$, ... to be be points of the form

$$\frac{2i-1}{2^{k+1}} \tag{3.5}$$

where for each $k = 0, 1, \ldots$ we evaluate at all values of $i$ where $i = 1, 2, \ldots, 2^k$. Then by Moriguti's algorithm given in Rychlik (2001), it is possible to project each of the successive Haar functions into the cone $C$ of non-decreasing functions. The resulting solutions are standardized to obtain a system of non-decreasing functions on $[0, 1]$ with an $L_2$-norm of unit value defined by

$$l_j(z) = -\sqrt{\frac{1-a_j}{a_j}} I\left(0 \leq z < a_j\right) + \sqrt{\frac{a_j}{1-a_j}} I\left(a_j \leq z \leq 1\right), \tag{3.6}$$

where $I(A)$ is the indicator function of the set $A$, that is it takes on the value 1 if $z \in A$ and takes on the value 0 if $z \notin A$. Now $\int_0^1 l_j \; d\mu = 0$ and $\int_0^1 l_j^2 \; d\mu = 1$ for all $j \in \mathbb{N}$. Define the $j$th Fourier coefficient of $\overline{s}$ as follows:

$$\gamma_j = \gamma_j(a_j) = \int_0^1 \overline{s}(z) l_j(z) \mu(dz) = \sqrt{\frac{1}{a_j(1-a_j)}} \{F_1(H_0^{-1}(a_j)) - F_2(H_0^{-1}(a_j))\}, \tag{3.7}$$

for each $j \in \mathbb{N}$. If $F_1 \geq F_2$, then $\gamma_j \geq 0$ for all $j$; but since the points $a_j$ are dense in the interval $[0, 1]$ and furthermore $F, G, H_0$ are continuous functions, we can say that

$F_1(x) \geq F_2(x)$ for each $x \in \mathbb{R}$ if and only if $\gamma_j \geq 0$ for all $j$ in the natural numbers. Specifically, if the strict inequality holds for some $x$, then there must exists a $j \in \mathbb{N}$ such that $\gamma_j$ is strictly greater than zero. Instead of testing the hypotheses $H$ versus $H_+$ on $F_1$ and $F_2$, we may instead test the equivalent hypotheses on the $\gamma_j$, that is, $H_+$ is equivalent to $\gamma_j \geq 0$ for $j = 1, 2, \ldots$ with at least one strict inequality and $H$ is the complement.

First, we look at the issue of testing the contrast function identically equal to zero versus $\bar{s}(z) \neq 0$ for some $z \in (0, 1)$. This is equivalent to testing the Fourier coefficients $\gamma_j = 0$ for all $j$ against the alternative that $\gamma_j \neq 0$ for some $j$. As the contrast function is unknown, we define $w(k) = 2^{k+1} - 1$, for $k = 0, 1, \ldots$ and define the set of dimensions $\mathcal{D}(n) = \{w(k) : k = 0, 1, \ldots, k(n)\}$ where $k(n)$ is a non-decreasing sequence of natural numbers. For $d \in \mathcal{D}$, define $\mathbf{m}_d = \langle l_1, l_2, \ldots, l_d \rangle$, $\mathcal{M}_d = \text{span}\{\mathbf{m}_d\}$, $d(N) = w(k(n))$ and $\boldsymbol{\gamma}(n) = \langle \gamma_1, \ldots, \gamma_{d(n)} \rangle$. The contrast function, $\bar{s}$, is well approximated by some element of $\mathcal{M}_d$ for some $n$ and $d \in \mathcal{D}(n)$. If $d$ is known, it would suffice to test $\gamma_1 = \cdots = \gamma_d = 0$ against $\gamma_j \neq 0$ for some $j \in \{1, \cdots, d\}$. This problem can be solved by applying a score test based on the test statistic

$$W_d = \boldsymbol{L}_d \boldsymbol{\Sigma}_d^{-1} \boldsymbol{L}_d' \tag{3.8}$$

where the score vector is given by

$$\boldsymbol{L}_d = \langle L_1, \cdots, L_d \rangle \tag{3.9}$$

with each $L_j$ defined by

$$L_j = \sum_{i=1}^{n} c_{ni} l_j \left( \frac{R_i - 0.5}{n} \right) \qquad \text{with} \qquad c_{ni} = \sqrt{\frac{n_1 n_2}{n}} \begin{cases} -n_1^{-1} & 1 \leq i \leq n_1 \\ n_2^{-1} & n_1 < i \leq n \end{cases} \tag{3.10}$$

where $R_1, \ldots R_n$ are the ranks of $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n_2}$ in the pooled sample of size $n$. $\boldsymbol{\Sigma}_d^{-1}$ is the inverse of the limiting covariance matrix, $\boldsymbol{\Sigma}_d$, of the score vector under the assumption that $F_1 = F_2$. Unfortunately, $d$ is not generally known and decisions on $d$ must be based off of the data. Janic-Wroblewska and Ledwina (2000) suggest trying

$$S = \min\{d \in \mathcal{D}(n) : W_d - d\log n \geq W_j - j\log n, j \in \mathcal{D}(n)\}, \tag{3.11}$$

to find a suitable choice of $d$ called $S$, which mimics Schwarz's BIC. As very distinct alternatives are easily detected, we want to focus on methods that are able to detect alternatives that are not very far away from the null hypothesis, and Ledwina and Wylupek believed they could improve upon this choice of $d$. Define the indicator function

$$J(n) = I\left(\max_{1\leq j\leq d(n)}\{|L_j|\} \leq \sqrt{t\log n}\right) \tag{3.12}$$

for $t$ a positive constant called the tuning parameter. Ledwina and Wylupek (2012) discuss multiple selection methods for this tuning parameter. Under the assumption $F_1 = F_2$, $L_i$ for $i = 1, \ldots, d$, $d$ fixed, are asymptotically $N(0, 1)$ but they are not independent. When $F_1 \neq F_2$, define $j_0$ to be the smallest index such that $\gamma_{j_0} = \gamma_{j_0}(a_{j_0}) \neq 0$ and let $d_0$ be the smallest dimension of the dyadic partition of $[0, 1]$ containing $a_{j_0}$. Both $j_0$ and $d_0$ depend on $F_1$ and $F_2$ but not on $n$. Now we can use $J(n)$ to make decisions about the hypotheses according to the following Lemma.

Lemma 1 (Ledwina-Wylupek) *Assume that $F_1 = F_2$ and $d(n) = o(\log n)$. Then for any positive $t$*

$$\lim_{n\to\infty} P(J(n) = 0) = \lim_{n\to\infty} P\left(\max_{1\leq j\leq d(n)}\{|L_j|\} \geq \sqrt{t\log N}\right) = 0. \tag{3.13}$$

*If $F_1 \neq F_2$, then for any sequence $d(n) \to \infty$ as $n \to \infty$ and any positive $t$ it holds that*

$$P(J(n) = 0) = \lim_{n\to\infty} P\left(\max_{1\leq j\leq d(n)}\{|L_j|\} \geq \sqrt{t\log N}\right) = 1. \tag{3.14}$$

*In the case when $d(n) = D$, $D$ is a constant,* (3.14) *holds for all $F_1 \neq F_2$ such that $d_0 \leq D$.*

So when $J(n) = 1$ we may say that $\gamma(n) = 0$ and when $J(n) = 0$, $\gamma(n) \neq 0$, that is we deviate from the hypothesis $F_1 = F_2$. Define the penalty function

$$\Pi_d = \Pi_d(n) = \begin{cases} d \log n & \text{if } J(n) = 1, \\ 0 & \text{if } J(n) = 0. \end{cases} \tag{3.15}$$

Then we make our selection of $d$ called $T$ as follows:

$$T = \min\{d \in \mathcal{D}(n) : W_d - \Pi_d \geq W_j - \Pi_j, j \in \mathcal{D}(n)\}. \tag{3.16}$$

When $J(n) = 1$, the selection $T$ coincides with $S$ and will mimic Schwartz's BIC. We now give some properties of our choice of $T$.

Lemma 2 (Ledwina and Wylupek) *If $F_1 = F_2$ and $d(n) = o\left(\sqrt{\log n}\right)$, then*

$$\lim_{n\to\infty} P(T = 1) = 1. \tag{3.17}$$

*Assume that $F_1 \neq F_2$ and let $j_0$ and $d_0$ be define as before. Then,*

$$\lim_{n\to\infty} P(T \geq d_0) = 1 \tag{3.18}$$

*provided that $d(n) = D$, where $D$ is a constant, and $D \geq d_0$.*

Letting $W_d$ now be defined with $d = T$, so that we use $W_T$ in the test of $\gamma_1 = \cdots = \gamma_n = 0$ against the alternative that $\gamma_j \neq 0$ for some $j \in \{1, \ldots, n\}$, $W_T$ will have the following properties:

Corollary 1 (Ledwina and Wylupek) *Under the assumptions of lemma 2, $W_T \xrightarrow{\mathcal{L}} \chi_1^2$, under $\gamma_1 = \cdots = \gamma_n = 0$, and $W_T \xrightarrow{P} +\infty$, under $\gamma_j \neq 0$ for some $j \in \{1, \ldots, n\}$ where $\chi_1^2$ denotes a random variable with the central chi-square distribution with one degree of freedom while $\xrightarrow{\mathcal{L}}$ denotes convergence in law.*

If we restrict our attention to the problem of testing the null hypothesis $\gamma_j \geq 0$ against the alternative $\gamma_j < 0$, we accept the null hypothesis when $L_j > z(\alpha)$ where $P\left(L_j \leq z(\alpha)\right) = \alpha$

under $\gamma_j = 0$, and the $z(\alpha)$ values are found from the standard normal distribution tables. Now define

$$Z(n) = Z(n, \alpha) = \text{sign}\left[\underset{1\leq j\leq d(n)}{min} L_j - z(\alpha)\right], \tag{3.19}$$

so that $Z(n) = +1$ only when all single tests of $\gamma_j \geq 0$ against $\gamma_j < 0$, for $j = 1, \ldots, d(n)$ accept the null hypothesis.at level $\alpha$. The following lemma gives some properties of the function $Z(n)$.

Lemma 3 (Ledwina and Wylupek) *Assume that $F_1 = F_2$, $d(n) = o(n)$. Then it holds that*

$$\underset{1\leq j\leq d(n)}{max}\{-L_j\} = O_P\left(\sqrt{\log\log d(n)}\right). \tag{3.20}$$

*If, additionally, $d(n) \to \infty$ as $N \to \infty$, then*

$$\lim_{N\to\infty} P\left(Z(n) = -1\right) = 1. \tag{3.21}$$

Define $H_{(\xi)}(x) = \xi F_1(x) + (1-\xi)F_2(x)$ for $\xi \in (0,1)$ and let $K_\xi(z) = F_1 \circ H^{-1}_{(\xi)}(z)$ for $z \in (0,1)$. Furthermore, we assume that the functions $K_\xi$ have derivatives $k_\xi$ for all $z \in (0,1)$. For some $\xi' \in (0,1)$, $k_{\xi'}$ is continuous on $(0,1)$ and will have one sided limits at the boundary points. Pyke and Shorack (1975) give sufficient conditions ensuring that this assumption holds. We call this assumption 1.

Lemma 4 (Ledwina and Wylupek) *Suppose that $H$ is true and $F_1 \neq F_2$. Impose assumption 1 with $k_\xi$ bounded on $[0,1]$. If $d(n) \leq n$ and $d \to \infty$, then*

$$\lim_{n\to\infty} P(Z(n) = -1) = 1. \tag{3.22}$$

*In the case where $d(n) = D$, where $D$ is a constant,* (3.22) *holds for all $F_1 \neq F_2$ such that $d_0 \leq D$.*

Ledwina and Wylupek (2013) now propose to reject $H$ when each empirical Fourier coefficient can be expected to be non-negative and the value of $W_T$ is large. Define the test statistic

$$V_T = Z(n)W_T \tag{3.23}$$

so that we reject $H$ for large values of $V_T$.

A few more notations need to be defined prior to giving a theorem about the behavior of the new test statistic. Suppose $F_1 = F_2$. For a given $n$, define $v(\alpha) = v(\alpha, n)$ to be the smallest value such that $P(V_T > v(\alpha)|F_1 = F_2) \leq \alpha$. Let $\Gamma^o_{jj} = (1 - \lambda_n)\{s_2(a_j)\}^2 + \lambda_n\{s_1(a_j)\}^2$ and recall that $\Phi$ denotes the cumulative distribution function of the standard normal distribution and $\chi^2_1$ the central chi-squared random variable with one degree of freedom.

Theorem 1 (Ledwina and Wylupek) *Impose assumption 1 with $k_\xi$ bounded on $[0, 1]$.*

1. *Assume that $H$ is true. If $F_1 = F_2$, $d(n) = o\left(\sqrt{\log n}\right)$ and $d(n) \to \infty$, then*

$$\lim_{n\to\infty} P\left(V_T \leq x|F_1 = F_2\right) = P\left(-\chi^2_1 \leq x\right) \text{ for any } x \in \mathbb{R}. \tag{3.24}$$

*In the case $F_1 \neq F_2$, under the assumptions on $d(n)$ as for (3.14) and (3.22), it holds that*

$$\lim_{n\to\infty} P\left(V_T \leq x|H, F_1 \neq F_2\right) = 1 \text{ for any } x \in \mathbb{R}. \tag{3.25}$$

2. *Assume now that $H_+$ holds true, $d(n) = D$, where $D$ is a constant, and consider $F_1$ and $F_2$ such that $d_0 \leq D$. If $F_1 \geq F_2$ and all $\gamma_1, \ldots, \gamma_D$ are strictly positive, then*

$$\lim_{n\to\infty} P\left(V_T \leq v(\alpha)|H_+\right) = 0. \tag{3.26}$$

*In the case when $F_1 \geq F_2$ and there are some $\gamma_j = 0$ among $\gamma_1, \ldots, \gamma_D$, it holds that*

$$\lim_{n\to\infty} P\left(V_T \leq v(\alpha)|H_+\right) \leq 1 - \prod_{j=1:\gamma_j=0}^{D} \left(1 - \Phi\left(\frac{z(\alpha)}{\sqrt{\Gamma^o_{jj}}}\right)\right). \tag{3.27}$$

*Moreover, $1 \leq \Gamma^o_{jj} \leq \{\lambda_n(1 - \lambda_n)\}^{-1}$ and $\Gamma^o_{jj} = 1$ if and only if $s_1(a_j) = s_2(a_j) = 1$.*

While Ledwina and Wylupek's procedure was interesting, it was quite complex and computationally intensive. We propose a multiple decision procedure, similar to that proposed for ordering among means. We will define the decisions as follows:

$$D_0 : F_1(x) = F_2(x) \text{ for all } x \text{ cannot be rejected} \tag{3.28}$$

$$D_1 : F_1(x) \geq F_2(x) \text{ with strict inequality for at least one } x \tag{3.29}$$

$$D_2 : F_1(x) \leq F_2(x) \text{ with strict inequality for at least one } x \tag{3.30}$$

$$D_3 : D_1 \text{ and } D_2 \text{ both hold} \tag{3.31}$$

In order to test these decisions we must calculate the test statistics for $H_0$ versus $H_1 - H_0$ and $H_0$ versus $H_2 - H_0$ at the same time. For any of the tests described so far, we need only reverse the ordering of $F_1$ and $F_2$ to obtain the new test statistic. Then we conclude $D_0$ for sufficiently small values of both test statistics; conclude $D_1$ when the test statistic for the first one-sided test is sufficiently large and the second is sufficiently small; conclude $D_2$ when the test statistic for the first one-sided test is sufficiently small and the second is sufficiently large; and conclude $D_3$ when both test statistics are sufficiently large. This approach will offer increased protection from incorrect conclusions. For comparisons we ran the multiple decision procedure for both $S_n$ proposed by El Barmi and Mukerjee (2005) and $T_n$ proposed by El Barmi and McKeague (2013).

# CHAPTER 4

## SOME SIMULATIONS IN SUPPORT OF THE MULTIPLE DECISION PROCEDURE

Using 10,000 simulations in each case, we have computed the percentage of the time that the multiple decision procedure (MDP) using empirical likelihood ratio (ELR) test statistic $T_n$ and the Kolmogorov-Smirnov variant $S_n$ proposed by El Barmi and Mukerjee (EM) concludes $D_0$, $D_1$, $D_2$, and $D_3$ for a couple of distribution functions $F_1$ with sample size $n_1$ and $F_2$ with sample size $n_2$. In each case, we have also recorded the percentage of the time the traditional Kolmogorov-Smirnov (KS) test concludes $D_1$ and noted the reduction percentage of |KS - MDP|/KS×100.

To begin our simulation process we calculated the cut-off values for the Kolmogorov-Smirnov variant test for $\alpha = (0.01, 0.05, 0.10)$ as follows:

$$e^{-2t^2} = .01 \implies t = \left(-\left(\frac{1}{2}\right)\ln .01\right)^{\frac{1}{2}} = 1.5174,$$

$$e^{-2t^2} = .05 \implies t = \left(-\left(\frac{1}{2}\right)\ln .05\right)^{\frac{1}{2}} = 1.2239,$$

$$e^{-2t^2} = .10 \implies t = \left(-\left(\frac{1}{2}\right)\ln .10\right)^{\frac{1}{2}} = 1.0730.$$

Then we considered $n_1 = n_2 = (50, 100, 150, 200)$. Since $S_n = \sqrt{n_1 n_2/(n_1+n_2)}\,\sup(F_2 - F_1)$, we will have $\sqrt{n_1 n_2/(n_1+n_2)} = (5, \sqrt{50}, \sqrt{75}, 10)$. For $p$-values to be less than $\alpha$, $\sup(F_2 - F_1)$ must be more than $t_\alpha/(5, \sqrt{50}, \sqrt{75}, 10)$. For $\alpha = (0.01, 0.05, 0.10)$, these computations give

$$(.3035, .2146, .1752, .1517) \text{ for } \alpha = 0.01,$$

$$(.2248, .1731, .1413, .1224) \text{ for } \alpha = 0.05,$$

$$(.2146, .1517, .1239, .1073) \text{ for } \alpha = 0.10.$$

These results indicate that the distribution functions with the same supports will require very large sample sizes. We did exact computations for our three cases.

For case one, we considered the situation in which stochastic ordering was present. Let $F_1 = U(0,1)$ and let $F_2 = x^2$ so that $F_1 \geq F_2$ for all $x \in [0,1]$, in other words $D_1$ holds. Graphic representation of distribution functions $F_1$ and $F_2$ can be found in Figure 4.1. Next we needed to find the absolute value of the difference between the two functions so that we could determine the appropriate sample sizes with which to simulate.

$$(F_1(x) - F_2(x))' = 1 - 2x = 0$$

$$\implies x = \frac{1}{2}$$

$$\implies x - F_2(x) = \frac{1}{2} - \frac{1}{4} = .25$$

Due to this difference between the distribution functions, we used samples of sizes $n_1 = n_2 = (100, 50, 50)$ for $\alpha = (0.01, 0.05, 0.10)$. The resulting values of the simulations can be found in Tables 4.1, 4.2, and 4.3.

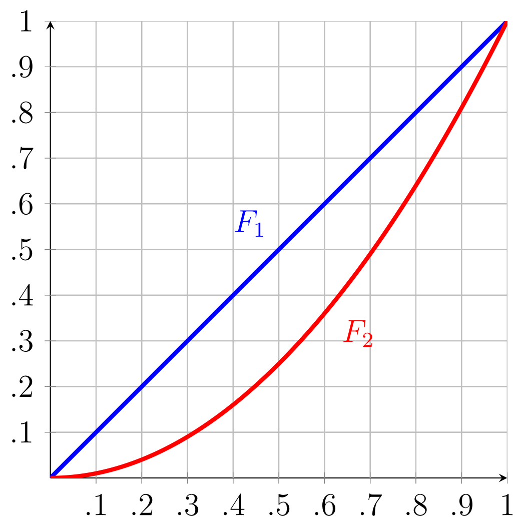


FIGURE 4.1

COMPARISON OF $F_1 = U(0,1)$ TO $F_2 = x^2$

TABLE 4.1

$F_1 = U(0,1)$ VS $F_2 = x^2$ WITH $\alpha_1 = \alpha_2 = 0.01$ AND $n_1 = n_2 = 100$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.9284 | | | |
| EM (MDP) | 0.0716 | 0.9284 | 0.0000 | 0.0000 | 0.00 |
| ELR (MDP) | 0.0192 | 0.9808 | 0.0000 | 0.0000 | 5.64 |

TABLE 4.2

$F_1 = U(0,1)$ VS $F_2 = x^2$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 50$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.8064 | | | |
| EM (MDP) | 0.1936 | 0.8064 | 0.0000 | 0.0000 | 0.00 |
| ELR (MDP) | 0.0739 | 0.9260 | 0.0001 | 0.0000 | 14.83 |

TABLE 4.3

$F_1 = U(0,1)$ VS $F_2 = x^2$ WITH $\alpha_1 = \alpha_2 = 0.10$ AND $n_1 = n_2 = 50$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.9216 | | | |
| EM (MDP) | 0.0783 | 0.9216 | 0.0000 | 0.0000 | 0.00 |
| ELR (MDP) | 0.0357 | 0.9642 | 0.0000 | 0.0001 | 4.62 |

For case two, we considered a symmetric situation in which no stochastic ordering was present. Let $F_1(x) = 4\left(x - \frac{1}{2}\right)^3 + \frac{1}{2}$ and let $F_2 = U(0,1)$ so that $D_3$ holds for all $x \in [0,1]$. Graphic representation of distribution functions $F_1$ and $F_2$ can be found in Figure 4.2. Next we needed to find the absolute value of the difference between the two functions so that we could determine the appropriate sample sizes with which to simulate.

$$(F_1(x) - F_2(x))' = 12\left(x - \frac{1}{2}\right)^2 - 1 = 0$$

$$\implies x = \frac{1}{2} \pm \frac{1}{\sqrt{12}}$$

$$\implies F_1(x) - x = 4\left(-\frac{1}{\sqrt{12}}\right)^3 + \frac{1}{\sqrt{12}} = 0.1925.$$

Due to this difference between the distribution functions, we used samples of sizes $n_1 =$

$n_2 = (150, 100, 100)$ for $\alpha = (0.01, 0.05, 0.10)$. The resulting values of the simulations can be found in Tables 4.4, 4.5, and 4.6.

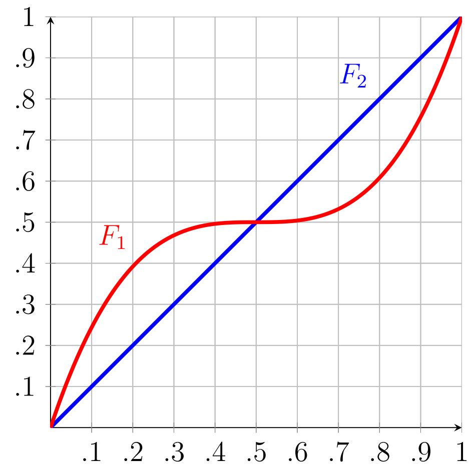


FIGURE 4.2

COMPARISON OF $F_1 = 4(x - .5)^3 + .5$ TO $F_2 = U(0, 1)$

TABLE 4.4

$F_1 = 4(x - .5)^3 + .5$ VS $F_2 = U(0, 1)$ WITH $\alpha_1 = \alpha_2 = 0.01$ AND $n_1 = n_2 = 150$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.8047 | | | |
| EM (MDP) | 0.0059 | 0.1900 | 0.1894 | 0.6147 | 76.39 |
| ELR (MDP) | 0.0231 | 0.2655 | 0.2588 | 0.4526 | 67.00 |

TABLE 4.5

$F_1 = 4(x - .5)^3 + .5$ VS $F_2 = U(0, 1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 150$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.7963 | | | |
| EM (MDP) | 0.0046 | 0.1937 | 0.1991 | 0.6026 | 75.67 |
| ELR (MDP) | 0.0129 | 0.2368 | 0.2420 | 0.5083 | 70.26 |

TABLE 4.6

$F_1 = 4(x - .5)^3 + .5$ VS $F_2 = U(0, 1)$ WITH $\alpha_1 = \alpha_2 = 0.10$ AND $n_1 = n_2 = 100$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.8803 | | | |
| EM (MDP) | 0.0013 | 0.1223 | 0.1184 | 0.7580 | 86.11 |
| ELR (MDP) | 0.0019 | 0.1336 | 0.1302 | 0.7343 | 84.82 |

For case three, we considered a second situation in which no stochastic ordering and no symmetries were present. Let $F_1(x) = \frac{16}{9}\left(x - \frac{3}{4}\right)^3 + \frac{3}{4}$ and let $F_2 = U(0,1)$ so that $D_3$ holds for all $x \in [0, 1.27]$. Graphic representation of distribution functions $F_1$ and $F_2$ can be found in Figure 4.3. Next we needed to find the absolute value of the difference between the two functions so that we could determine the appropriate sample sizes with which to simulate.

$$(F_1(x) - F_2(x))' = \frac{16}{3}\left(x - \frac{3}{4}\right)^2 - 1 = 0$$

$$\implies x = \frac{3}{4} \pm \frac{\sqrt{3}}{4}$$

$$\implies F_1(x) - x = \frac{16}{9}\left(-\frac{\sqrt{3}}{4}\right)^3 + \frac{\sqrt{3}}{4} = 0.2887.$$

Due to this difference between the distribution functions, $n_1 = n_2 = 100$ should work for all $\alpha$ levels. We used the two pairs of alpha values $(\alpha_1, \alpha_2) = (.05, .05)$ and $(.05, .10)$ with sample sizes $n_1 = n_2 = (100, 150, 200)$. The resulting values of the simulations can be found in Tables 4.7 - 4.12.

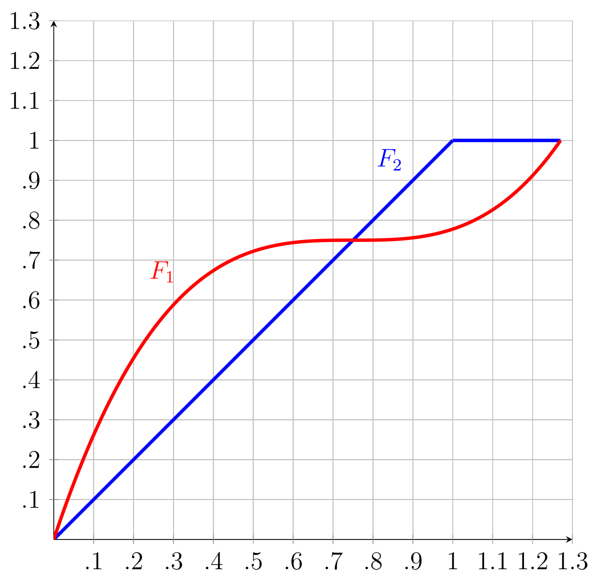


FIGURE 4.3

COMPARISON OF $F_1 = \frac{16}{9}(x - .75)^3 + .75$ TO $F_2 = U(0,1)$

TABLE 4.7

$F_1 = \frac{16}{9}(x - .75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 100$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.9917 | | | |
| EM (MDP) | 0.0000 | 0.1143 | 0.0083 | 0.8774 | 88.47 |
| ELR (MDP) | 0.0000 | 0.1283 | 0.0090 | 0.8627 | 87.06 |

TABLE 4.8

$F_1 = \frac{16}{9}(x - .75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 150$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.9999 | | | |
| EM (MDP) | 0.0000 | 0.0057 | 0.0001 | 0.9942 | 99.42 |
| ELR (MDP) | 0.0000 | 0.0076 | 0.0002 | 0.9922 | 99.24 |

TABLE 4.9

$F_1 = \frac{16}{9}(x - .75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = \alpha_2 = 0.05$ AND $n_1 = n_2 = 200$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | | 1.0000 | | |
| EM (MDP) | 0.0000 | 0.0002 | 0.0000 | 0.9998 | 99.98 |
| ELR (MDP) | 0.0000 | 0.0003 | 0.0000 | 0.9997 | 99.97 |

TABLE 4.10

$F_1 = \frac{16}{9}(x - .75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = 0.05$ AND $\alpha_2 = 0.10$ AND

$n_1 = n_2 = 100$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.9930 | | | |
| EM (MDP) | 0.0000 | 0.1153 | 0.0022 | 0.8825 | 88.39 |
| ELR (MDP) | 0.0000 | 0.0032 | 0.0076 | 0.9604 | 96.78 |

TABLE 4.11

$F_1 = \frac{16}{9}(x - .75)^3 + .75$ VS $F_2 = U(0,1)$ WITH $\alpha_1 = 0.05$ AND $\alpha_2 = 0.10$ AND

$n_1 = n_2 = 150$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 0.9991 | | | |
| EM (MDP) | 0.0000 | 0.0088 | 0.0002 | 0.9910 | 99.12 |
| ELR (MDP) | 0.0000 | 0.0013 | 0.0007 | 0.9980 | 99.98 |

TABLE 4.12

$F_1 = \frac{16}{9}(x - .75)^3 + .75$ VS $F_2 = U(0, 1)$ WITH $\alpha_1 = 0.05$ AND $\alpha_2 = 0.10$ AND

$n_1 = n_2 = 200$

| Test | $D_0$ | $D_1$ | $D_2$ | $D_3$ | RED% |
|---|---|---|---|---|---|
| KS | | 1.0000 | | | |
| EM (MDP) | 0.0000 | 0.0001 | 0.0000 | 0.9999 | 99.99 |
| ELR (MDP) | 0.0000 | 0.0000 | 0.0000 | 1.0000 | 100.00 |

# CHAPTER 5

## AN EXAMPLE FOR THE MULTIPLE DECISION PROCEDURE

An article by Khamaladze, Brownrigg, and Haywood concluded that the lengths of rule of Roman Emperors were exponentially distributed. This implied that their reigns ceased unexpectedly. We used the same list of $n = 70$ Roman Emperors from Augustus to Theodossius, 27 BC to 395 AD, from the chronology due to Parkin. We first considered whether there is an effect on duration of rule due to the Crisis of the Third Century when the Roman Empire neared collapse due to civil war (235 - 284 AD). During the Principate (27 BC - 235 AD), the Empire was relatively stable ($m = 29$). During the period following ($n = 41$) is when we believe there might be a decline. El Barmi and McKeague claim that both distributions appear to be exponential. Using their test to compare the Principate to an exponential model, we conclude $D_0$ (the same distribution) with large $p$-values of 0.84 and 0.55. Similarly we conclude $D_0$ for the period following with large $p$ values of 0.71 and 0.99. Using their test statistic to test for stochastic dominance (decline after the Principate), we obtain a test statistic $T_n = 0.398$. This yields a large $p$-value of 0.625. Similarly, the test based off of El Barmi and Mukerjee's test procedure yields a large $p$-value of 0.578.

# CHAPTER 6

# SUMMARY AND CONCLUDING REMARKS FOR THE MULTIPLE DECISION PROCEDURE

As with the multiple decision procedure used in testing for ordered alternatives in one-way ANOVA, the multiple decision procedure introduced for testing for stochastic ordering between two distribution functions performs much better than the traditional test as the data results show. Minimal power is lost when comparing $D_1$ should be concluded, and more information is gained when $D_3$ should be decided. The traditional test will always conclude $D_1$ in these cases, whereas the multiple decision procedure will make a distinction between the many possibilities.

# CHAPTER 7

# INTRODUCTION TO AN IMPROVED ESTIMATION OF A DECREASING DENSITY OF A RANDOM VARIABLE

Let $X, X_1, X_2, ...$ be independent identically distributed (IID) random variables (RVs) with distribution function (DF) $F$ and a decreasing density $f$ with support $[0, \tau], \tau < \infty$ or $[0, \infty)$; throughout, we write decreasing (increasing) to mean nonincreasing (nondecreasing). Grenander (1956) showed that the nonparametric maximum likelihood estimator (NPMLE) of $F$ under this restriction is the least concave mojorant $\hat{F}_n$ of $F_n$, the empirical DF based on the order statistics $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$, and the estimate of $f(x)$ is given by $\hat{f}_n(x)$, the left derivative of $\hat{F}_n$ at $x$. There is an equivalent way of defining $\hat{f}_n$. Let $f_n$ be the (very jagged) naive estimator of $f$:

$$f_n(x) = \frac{F_n(X_{(i)}) - F_n(X_{(i-1)})}{X_{(i)} - X_{(i-1)}} I(X_{(i-1)} < x \le X_{(i)}), \quad 1 \le i \le n,$$

with $X_{(0)} = 0$. The numerator above is $1/n$ with probability 1 by our continuity assumption. The NPMLE of $f$ can be viewed as the least squares isotonic regression of $\{(f_n(X_{(i)}), w_i)\}$ on the linearly ordered set $(1 < 2 < \cdots < n)$ with the weight vector $\{w_i = X_{(i)} - X_{(i-1)}\}$, subject to $\{f_n(X_{(i)})\}$ decreasing. An explicit expression for $\hat{f}_n(X_{(i)})$ is given by the so-called max-min formula:

$$\hat{f}_n(X_{(i)}) = \max_{s \ge i} \min_{r \le i} \frac{\sum_{j=r}^{s} w_j f_n(X_{(j)})}{\sum_{j=r}^{s} w_j} = \max_{s \ge i} \min_{r \le i} \frac{s - r + 1}{n(X_{(s)} - X_{(r-1)})}, \quad 1 \le i \le n, \qquad (7.1)$$

which is extended as a left-continuous step function to $(0, \infty)$ that is constant between successive order statistics, and then to $[0, \infty)$ by defining $\hat{f}_n(0) = \hat{f}_n(0+)$. Throughout, we will encounter functions defined on $\{x > 0\}$ that will have right hand limits at 0, and we will define such a function $g$ at 0 by $g(0+)$.

If $\hat{f}_n(X_i) = [F_n(X_u) - F_n(X_l)]/(X_u - X_l)$, then $\hat{f}_n(x) = \hat{f}_n(X_i)$ for all $x \in (X_l, X_u]$. This gives another expression for $\hat{f}_n$:

$$\hat{f}_n(x) = \max_{s \geq x} \min_{0 \leq r \leq x} \bar{f}_n((r, s]),\ x \geq 0,\ \text{where } \bar{f}_n(I) = F_n(I)/l(I),\ I \text{ an interval,} \tag{7.2}$$

with $F_n(I) = F_n$-measure of $I$ and $l(I) =$ length of $I$. It is an unbiased estimator of $\bar{f}(I) = P(X \in I)/l(I)$. Here 0/0 is defined to be 0.

The estimator $\hat{f}_n$ takes on $m_n$ (a random number) of distinct values, constant on intervals $B_k \equiv (X_{(i_{k-1})}, X_{(i_k)}], 1 \leq k \leq m_n$, where $X_{(i_0)} = 0$. The $B_k$s will be called the level sets of $\hat{f}_n$. We use expressions like $B_i > (<) B_j$ and $x > (<) B_i$ with obvious meanings.

The estimators $\hat{F}_n$ and $\hat{f}_n$ have been studied extensively. Prakasa Rao (1969) showed

$$[4f(x)f'(x)]^{-1/3} n^{1/3} [\hat{f}_n(x) - f(x)] \xrightarrow{d} \text{argmax}_x [W(x) - x^2],$$

when $f'(x) < 0$, where $W$ is a 2-sided standard Brownian motion starting at 0. Groeneboom (1985) gave an elegant derivation of this, and derived the density of the limiting distribution. Groeneboom and Wellner (2001) tabled values of the density and quantiles of the distribution with a norming that depends on the unknown $f(x)$ and $f'(x)$. It would have been nice to be able to find confidence intervals by bootstrapping. Unfortunately, this was shown to be inconsistent when bootstrapping the empirical by Kosorok (2008). Sen et al. (2010) showed the same inconsistency even when bootstrapping $\hat{F}_n$, but an $m$ out of $n$ bootstrap of $F_n$ gave consistent results.

The Grenander estimator partitions $(0, X_{(n)}]$ into $m_n$ bins with random widths. Kiefer and Wolfowitz (1976) showed that if we choose a fixed number, $k_n$, of bins $(F_n^{-1}((j-1)/k_n), F_n^{-1}(j/k_n)]$ for $1 \leq j \leq k_n$, and estimate $f$ in each bin by the reciprocal of its length divided by $k_n$, then the resulting estimator of $F$ will be asymptotically concave and $||\hat{F}_n - F_n|| = O_{a.s.}((\log n/n)^{2/3}))$, if $k_n$ is chosen appropriately, albeit under some severe restrictions.

**Theorem 1** *(Kiefer-Wolfowitz, 1976). If $F$ has compact support, and within the support, $f'$ is uniformly continuous, and $f$ and $f'$ are bounded away from 0, then $||\hat{F}_n - F|| = O_{a.s.}((\log n/n)^{2/3})$.*

Balabdaoui and Wellner (2007) give a streamlined proof. It is curious that this theorem has been used by many authors for a large variety of asymptotic results, the conditions of the theorem are not satisfied by most parametric models used in reliability and survival analysis.

There are several papers dealing with global measures of the deviations, $|\hat{F}_n - F_n|$, $|\hat{F}_n - F|$ and $|\hat{f}_n - f|$ in $L_p$ norms, $1 \leq p < \infty$, and the sup-norm; see Wang (1994), Groeneboom et al. (1999), Kulikov and Lophuaä (2005, 2006) and the references therein. All of these results are based on the restrictive assumptions of Theorem 1. We will not pursue the subject of global differences of DFs in this paper.

### 7.1 Estimation of $f(0)$

The estimation of a decreasing density at zero has drawn special attention because of some interesting applications,– see Woodroofe and Sun (1993), Kulikov and Lophuaä (2006), and Balabdaoui et al. (2011). Woodroofe and Sun (1993) showed that $\hat{f}_n(0)$ is inconsistent even though

$$\sup_{x \geq \epsilon} |\hat{f}_n(x) - f(x)| \xrightarrow{a.s.} 0 \quad \text{for all } \epsilon > 0 \text{ if } f \text{ is contnuous.}$$

in fact, $\hat{f}_n(0) \xrightarrow{d} f(0)/U$, where $U \sim \sup_{k \geq 1} \frac{k}{\Gamma(k)}$ with $\Gamma(k)$ being the sum of $k$ IID Exp(1) RVs. The inconsistency could be understood from the max-min formula for $\hat{f}_n(x)$ in equation (7.2), where the 'max' to the right has little or none of the counterbalance of the 'min' to the left when $x$ is near 0.

Woodroofe and Sun (1993) provide a penalized maximum likelihood solution, by maximizing the penalized log likelihood function

$$l_\alpha = \sum_{i=1}^{n} \log f(X_{(i)}) - n\alpha_n f(0)$$

subject to $f(X_{(i)})$ decreasing in $i$ and $\int_0^{X_{(n)}} f(u)du = 1$. Here $\alpha_n \downarrow 0$ is a smoothing parameter. The resulting estimator $\hat{f}_n^{PL}$ is weakly consistent at 0. Although $\hat{f}_n^{PL}$ and $\hat{f}_n$ vary greatly near 0, they are very close globally in the sense that the Hellinger distance between the two goes to 0 in probability. They also consider the asymptotic distribution of $\hat{f}_n^{PL}(0)$ under the assumption that

$$f(x) = f(0) - c_1 x^p((1 + o(1)) \quad \text{for some } c_1 > 0 \text{ and } p > 0.$$

They also require that $\alpha_n = c_2 n^{-(p+1)/(2p+1)}$ in which case

$$n^{p/(p+1)}[\hat{f}_n^{PL}(0) - f(0)] \xrightarrow{d} \sup_{t>0} \frac{W(t) - [c_2 + f(0)c_1^p t^{p+1}]}{t},$$

where $W$ is a standard Brownian motion. From simulations they note that the norming $n^{1/3}$ tends to over-smooth when $F \sim Exp(1)$ and $p = 1$, but it seems reasonable for the half-normal for which $p = 2$ and the norming is $n^{2/5}$. The parameter $p$ is unknown, and no sample estimation procedure has been discussed by the authors. Sun and Woodroofe (1996) do discuss adaptive choices of $c$ in $\alpha_n = cn^{-1/3}$.

Kulikov and Lophuaä (2006) make a thorough study of the behavior of

$$U_c = n^\beta[\hat{f}_n(cn^\alpha) - f(cn^\alpha], \quad 0 < \alpha < 1, \tag{7.3}$$

with a $\beta$ that makes $U_c$ converge in distribution when $f(0) < \infty$ and

$$0 < |f^{(k)}(0)| \leq \sup_{s\geq 0} |f^{(k)}(0)| < \infty \text{ and } f^{(j)}(0) = 0, \quad 1 \leq j < k. \tag{7.4}$$

They derive the necessary $\beta$ for various ranges of $\alpha$ and the limiting distributions of $U_c$ that turn out to be non-normal, and depend on functionals of $f$ that need to be estimated. They also suggest estimating $f(0)$ by $\hat{f}_n(cn^{-1/3})$ and an adaptive estimator to reduce the MSE, both of which are weakly consistent with a rate of $n^{-1/3}$.

Groeneboom and Jongbloed (2014) suggest the bin estimator $F_n(h_n)/h_n$ for $f(0)$, where the optimal bandwidth $h_n \approx n^{-1/3}$. Balabdaoui et al. (2011) consider the case when $f(0) = \infty$ for various growth rates of $f$ near 0 and derive the limiting behavior of $\hat{f}_n$ near 0 for these growth rates. We will show that our estimator $f_n^*(0)$ of $f(0)$, given in equations (7.6) – (7.8) below, is not only consistent when $f(0) < \infty$, but also $n^{2/5}[f_n^*(0) - f(0)]$ has a limiting normal distribution.

### 7.2 A hybrid estimator

From now on we assume that $f(0) < \infty$, $f$ is differentiable everywhere, and $\sup_x |f'(x)| < \infty$ without explicit mention.

The NPMLE does not work out well in some order restricted estimation problems. For example, it is inconsistent for some DFs with increasing failure rate average (Boyles, et al. (1986)) or some star-shaped DFs (Barlow et al. (1972)). Even when the NPMLE is consistent, other estimators can turn out to be superior in a practical sense. For example, the NPMLE for two stochastically ordered DFs (Brunk et al. (1966)) is consistent; however, an ad hoc but intuitive estimator of Hogg (1962) has been shown to be strictly better in terms of MSE at all quantiles by simulations for the distributions considered (Elbarmi and Mukerjee (2005)). Moreover, this estimator easily generalizes to the $k$-sample case, while the NPMLE is not known for $k \geq 3$. Our take is that the NPMLE, if it exists, may be a good starting point, but one should look for possible improvements. We think that the Grenander density estimator has 3 major drawbacks:

(i) The convergence rate is only $n^{-1/3}$,

(ii) the limiting distribution is not normal, and

(iii) bootstrap estimates of the unknown parameters of the limiting distribution are inconsistent.

On the other hand, a kernel estimator $\tilde{f}_{nk_n}$ with a kernel $k_n$ could have $n^{2/5}[\tilde{f}_{nk_n}(x) - f(x)]$ converging to a normal distribution. However, $\tilde{f}_{nk_n}$ is not necessarily decreasing.

From Prakasa Rao's (1969) derivation of the asymptotic distribution of $\hat{f}_n(x)$, the Grenander estimator could be viewed as a bin estimator with a random variable bandwidth of $O_p(n^{-1/3})$. On the other hand, an optimal kernel estimator uses a bandwidth of $O(n^{-1/5})$. Let $h_n \downarrow 0$ be a positive sequence and let

$$\tilde{f}_n(x) = \bar{f}_n((x - h_n, x + h_n]), \quad h_n \leq x \leq X_{(n)} - h_n. \tag{7.5}$$

This is a uniform kernel estimator of $f$ on $[h_n, X_{(n)} - h_n]$ with the kernel $u(x) = (1/2)I(-1 < x \leq 1)$ and a bandwidth of $h_n$.

We propose a hybrid estimator that adds this fixed symmetric band to the average in equation (7.2) in addition to the random band from the max-min process. We extend it to $[0, X_{(n)}]$ with linear tails which provide a first order bias correction. This idea was first used by Mukerjee (1993) where a fast convergence rate was necessary to get a Bahadur-type representation of monotone conditional quantile estimates. Define

$$f_n^*(x) = \begin{cases} \max_{s \geq x+h_n} \min_{0 \leq r \leq x-h_n} \bar{f}_n((r,s]), & x \in [h_n, X_{(n)} - h_n] \\ f_n^*(h_n) + (x - h_n) f_n^{*'}(h_n), & x \in [0, h_n) \\ f_n^*(X_{(n)} - h_n) + (x - X_n + h_n) f_n^{*'}(X_{(n)} - h_n), & x \in (X_{(n)} - h_n, X_{(n)}] \end{cases} \tag{7.6}$$

where

$$f_n^{*'}(h_n) = [f_n^*(2h_n) - f_n^*(h_n)]/h_n \tag{7.7}$$

and

$$f_n^{*'}(X_{(n)} - h_n) = \{[f_n^*(X_{(n)} - h_n) - f_n^*(X_{(n)} - 2h_n)] \vee [f_n^*(X_{(n)}) = f_n^*(X_{(n)} - h_n)]\}/h_n, \tag{7.8}$$

are the estimates of $f'$ at $h_n$ and $X_{(n)}-h_n$, respectively, the $'\vee'$ guarantees that $f_n^*(X_{(n)}) \geq 0$. Note that we need that $X_{(n)} \geq 4h_n$ to define our estimates in the tails.

For $x_1, x_2 \in [h_n, X_{(n)} - h_n]$, if $x_1 < x_2$ and $f_n^*(x_i) = \bar{f}_n([r_i, s_i])$ for $i = 1, 2$, then

$$\bar{f}_n([r_1, s_1]) \geq \bar{f}_n([r_1, s_2]) \geq \bar{f}_n(r_2, s_2])$$

by the definitions of $f_n^*(x_1)$ for the first inequality and of $f_n^*(x_2)$ for the second inequality. From the forms of the extensions in the tails it is easy to see that $f_n^*$ is decreasing and strictly positive on $[0, X_n)$. This $f_n^*$ may not integrate to 1 but it will be very close, and could be normalized by dividing the estimator by its integral.

# CHAPTER 8

## ASYMPTOTIC EQUIVALENCE WITH A KERNEL ESTIMATOR

The random bandwidth of the NPMLE of $f(x)$ is $O_p(n^{-1/3})$ when $f'(x) < 0$. The fixed band of width $2h_n$ in defining $f_n^*$ needs to of a higher order of magnitude to be useful. In this Section we analyze how $h_n$ affects the width of the random "skin" that is added to the fixed band. Our key tool will be the exponential bounds on exceedance probabilities of centered sums of independent Bernoulli random variables due to Bernstein (see Uspensky (1937), page 205).

Define the uniform kernel estimator of $f$ with bandwidth $h_n$ by

$$\tilde{f}_{nu}(x) = \bar{f}_n((x - h_n, x + h_n])$$

for $x \in [h_n, X_n - h_n]$. Our hybrid estimator can also be expressed as a uniform kernel estimator with a random bandwidth:

$$f_n^*(x) = \bar{f}_n((x - h_n - B_{1n}, x + h_n + B_{2n}]),$$

where $N_{1n}$ and $B_{2n}$ are widths of randoms bands (that depend on each sample sequence) added to the fixed band of width $2h_n$. Our first objective is to find how the fixed and the random bandwidths are related.

For a fixed positive sequence $b_n \downarrow 0$, consider the localized version of our estimator in equation (7.3):

$$f_n^\dagger(x) = \max_{x+h_n \le s \le x+h_n+b_n} \; \min_{(x-h_n-b_n)\vee 0 \le r \le x-h_n} \bar{f}_n((r, s]), \quad x \in [h_n, X_{(n)} - h_n]. \tag{8.1}$$

Note that $x - h_n - b_n$ may be less than 0; $x + h_n + b_n > X_{(n)}$ creates no problem since the maximum over $s$ will occur at some $s_0 \le X_{(n)}$ by definition in equation (7.2). For simplicity in notation, $\min_{r \le x-h_n}$ will stand for $\min_{0 \le r \le x-h_n}$ and $\min_{r \ge x-h_n-b_n}$ will stand for $\min_{r \ge (x-h_n-b_n)\vee 0}$ from now on.

Lemmas 1 and 2 below deal with a strong form of equivalence of $f_n^\dagger(x)$ and $f_n^*(x)$ for $x \in [h_n, X_{(n)} - h_n]$, while Lemma 3 deals with a weak form of equivalence under various conditions on $h_n$ and $b_n$. Note that $f_n^*$ is determined in the tails by its values on $[h_n, X_{(n)} - h_n]$.

**Lemma 1** *Assume $f'(x) < 0$ and is continuous in a neighborhood of $x \in [h_n, X_{(n)} - h_n]$ and $nh_n^3/\log n \to \infty$. Then*

$$P[f_n^*(x) \neq f_n^\dagger(x) \quad i.o.] = 0 \tag{8.2}$$

*if $b_n \geq d \log n/(nh_n^2)$ for some $d > 0$. This implies that the widths of the random bands, $B_{1n}$ and $B_{2n}$ added to the fixed band $(x - h_n, x + h_n]$, in $f_n^*(x) = \bar{f}_n((x - h_n - B_{1n}, x + h_n + B_{2n}])$ are $O_{a.s.}((\log n/(nh_n^2))$.*

Proof. We first show that that $\{f_n^*(x) = f_n^\dagger(x)\}$ contains the intersection of the two events,

$$\min_{r \leq x - h_n} \bar{f}_n((r, x + h_n]) \geq \max_{s \geq x + h_n + b_n} \bar{f}_n((x + h_n, s]) \tag{8.3}$$

and

$$\min_{r \leq x - h_n - b_n} \bar{f}_n((r, x - h_n]) \geq \max_{s \geq x + h_n} \bar{f}_n((x - h_n, s]). \tag{8.4}$$

Suppose $f_n^*(x) = \bar{f}_n((r_0, s_0])$. If (8.3) holds then, by the Cauchy Mean Value property of averages (Robertson and Wright (1974)), for all $s \geq x + h_n + b_n$,

$$\bar{f}_n((r_0, s]) \leq \bar{f}_n((r_0, x + h_n]) \leq \max_{t \geq x + h_n} \bar{f}_n((r_0, t]) = \bar{f}_n((r_0, s_0]),$$

implying $s_0 \leq x + h_n + b_n$. Using a similar argument, if (8.4) holds then $r_0 \geq x - h_n - b_n$. Thus, (8.3) and (8.4) imply $f_n^*(x) = f_n^\dagger(x)$. Hence, $P[f_n^*(x) \neq f_n^\dagger(x)]$ is bounded above by the sum of

$$P\left[\min_{r \leq x - h_n} \bar{f}_n((r, x + h_n]) < \max_{s \geq x + h_n + b_n} \bar{f}_n((x + h_n, s])\right] \tag{8.5}$$

and

$$P\left[\min_{r \leq x - h_n - b_n} \bar{f}_n((r, x - h_n]) < \max_{s \geq x + h_n} \bar{f}_n((x - h_n, s])\right]. \tag{8.6}$$

First, consider the probability in (8.5). It is bounded above by the sum of

$$P\left[\max_{s\geq x+h_n+b_n} \bar{f}_n((x+h_n,s])\geq f(x+h_n/2)\right] \tag{8.7}$$

and

$$P\left[\min_{r\leq x-h_n} \bar{f}_n((r,x+h_n])\leq f(x+h_n/2)\right]. \tag{8.8}$$

We first upper bound the probability in (8.7). For each $s\geq x+h_n+b_n$, let

$$Y_{ni}(s)=I(x+h_n<X_i\leq s),\ 1\leq i\leq n,$$

and $\bar{Y}_n(s)=\sum_{i=1}^n Y_{ni}(s)/n=(s-x-h_n)\bar{f}_n((x+h_n,s])$. Now, $\bar{Y}_n(s)/(s-x-h_n)$ is an average of IID RVs with values in $\{0,1/(s-x-h_n)\}$ and mean $\mu_n(s)=[F(s)-F(x+h_n)]/(s-x-h_n)=\bar{f}((x+h_n,s])$. Since $f'(x)<0$, $\mu_n(s)$ is decreasing also for $s$ in a neighborhood of $x$. Moreover, $1/(s-x-h_n)$ is decreasing as $s$ goes up. Thus, $Y_{ni}(s)$ is stochastically decreasing as $s$ goes up. From preservation of stochastic ordering under convolutions, $\bar{Y}_n(x+h_n+b_n)$ is stochastically larger than $\bar{Y}_n(s)$ for all $s\geq x+h_n+b_n$. Hence,

$$\max_{s\geq x+h_n+b_n} P(\bar{f}_n(x+h_n,s])\geq f(x+h_n/2))=P(\bar{Y}_n(x+h_n+b_n)/b_n\geq f(x+h_n/2)).$$

Let $\bar{Y}_n=\bar{Y}_n(x+h_n+b_n)$ and $\mu_n=\mu_n(x+h_n+b_n)$. Now,

$$\begin{aligned}
E(\bar{Y}_n) &= b_n\mu_n=F(x+h_n+b_n)-F(x+h_n)\\
&= \int_{x+h_n}^{x+h_n+b_n}[f(x+h_n/2)+f'(x)(u-x-h_n/2)(1+o(1))]du\\
&= b_nf(x+h_n/2)+(1/2)f'(x)b_n(b_n+h_n)(1+o(1))\\
&= b_n[f(x+h_n/2)+(1/2)f'(x)(h_n+b_n)(1+o(1))],
\end{aligned}$$

so that $\mu_n=f(x+h_n/2)-t_n$, where $t_n=-(1/2)f'(x)(h_n+b_n)(1+o(1))$. The same result could be obtained by noticing the approximate linearity of $f$ and continuity of $f'$ near $x$, and by a Taylor expansion around $x+h_n/2$:

$$\mu_n = E[\bar{f}_n((x + h_n, x + h_n + b_n])]$$
$$= (1/2)[f(x + h_n) + f(x + h_n + b_n)](1 + o(1))$$
$$= f(x + h_n/2) + (1/2)f'(x)[h_n/2 + (h_n/2 + b_n)](1 + o(1))$$
$$= f(x + h_n/2) - t_n. \tag{8.9}$$

By Bernstein's inequality for independent Bernoulli variables,

$$\begin{aligned} P(\bar{Y}_n/b_n \geq f(x + h_n/2)) &= P(\bar{Y}_n - b_n\mu_n \geq b_n[f(x + h_n/2) - \mu_n]) \\ &= P(\bar{Y}_n - b_n\mu_n \geq b_n t_n) \\ &\leq \exp\{-(1/2)nb_n^2 t_n^2/(b_n\mu_n + b_n t_n/3)\} \\ &\leq \exp\{-(1/2)nb_n t_n^2/(\mu_n + t_n)\}. \end{aligned}$$

Since $t_n = o(\mu_n)$, we have for all $n$ large enough,

$$P(\bar{Y}_n/b_n \geq f(x + h_n/2)) \leq \exp\{-cnb_n(b_n + h_n)^2\} \leq \exp\{-cnb_n h_n^2\} \tag{8.10}$$

for some $c > 0$ that depends only on $f'(x)$. Since this inequality holds for every $s$ or $X_i$ greater than or equal to $x + h_n + b_n$,

$$P[\max_{s \geq x + h_n + b_n} \bar{f}_n((x + h_n, s]) \geq f(x + h_n/2)] \leq n \exp\{-cnb_n h_n^2\}, \tag{8.11}$$

which is summable if $b_n \geq d \log n/(nh_n^2)$ for all $d$ large enough.

An upper bound for (8.8) can be derived similarly. For each $r \leq x - h_n$, let $Y_{ni}(r) = I(r \leq X_i \leq x + h_n) = (x + h_n - r)\bar{f}_n([r, x + h_n])$ with mean $(x + h_n - r)\mu_n(r)$, and its average $\bar{Y}_n(r) = \sum_{i=1}^n Y_n(r)/n$. Using the stochastic ordering argument as above, we see that $P[\bar{f}_n([r, x + h_n]) \leq f(x + h_n/2)]$ is maximized when $r = x - h_n$. Let $\bar{Y}_n = \bar{Y}_n(x - h_n)$ and $\mu_n = \mu_n(x - h_n)$. Following the steps to obtain (8.9), we have $E(\bar{Y}_n) = 2h_n\mu_n$, where

$$\begin{aligned} \mu_n &= (1/2)[f(x - h_n) + f(x + h_n)](1 + o(1)) \\ &= f(x + h_n/2) - f'(x)h_n(1 + o(1)) \equiv f(x + h_n/2) + t_n. \end{aligned}$$

Proceeding as before using Bernstein's inequality, the probability in (8.8) is bounded above by

$$\begin{aligned} nP[\bar{f}_n([x-h_n, x+h_n]) \leq f(x+h_n/2)] &= nP[\bar{Y}_n - 2h_n\mu_n \leq 2h_n[f(x+h_n/2) - \mu_n] \\ &= nP[\bar{Y}_n - 2h_n\mu_n \leq -2h_nt_n] \\ &\leq n\exp\{-2nh_n^2t_n^2/(2h_n\mu_n + t_n)\} \\ &\leq n\exp\{-cnh_n^3\} \end{aligned}$$

for some $c$ (depending only on $f'(x)$) $> 0$ since $t_n = o(\mu_n)$ and $t_n \propto h_n(1+o(1))$. The last expression is summable if $nh_n^3/\log n \to \infty$.

Combining these 2 results we see that the probability of the event in (8.5) occurring infinitely often is 0. A symmetric argument shows that the event in (8.6) occurring infinitely often is 0 also. This completes the proof of the lemma. ∎

**Remark 1**. Lemma 1 holds if $h_n \downarrow 0$ arbitrarily slowly. If $\delta$ in $h_n = cn^{-\delta}$ goes to 0, the limiting random bandwidth gets close to $O_{a.s.}(\log n/n)$, the maximal spacing of the order statistics in a region where $f$ is bounded away from 0.

The next lemma assumes $h_n = O(n^{-1/3}\log n)$, that includes the case $h_n = 0$ when $f_n^*(x) = \hat{f}_n(x)$. It provides an explanation for the convergence rate of $n^{-1/3}$ in probability when $h_n = 0$; the factor $\log n$ comes in when we go form 'in probability' to 'almost sure' convergence.

**Lemma 2** *Assume $f'(x) < 0$ and continuous in a neighborhood of $x \in [h_n, X_{(n)} - h_n]$ and $h_n = O((\log n/n)^{-1/3})$. Then*

$$P[f_n^*(x) \neq f_n^\dagger(x) \quad i.o.] = 0$$

*if $b_n \geq d(\log n/n)^{1/3}$ for some $d > 0$.*

Proof. The technique used in the proof of Lemma 1 will fail when $h_n = 0$ (or $h_n$ is very small) because $P[\min_{r\leq x-h_n} \bar{f}_n([r, x+h_n]) \leq f(x+h_n/2)]$ in (8.8) reduces to $P[\min_{r\leq x} \bar{f}_n([r, x]) \leq$

$f(x)]$ that could be $P[f_n(x) \le f(x)]$ when $h_n = 0$ and $r = x$, and this probability could be very large. The steps of the proof are exactly the same that were used in the proof of Lemma 1, but with some modifications. We use the upper bound of $P[f_n^*(x) \ne f_n^\dagger(x)]$ as the sum of

$$P[\min_{r \le x - h_n} \bar{f}_n([r, x + h_n + b_n/2]) < \max_{s \ge x + h_n + b_n} \bar{f}_n((x + h_n + b_n/2, s])] \tag{8.12}$$

and

$$P[\min_{r \le x - h_n - b_n} \bar{f}_n([r, x - h_n - b_n/2]) < \max_{s \ge x + h_n + b_n} \bar{f}_n((x - h_n = b_n/2, s])] \tag{8.13}$$

using the same argument in the proof of Lemma 1. The probability in (8.12) is bounded above by the sum of

$$P[\max_{s \ge x + h_n + b_n} \bar{f}_n([x + h_n + b_n/2, s]) \ge f(x + h_n + b_n/2)] \tag{8.14}$$

and

$$P[\min_{r \le x - h_n} \bar{f}_n([r, x + h_n + b_n/2]) \le f(x + h_n + b_n/2)]. \tag{8.15}$$

To bound the probability in (8.14), define $Y_{ni}(s) = I(x + h_n + b_n/2 \le X_i \le s)$ and $\bar{Y}_n(s) = \sum_{i=1}^n Y_{ni}(s)/n$. Using the same stochastic ordering argument in the proof of Lemma 1, the probability in (8.14) is a maximum when $s = x + h_n + b_n$. Let $\bar{Y}_n = \bar{Y}_n(x + h_n + b_n)$ with mean $E(\bar{Y}_n) \equiv (b_n/2)\mu_n$, where

$$\begin{aligned} \mu_n &= (1/2)[f(x + h_n + b_n/2) + f(x + h_n + b_n)](1 + o(1)) \\ &= f(x + h_n + b_n/2) + (b_n/4)f'(x)(1 + o(1)) \equiv f(x + h_n + b_n/2) - t_n. \end{aligned}$$

Proceeding as in the proof of Lemma 1, the probability in (8.14) is bounded above by

$$\begin{aligned} nP(\bar{Y}_n - b_n\mu_n/2 \ge b_n t_n/2) &\le n \exp\{-n b_n^2 t_n^2/(2 b_n \mu_n + 2 b_n t_n)\} \\ &\le n \exp\{-c n b_n^3\} \end{aligned}$$

for some $c$ (depending only on $f'(x)$) $> 0$ since $t_n = o(\mu_n)$ and $b_n \propto t_n$. The last expression is summable if $b_n = d(\log n/n)^{-1/3}$ for a sufficiently large $d$.

Noting that

$$[f(x-h_n)+f(x+h_n+b_n/2)] = f(x+h_n+b_n/2) - (1/2)f'(x)(2h_n+b_n/2)(1+o(1)),$$

it is easy to show using the steps above that the probability in (8.15) is bounded above by

$$n\exp\{-cn(h_n+b_n/4)^3\}$$

for some $c$ (depending only on $f'(x)$) $> 0$. Since $h_n = O((\log n/n)^{-1/3})$, this expression is summable if $b_n = d(\log n/n)^{-1/3}$ for a sufficiently large $d$. Putting all of the results above, we have the probability of the event in (8.12) occurring infinitely often is 0. A symmetric argument shows this to be true for the event in (8.13). This completes the proof. ■

**Lemma 3** *Assume $|f'(x)| < 0$ and continuous in a neighborhood of $x$. Then*

$$\lim_{c\to\infty}\limsup_{n\to\infty} P[f_n^*(x) \neq f_n^{\dagger}(x)] = 0 \tag{8.16}$$

*if*

*(i) $nh_n^3 \to \infty$ and $b_n = c/(nh_n^2)$ for some $c > 0$ or*

*(ii) $h_n = O(n^{-1/3})$ and $b_n = dn^{-1/3}$ for some $d > 0$.*

*This implies that the widths of the random bands, $B_{1n}$ and $B_{2n}$ added to the fixed band $(x-h_n, x+h_n]$, in $f_n^*(x) = \bar{f}_n((x-h_n-B_{1n}, x+h_n+B_{2n}])$ are $O_p((1/(nh_n^2))$ in case (i) and $O_p(n^{-1/3})$ in case (ii).*

Proof. It is clear from the proofs of Lemmas 1 and 2 that $P[f_n^*(x) \neq \tilde{f}_n(x)]$ is bounded above by $\exp\{-dnb_nh_n^2\}$ under condition (i) for some $d > 0$ and by $\exp\{-dnb_n^3\}$ under condition (ii) for some $d > 0$ for all large $n$. This lemma follows from these observations. ■

We next show that $f_n^*$ and the kernel estimator $\tilde{f}_{nu}$ are very close. We do this only under the assumptions of Lemma 1 since that is the only case of importance to us.

**Theorem 1** *Assume* $0 < \inf_{0<x<t} |f'(x)|$ *and* $f'$ *is continuous on* $[0, t]$.

*(i) If* $nh_n^3/\log n \to \infty$ *then*

$$\sup_{h_n \le x \le t-h_n} |f_n^*(x) - \tilde{f}_{nu}(x)| = O_{a.s.}(\log n/(nh_n^2)).$$

*(ii) If* $nh_n^3 \to \infty$ *then* $|f_n^*(x) - \tilde{f}_{nu}(x)| = O_p(1/(nh_n^2))$ *uniformly in* $x \in (0, t)$.

PROOF. The estimator $f_n^*$ is a step function that jumps only if $x - h_n$ or $x + h_n$ is an observation point for $x \in [0, t]$. So it has at most $2n$ jump points. From the proof of Lemma 1, it is easy to see that

$$P[\cup_{\{h_n \le x \le X_{(n)} - h_n\}}\{f_n^*(x) \ne f_n^\dagger(x) \ i.o.\}] \le 8n^2 \exp\{-cnh_n^2 b_n\},$$

which is summable if $b_n = d \log n/(nh_n^2)$ for $d$ large enough. Now apply Glivenko-Cantelli lemma.

Now $f_n^*(x) = \bar{f}_n((x - h_n - B_{1n}, x + h_n + B_{2n}])$, where $B_{1n}$ and $B_{2n}$ are random band widths added to the fixed band $(x - h_n, x + h_n]$. From Lemma 1, these random bandwidths are $O_{a.s.}(\log n/(nh_n^2)) = o_{a.s.}(h_n)$ under our assumptions in (i), and they are $O_p(1/(nh_n^2)) = o_p(h_n)$ under our assumptions in (ii). Thus, $|f_n^*(x) - f^\dagger(x)|$ and $|f_n^*(x) - \tilde{f}_{nu}(x)|$ are of the same asymptotic order. This completes the proof. ■

# CHAPTER 9

## ESTIMATION OF $f(0)$

A kernel estimator of $f(0)$ is biased up because of lack of observations to the left of 0. Our estimator reduces this bias.

**Theorem 2** *Assume $f''$ exists and is continuous near 0. Then*

*(i)* $f_n^*(0) \xrightarrow{a.s.} f(0)$ *if* $nh_n^3/\log n \to \infty$ *and*

*(ii)* $f_n^*(0) \xrightarrow{p} f(0)$ *if* $nh_n^3 \to \infty$.

Proof. Assume $X_n > 2h_n$. Using Theorem 1 and Law of the iterated logarithm, we have

$$\begin{aligned}
f_n^*(0) &= 2f_n^*(h_n) - f_n^*(2h_n) \\
&= 2\frac{F_n(2h_n)}{2h_n} - \frac{F_n(3h_n) - F_n(h_n)}{2h_n} + O_{a.s.}(\log n/(nh_n^2)) \\
&= \frac{F(h_n) + 2F(2h_n) - F(3h_n) + O_{a.s.}(\sqrt{\log\log n/n})}{2h_n} + O_{a.s.}(\log n/(nh_n^2)) \\
&= \frac{F(h_n) + 2F(2h_n) - F(3h_n)}{2h_n} + O_{a.s.}(\log n/(nh_n^2)) \\
&\equiv Y_n + O_{a.s.}(\log n/(nh_n^2)). \qquad (9.1)
\end{aligned}$$

By Taylor series expansion, we have

$$f_n^*(0) = f(0) - (5/6)f''(0)h_n^2(1+o(1)) + O_{a.s.}(\log n/(nh_n^2)) \xrightarrow{a.s.} f(0), \qquad (9.2)$$

proving (i). Note that our bias-corrected definition of $f_n^*(0)$ removes the first order bias term.

We could replace the last display by

$$f_n^*(0) = f(0) - (5/6)f''(0)h_n^2(1+o(1)) + O_p(1/(nh_n^2)) \xrightarrow{p} f(0), \qquad (9.3)$$

using Lemma 3 instead of Lemma 1 and the fact that $||F_n - F|| = O_p(1/\sqrt{n})$. This completes the proof of the theorem. ■

## 9.1 Asymptotic distribution of $f_n^*(0)$

Note that $Y_n$ in (9.1) could be written as

$$Y_n = [F_n(h_n) + 2F_n(2h_n) - F_n(3h_n)]/(2h_n) =$$

$$\frac{1}{2nh_n}\sum_{i=1}^{n}[2I(0 < X_i < h_n) + I(h_n \le X_i < 2h_n) - I(2h_n \le X_i \le 3h_n)]$$

is the average of independent RVs. Since the summands in the definition of $Y_n$ are uniformly bounded, Lyapunov's central limit theorem yields

$$\sqrt{nh_n}[Y_n - E(Y_n)] \text{ is asymptotically normal with mean 0 and variance } Var(f_n^*(0)). \quad (9.4)$$

Using the boundedness of the summands in the definition of $Y_n$, the expression for $f_n^*(0)$ in (9.3), and Taylor's expansion of $F(x)$ near 0,

$$\begin{aligned} E[f_n^*(0)] &= f(0) - (5/6)f''(0)h_n^2(1 + o(1)) + O(1/(nh_n^2)) \\ &= f(0) - (5/6)f''(0)h_n^2(1 + o(1)) \quad \text{if } nh_n^4 \to \infty. \end{aligned} \quad (9.5)$$

For future reference, note that $\sqrt{nh_n}E[f_n^*(0) - f(0)] = \sqrt{nh_n^5}(5/6)f''(0)(1 + o(1))$.

We now compute the variance of $f_n^*(0)$ assuming $nh_n^4 \to \infty$. In this case

$$4nh_n^2 Var[f_n^*(0)] \;=\; 4nh_n^2 Var(Y_n) = 6h_n f(0)(1 + o(1))$$

using the first order Taylor expansion of $F(x)$ near 0. Putting this together with (9.4) and (9.5) yields the following theorem.

**Theorem 3** *Assume $f''(x)$ exists and is continuous near 0, $nh_n^4 \to \infty$, and $h_n^5 = cn^{-1/5}$. Then*

$$\sqrt{nh_n}[f_n^*(0) - f(0)] \xrightarrow{d} N(-(5/6)cf''(0), (3/2)f(0)).$$

*If $nh_n^5 \to 0$, then the asymptotic mean is 0.*

This shows that $f_n^*(0)$ substantially improves on past estimators of $f(0)$. In addition, asymptotic normality implies the asymptotic bias could be estimated by standard bootstrap methods.

# CHAPTER 10

## ASYMPTOTIC DISTRIBUTION OF $f_n^*(x)$

We now consider the asymptotic distribution of $f_n^*(x)$ when $x$ is an interior point of the support of $F$. The following theorem is a direct consequence of the equivalence theorem, Theorem 2 and standard results on kernel density estimation with the uniform kernel.

**Theorem 4** *Let* $h_n = cn^{-\delta}$ *with* $1/5 \le \delta < 1/3$ *for some* $c > 0$. *Assume* $f''$ *exists continuously in a neighborhood of* $x$ *in the interior of the support of* $F$ *where* $f'(x) < 0$. *Then*

$$\sqrt{nh_n}[f_n^*(x) - f(x)] \xrightarrow{d} N(\mu(x), (1/2)f(x)),$$

*where* $\mu(x) = (1/2)f''(x)$ *if* $\delta = 1/5$, *and it is 0 if* $1/5 < \delta < 1/3$.

PROOF. Since $\sqrt{nh_n}|f_n^*(x) - \hat{f}_{nu}(x)| = o_p(1)$ by our assumptions and Theorem 2.1, the theorem follows from the fact the asymptotic distribution holds if $f_n^*$ is replaced by $\hat{f}_{nu}$ from standard theory of kernel density estimation. ■

**Remark 2** (i) The asymptotic normality implies we could estimate the asymptotic bias by standard bootstrap methods.

(ii) Note that the asymptotic variance of $f_n^*(0)$ is 3 times larger than that of $f_n^*(x)$ when $x$ is an interior point. This is because the advantage of a symmetric kernel cannot be utilized in the former case.

# CHAPTER 11

## SIMULATIONS AND GRAPHS

In this section we compare our estimator of $f(0)$ with that of Woodroofe and Sun (1993). The superiority of our estimator is obvious. For each of the following data tables, $f_n^*(0+)$ was calculated as follows:

$$f_n^*(0+) = \frac{F_n(h_n) + 2F_n(2h_n) - F_n(3h_n)}{2h_n},$$

where $h_n$ was first chosen to be a fixed bandwidth of .1, .2, .3, .4, .5, .6 for both the standard exponential case and the half-normal case (10,000 Simulations for all trials). The first row of data for each sample size records the average of the ratio and the second row records the standard deviation of the ratio. Then to attempt to find a more optimal bandwidth but still fixed for all simulations and all sample sizes, we ran 1,000,000 simulations and found the average bandwidth R would choose for each case. For example, for the standard exponential case and samples of size 200, the average bandwidth of 1,000,000 simulations was $h_n = .253939$. We then re-ran the ratio to see how these bandwidths performed.

TABLE 11.1

$f_n^*(0+)/f(0+)$ FOR THE STANDARD EXPONENTIAL CASE; BANDWIDTH FIXED

| Sample Size | $h_n = .1$ | $h_n = .2$ | $h_n = .3$ | $h_n = .4$ | $h_n = .5$ | $h_n = .6$ |
|---|---|---|---|---|---|---|
| $n = 25$ | .99176 | .97436 | .9477267 | .912525 | .879508 | .84562 |
| | .7142822 | .4605647 | .3400202 | .270054 | .2165884 | .179487 |
| $n = 50$ | .99546 | .97292 | .9460133 | .9204925 | .879708 | .84495 |
| | .4992918 | .3245415 | .2392624 | 0.1898927 | .1533053 | .126064 |
| $n = 100$ | .997955 | .970445 | .94669 | .9167112 | .881658 | .8462308 |
| | .3599723 | .2299077 | .1722326 | .1351591 | .1077855 | .08956797 |
| $n = 200$ | .9937 | .97367 | .9460083 | .9149131 | .8786635 | .8450104 |
| | .2523894 | .1628617 | .1218916 | 0.0952496 | .07674539 | 0.06411635 |

TABLE 11.2

$f_n^*(0+)/f(0+)$ FOR THE STANDARD EXPONENTIAL CASE; BANDWIDTH SIMULATED

| Sample Size | $h_n = .253939$ | $h_n = .2887531$ | $h_n = ..3242278$ | $h_n = .356044$ |
|---|---|---|---|---|
| $n = 25$ | .9648301 | .9550512 | .9424732 | .9267619 |
| | 0.387007 | .3520133 | .322408 | .2999453 |
| $n = 50$ | .9594155 | .9529872 | .9405671 | .9305591 |
| | .2722727 | .2483402 | .227145 | .209288 |
| $n = 100$ | .9606126 | .9521387 | .9406858 | .9301308 |
| | .1931686 | .1747815 | .1607034 | .1488294 |
| $n = 200$ | .9628375 | .9507604 | .9386071 | .9308512 |
| | .1388276 | .1231871 | .1130876 | .1042787 |

TABLE 11.3

$f_n^*(0+)/f(0+)$ FOR THE HALF-NORMAL CASE; BANDWIDTH FIXED

| Sample Size | $h_n = .1$ | $h_n = .2$ | $h_n = .3$ | $h_n = .4$ | $h_n = .5$ | $h_n = .6$ |
|---|---|---|---|---|---|---|
| $n = 25$ | 1.014884 | 1.032844 | 1.055357 | 1.092821 | 1.104365 | 1.109743 |
| | 0.8378116 | 0.5674886 | 0.4350507 | 0.3454052 | 0.2802226 | 0.2242692 |
| $n = 50$ | 1.012703 | 1.030469 | 1.059163 | 1.084339 | 1.10697 | 1.108613 |
| | 0.5926255 | 0.3985238 | 0.3084085 | 0.246492 | 0.1973042 | 0.1590009 |
| $n = 100$ | 1.00962 | 1.029275 | 1.058589 | 1.085672 | 1.105091 | 1.110593 |
| | 0.4194865 | 0.2829294 | 0.2172678 | 0.1755953 | 0.1390437 | 0.1124586 |
| $n = 200$ | 1.008774 | 1.029386 | 1.058021 | 1.086631 | 1.104311 | 1.109237 |
| | 0.2967001 | 0.2003904 | 0.1533127 | 0.1241067 | 0.09900316 | 0.07914662 |

TABLE 11.4

$f_n^*(0+)/f(0+)$ FOR THE HALF-NORMAL CASE; BANDWIDTH SIMULATED

| Sample Size | $h_n = .2306908$ | $h_n = .2606989$ | $h_n = .2918519$ | $h_n = .3203207$ |
|---|---|---|---|---|
| $n = 25$ | 1.041841 | 1.051038 | 1.061735 | 1.064282 |
| | 0.5173063 | 0.4815065 | 0.4362699 | 0.4095538 |
| $n = 50$ | 1.033702 | 1.046808 | 1.05404 | 1.067881 |
| | 0.3686704 | 0.3383564 | 0.3106273 | 0.2910296 |
| $n = 100$ | 1.036652 | 1.048745 | 1.055659 | 1.063368 |
| | 0.2582715 | 0.2398753 | 0.2223498 | 0.2092851 |
| $n = 200$ | 1.037569 | 1.048787 | 1.058102 | 1.064027 |
| | 0.1808386 | 0.1684117 | 0.1569683 | 0.1465812 |

TABLE 11.5

THE HALF-NORMAL CASE

| | $\int \|\hat{f} - f\| \, dx$ | $\int \|f_n^* - f\| \, dx$ | $\int \|f_n^{PL} - f\| \, dx$ | $\frac{\hat{f}(0+)}{f(0+)}$ | $\frac{f_n^*(0+)}{f(0+)}$ | $\frac{f_n^{PL}(0+)}{f(0+)}$ |
|---|---|---|---|---|---|---|
| n | | $h_n = n^{-1/5}$ | $\alpha = \frac{\log n}{n}$ | | $h_n = n^{-1/5}$ | $\alpha = \frac{\log n}{n}$ |
| 50 | 0.223 | 0.147 | 0.191 | 8.310 | 1.084 | 1.061 |
| | 0.063 | 0.058 | 0.057 | 65.07 | 0.246 | 0.188 |
| 100 | 0.177 | 0.119 | 0.159 | 7.904 | 1.086 | 1.059 |
| | 0.045 | 0.043 | 0.041 | 54.96 | 0.176 | 0.155 |
| 200 | 0.138 | 0.095 | 0.130 | 7.637 | 1.064 | 1.060 |
| | 0.032 | 0.033 | 0.030 | 55.12 | 0.147 | 0.130 |

TABLE 11.6

THE STANDARD EXPONENTIAL CASE

| | $\int \|\hat{f} - f\| \, dx$ | $\int \|f_n^* - f\| \, dx$ | $\int \|f_n^{PL} - f\| \, dx$ | $\frac{\hat{f}(0+)}{f(0+)}$ | $\frac{f_n^*(0+)}{f(0+)}$ | $\frac{f_n^{PL}(0+)}{f(0+)}$ |
|---|---|---|---|---|---|---|
| n | | $h_n = n^{-1/5}$ | | | $h_n = n^{-1/5}$ | $\alpha = \frac{\log n}{n}$ |
| 50 | 0.247 | 0.191 | 0.231 | 8.61 | 0.920 | 0.887 |
| | 0.058 | 0.063 | 0.052 | 67.60 | 0.190 | 0.184 |
| 100 | 0.199 | 0.147 | 0.191 | 14.26 | 0.917 | 0.913 |
| | 0.041 | 0.044 | 0.038 | 352.74 | 0.135 | 0.157 |
| 200 | 0.160 | 0.113 | 0.156 | 12.17 | 0.931 | 0.942 |
| | 0.029 | 0.032 | 0.028 | 218.74 | 0.104 | 0.137 |

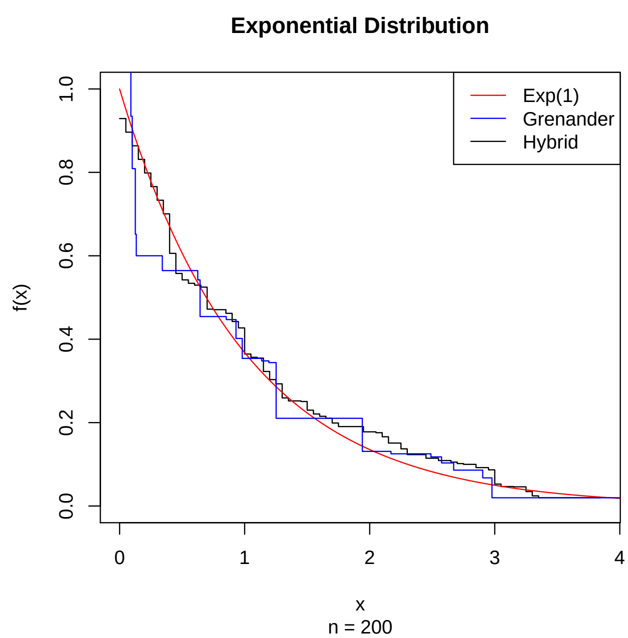


FIGURE 11.1

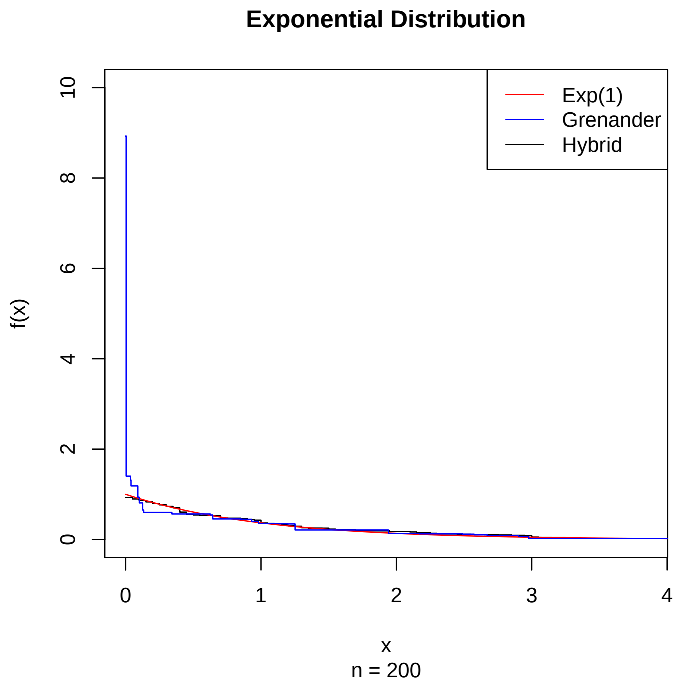


FIGURE 11.2

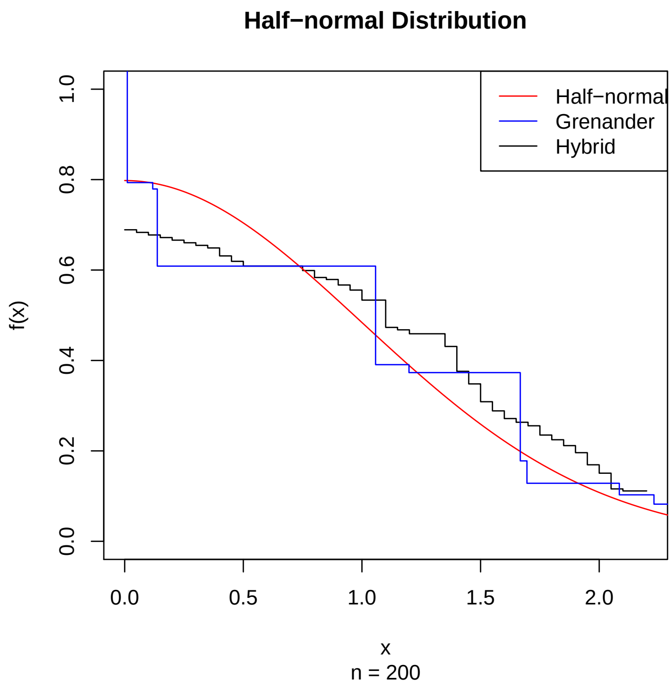


FIGURE 11.3

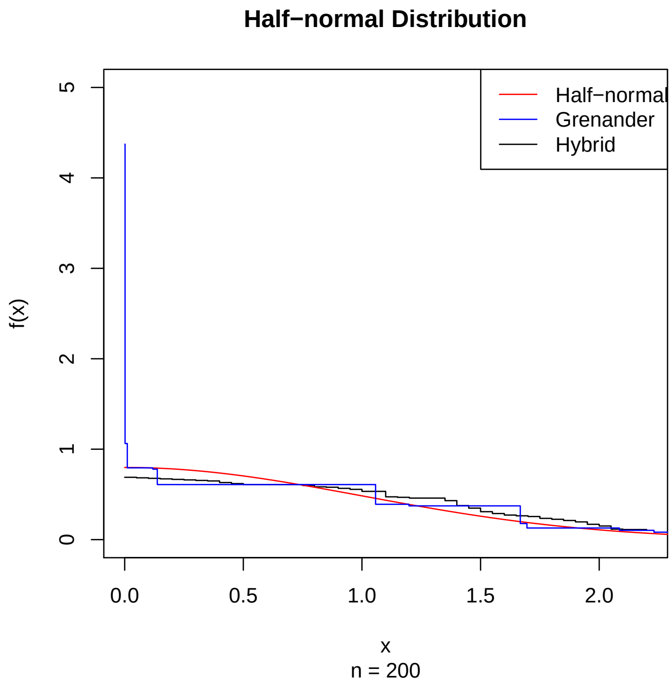


FIGURE 11.4

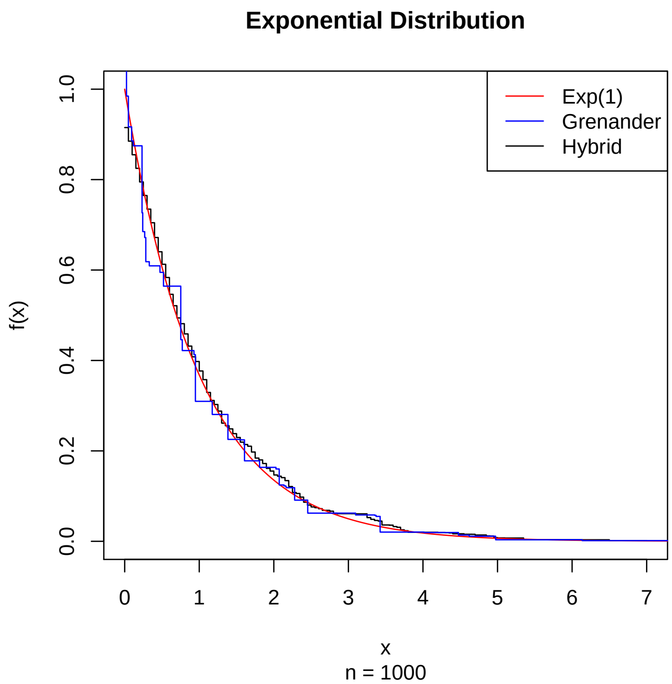


FIGURE 11.5

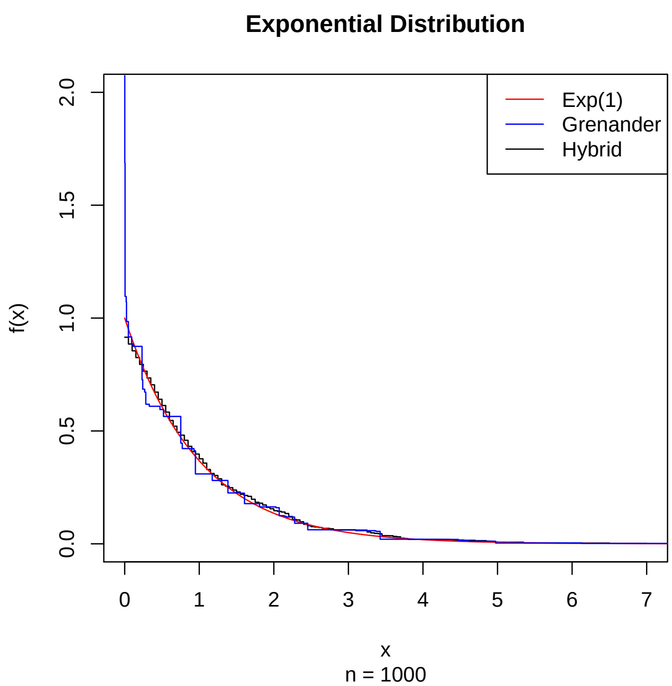


FIGURE 11.6

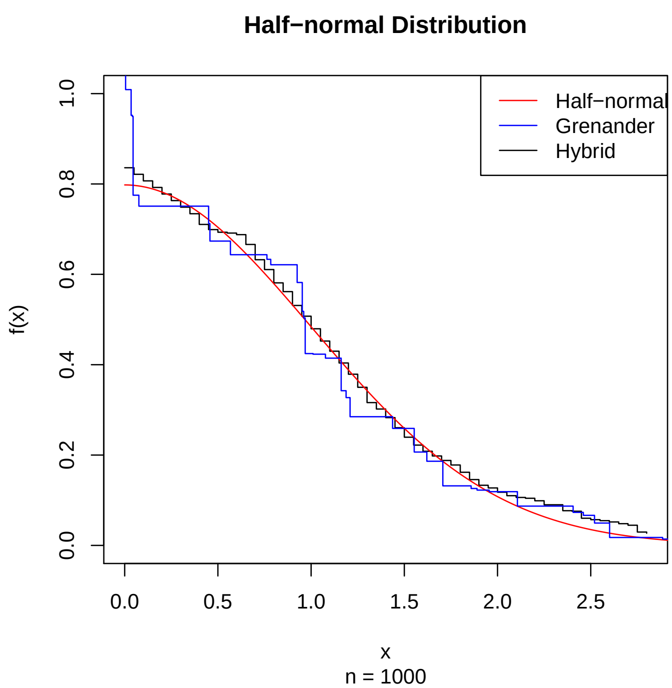


FIGURE 11.7

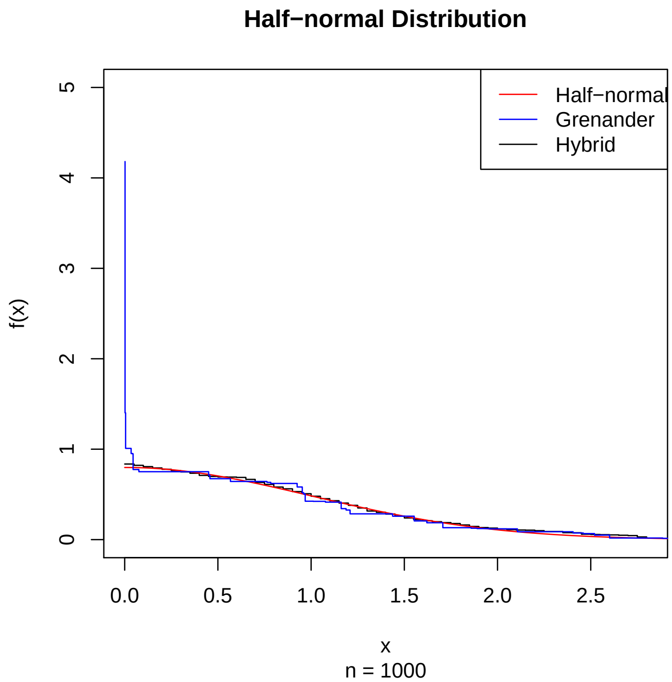


FIGURE 11.8

TABLE 11.7

THE HALF-NORMAL $E[f_n^*(x) - f(x)]^2$ WITH A SAMPLE OF SIZE 200 AND $h_n = .35$

| $x$ | MSE |
|---|---|
| 0.3 | 0.0030 |
| 0.6 | 0.0019 |
| 1 | 0.0019 |
| 2 | 0.0012 |
| 3 | 0.0003 |

TABLE 11.8

THE STANDARD EXPONENTIAL $E[f_n^*(x) - f(x)]^2$ WITH A SAMPLE OF SIZE 200 AND $h_n = .35$

| $x$ | MSE |
|---|---|
| 0.3 | 0.0037 |
| 0.6 | 0.0024 |
| 1 | 0.0019 |
| 2 | 0.0008 |
| 3 | 0.0003 |

# CHAPTER 12

# CONCLUDING REMARKS AND FUTURE WORK

The Grenander (1956) NPMLE of a decreasing density has an $n^{-1/3}$ convergence rate and has an unfamiliar asymptotic distribution, making statistical inferences difficult. Nonparametric kernel estimators on the other hand has $n^{-2/5}$ convergence rate and is asymptotically normal. We provide a hybrid estimator that utilizes both the concepts of a nonparametric density estimation and the Grenader estimation that guarantees monotonicity of the estimator. By analysis and simulations we were able to show a quantum improvement. Our methods could be used to improve other nonparametric estimations of monotone functions, e.g., a monotone hazard rate.

BIBLIOGRAPHY

## BIBLIOGRAPHY


Balabdaoui, F., Jankowski, H., Pavlides, M., Seregin, A., & Wellner, J. (2011). On the Grenander estimator at zero. *Statistica Sinica*, 21(2), 873.

Balabdaoui, F., & Wellner, J. A. (2007). A Kiefer-Wolfowitz theorem for convex densities. *Institute of Mathematical Statistics Lecture Notes - Monograph Series Asymptotics: Particles, Processes and Inverse Problems*, 1 - 31.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

Behnen, K., & Neuhaus, G. (1983). Galton's test as a linear rank test with estimated scores and its local asymptotic efficiency. *Ann. Statist.* **11**, 588-599.

Bohn, L. L., & Wolfe, D. A. (1992). Nonparametric two-sample procedures for ranked-set samples data. *J. Amer. Statist. Assoc.* **87**, 552-561.

Boyles, R. A., Marshall, A. W., & Proschan, F. (1985). Inconsistency of the Maximum Likelihood Estimator of a Distribution Having Increasing Failure Rate Average. *The Annals of Statistics*, **13**(1), 413-417.

Brunk, H. D., Franck, W. E., Hanson, D. L. & Hogg, R. V. (1966). Maximum likelihood estimation of the distributions of two stochastically ordered random variables. *J. Am. Statist. Assoc.* **61**, 1067-80.

Chacko, V. J. (1963). Testing homogeneity against ordered alternatives. *Ann. Math. Statist.*, **34**, 945-956.

Cuzick, J. (1985). A Wilcoxon-type test for trend. *Statistics in Medicine*, **4**, 87-90.

Desu, M. M., Park, S., & Chakraborti, S. (1996). Linear rank statistics for the simple tree alternatives. *Biometrical J.*, **38**(3), 359-373.

Dykstra, R. L. Madsen, R.W., & Fairbanks, K. (1983). A nonparametric likelihood ratio test. *J. Statist. Comput. Simulation*, **18**, 247-264.

El Barmi, H. & McKeague, I. W. (2013). Empirical likelihood-based test for stochastic ordering. *Bernoulli*, **19**(1), 295-307.

El Barmi, H. & Mukerjee, H. (2005). Inferences under a stochastic ordering constraint: The $k$-sample case. *J. Amer. Statist. Assoc.* **100**, 252?261.

Fligner, M. A., & Wolfe, D. A. (1982). Distribution-free tests for comparing several treatments with a control. *Statist. Neerlandica*, **36**, 119-127.

Franck, W. E. (1984). A likelihood ratio test for stochastic ordering. *J. Amer. Statist. Assoc.* **79** 686-691.

Grenander, U. (1956). On the theory of mortality measurement. II. *Scandinavian Actuarial Journal*, 1956(2), 125-153.

Groeneboom, P. (1985), Estimating a monotone density, in: *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer Vol. II* (Berkeley, CA. 1983), Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA.

Groeneboom, P., Hooghiemstra, G., & Lopuhaä, H. P. (1999). Asymptotic normality of the $L_1$ error of the Grenander estimator. The Annals of Statistics, **27**(4), 1316-1347.

Groeneboom, P., & Jongbloed, G. (2014). *Nonparametric estimation under shape constraints estimators, algorithms, and asymptotics.* New York, NY, USA: Cambridge University Press.

Groeneboom, P., G. Jongbloed, & J. A. Wellner (2001), Estimation of a convex function: characterizations and asymptotic theory, *The Annals of Statistics* **29**, 1653 - 1698.

Handcock, M. S., & Morris, M. (1999). *Relative distribution methods in the social sciences.* Springer-Verlag, New York.

Hogg, R.V. (1962). Iterated tests of the equality of several distributions. *J. Amer. Statist. Assoc.* **57**, 579?585.

Janic-Wróblewska, A. & Ledwina, T. (2000). Data driven rank test for two-sample problem. *Scand. J.Statist.* **27**, 281?297.

Jonkheere, A. R. (1954). A distribution-free $k$-sample test against ordered alternatives. *Biometrika*, **41**, 133-145.

Kiefer, J., & Wolfowitz, J. (1976). Asymptotically Minimax Estimation of Concave and Convex Distribution Functions. *Collected Papers*, 739-751.

Kosorok, M. R. (2008). Bootstrapping the Grenander estimator. *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen Collections*, 282-292.

Kulikov, V. N., & Lopuhaä, H. P. (2005). Asymptotic normality of the L k-error of the Grenander estimator. *The Annals of Statistics*, **33**(5), 2228-2255.

Kulikov, V. N., & Lopuhaä, H. P. (2006). The limit process of the difference between the empirical distribution function and its concave majorant. *Statistics & Probability Letters,* **76**(16), 1781 - 1786.

Kulikov, V. N., & Lopuhaä, H. P. (2007). Distribution of Global Measures of Deviation Between the Empirical Distribution Function and Its Concave Majorant. *Journal of*

*Theoretical Probability*, **21**(2), 356-377.

Le, C.T. (1988) A new rank test against ordered alternatives in $k$-sample problems. *Biometrical Journal, Journal of Mathematical Methods in Biosciences*, **30**, 87-92.

Ledwina, T., & Wy?upek, G. (2012). Two-Sample Test Against One-Sided Alternatives. *Scandinavian Journal of Statistics*, **39**(2), 358-381.

Lee, Y. J. & Wolfe, D. A. (1976). A distribution-free test for stochastic ordering. *J. Amer. Statist. Assoc.* **71**, 722-727.

Lehmann, E. L. (1955). Ordered families of distributions. *Ann. Math. Statist.* **26**, 399-419.

Mack, G. A. & Wolfe, D. A. (1981). $k$-sample rank test for umbrella alternatives (corr: V77 p954). *Journal of the American Statistical Association*, **76**, 175-181.

Mahrer, J. & Magel, R. (1995). A comparison of tests for the $k$-sample nondecreasing alternative. *Statistics in Medicine*, **14** 863-871.

Magel, R. (1988). A test for the equality of $k$ medians against the simple tree alternative under right censorship. *Communications in Statistics, Part B-Simulation and Computation*, **17**, 917-925.

Mau, J. (1988). A generalization of a nonparametric test for stochastically ordered distributions to censored survival data. *J. Roy. Statist. Soc. Ser. B*, **50**, 403?412.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser.A.* **31**(1), 23-36

Robertson, T., & Wright, F. T. (1974). A Norm Reducing Property for Isotonized Cauchy Mean Value Functions. *The Annals of Statistics*, **2**(6), 1302-1307.

Robertson, T., & Wright, F. T. (1981). Likelihood ratio tests for and against a stochastic ordering between multinomial populations. *Ann. Statist.* **9**, 1248-1257.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inferences*, Wiley, New York.

Rychlik, T. (2001). *Projecting statistical functionals.* Springer, New York.

Sen, B., Banerjee, M., & Woodroofe, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *The Annals of Statistics*, **38**(4), 1953-1977.

Sun, J., & Woodroofe, M. (1993). A penalized maximum likelihood estimate of $f(0+)$ when $f$ is non-increasing. *Statistica Sinica*, **3**, 501-515.

Sun, J., & Woodroofe, M. (1996). Adaptive smoothing for a penalized NPMLE of a non-increasing density. *Journal of Statistical Planning and Inference*, **52**(2), 143-159.

Terpstra, T. (1952). Asymptotic normality and consistency of Kendall's test against trend when ties are present in one ranking. *Indagationes Mathematica*, **14**, 337-333.

Terpstra, J. T. & Magel, R. C. (2003). A new nonparametric test for the ordered alternative problem. *Nonparametric Statistics*, **15**(3), 289-301.

Uspenskij, J. V. (1937). Introduction to mathematical probability. New York: McGraw-Hill.

Wang, Y. (1994). The limit distribution of the concave majorant of an empirical distribution function. *Statistics & Probability Letters*, **20**(1), 81-84.

Wy?upek, G. (2010). Data driven k-sample tests. *Technometrics*, **52**, 107?123.

APPENDICES

# APPENDIX A

## CODE FOR OBTAINING TEST STATISTIC VALUES VIA BROWNIAN BRIDGE SIMULATIONS

The following code was created by implementing the procedure described in Theorem 1 of El Barmi and McKeague (2013) and described in (2.32). The reported simulation values are used as the test statistic values for $T_n$ described in (2.30).

```
# Represents the number of simulations run.
nsim = 100000
# Ininitializes a storage space for the values of the integral.
I = rep(0, nsim)
# Creates a loop to calculate the quantile values for the test statistic
for(i in 1:nsim){
N = 100000
t = seq(0, 1, 1/N)
x < − c(0, cumsum(rnorm(N))/sqrt(N))
b < − x − t*x[N+1]
bp < − apply(cbind(b,0), 1, max)
I[i] < − sum(((bp)ˆ2/(t*(1-t))*1/N)[2:N])
}
quantile(I, c(.90, .95, .99))
```

# APPENDIX B

## CODE FOR CORRECTING TABLE 2 OF EL BARMI AND MCKEAGUE'S PAPER

The following code was created to calculate values of the test statistic proposed in (2.30). To obtain all other relevant values, vary sample sizes and distribution functions as needed. The author created this code after realizing that the power comparisons in Table 2 of El Barmi and McKeague's 2013 paper were inaccurate. A corrected table appears in Chapter 2.

Alpha = 0.05

t = 1.872133

m = 50 # Sample Size for $F_1$

n = 30 # Sample Size for $F_2$

Sim = 10000 # Number of Simulations

Xm $< -$ array(0, c(Sim, m)) # Initialize an array to store the data corresponding to $F_1$ (Sim by m)

Yn $< -$ array(0, c(Sim, n)) # Initialize an array to store the data corresponding to $F_2$ (Sim by n)

TD $< -$ array(0, c(Sim, m + n)) # Initialize an array to store the total data set (Sim by m+n)

TDA < − array(0, c(Sim, m + n)) # Initialize an array to store the ascending total data set (Sim by m+n)

FH < − array(0, c(Sim, m + n)) # Initialize an array to store the empirical cdf of the pooled samples (Sim by m+n)

F1H < − array(0, c(Sim, m + n)) # Initialize an array to store the empirical cdf of the 1st samples (Sim by m+n)

F2H < − array(0, c(Sim, m + n)) # Initialize an array to store the empirical cdf of the 2nd samples (Sim by m+n)

F1T < − array(0, c(Sim, m + n)) # Initialize an array to store the weighted projection of $\widehat{F}_1$ (Sim by m+n)

F2T < − array(0, c(Sim, m + n)) # Initialize an array to store the weighted projection of $\widehat{F}_2$ (Sim by m+n)

R < − array(0, c(Sim, m + n)) # Initialize an array to store the values of R for each x

T12 < − array(0, c(Sim, 2)) # Initialize an array to store the values of the test statistic and the decision (1 or 0)

for (i in 1:Sim){

Xm[i ,] < − (runif(m, 0, 1)) # Choose $F_1$ = Uniform(0, 1)

Yn[i ,] < − (runif(n)) # Choose $F_2$ = Uniform(0, 1)

TD[i ,] < − c(Xm[i ,], Yn[i ,]) # Combine all the data for each simulation

TDA[i ,] < − rbind(TD[i ,][order(TD[i ,])]) # Order the data for each simulation

}

for (i in 1:Sim){

FH[i ,] < − order(TDA[i ,])/(m + n) # Calculate the empirical cdf of the pooled samples

for(j in 1:(m + n)){

F1H[i, j] < − sum(Xm[i ,] <= TDA[i, j])/m # Calculate the empirical cdf of the 1st samples

F2H[i, j] < − sum(Yn[i ,] <= TDA[i, j])/n # Calculate the empirical cdf of the 2nd samples

}

F1T[i ,] = ifelse(F1H[i ,] <= F2H[i ,], F1H[i ,], (F1H[i ,] + F2H[i ,])/2) # Calculate the weighted projection of $\widehat{F}_1$

F2T[i ,] = ifelse(F1H[i ,] <= F2H[i ,], F2H[i ,], (F1H[i ,] + F2H[i ,])/2) # Calculate the weighted projection of $\widehat{F}_2$

}

for (i in 1:Sim){

R[i ,] < − ( (FH[i ,]/F1T[i ,])ˆ(m*F1H[i ,]) )*( ((1 - FH[i ,])/(1 - F1T[i ,]))ˆ(m*(1 - F1H[i ,])) )*( (FH[i ,]/F2T[i ,])ˆ(n*F2H[i ,]) )*( ((1 - FH[i ,])/(1 - F2T[i ,]))ˆ(n*(1 - F2H[i ,])) )# Calculate the values of R for each x

}

for(i in 1:Sim){

T12[i, 1] = -2*sum(log(R[i ,])/(m + n)) # Calculates the value of the test statistic for each simulation (using the natural logarithm)

T12[i, 2] = ifelse(T12[i, 1] ¿ t, 1, 0) # Returns 1 if we reject the null hypothesis, Returns 0 if we fail to reject the null hypothesis (using an alpha level of 0.05)

}

P12 = sum(T12[, 2])/Sim; P12 # Calculates the percentage of the time we reject the null hypothesis